

**Anais da Conferência
Nacional em
Comunicações, Redes e
Segurança da
Informação (ENCOM)**

Anais da Conferência Nacional em
Comunicações, Redes e Segurança
da Informação (ENCOM).

Anais...Natal(RN) IFRN, 2024

Dados Internacionais de Catalogação na Publicação (CIP)

AN532 Anais da Conferência Nacional em Comunicações, Redes e Segurança da Informação (ENCOM). Anais...Natal(RN) IFRN, 2024

Disponível em www.even3.com.br/anais/encom2024

ISSN: 2237-5112

DOI: doi.org/10.29327/22375112

1. Ciência da computação, informação e obras gerais 2. Tecnologia
(ciências aplicadas)

IFRN

CDD - 370

CORPO EDITORAL

COORDENAÇÃO GERAL

FRANCISCO SALES DE LIMA FILHO (IFRN)

ALESSANDRO JOSÉ DE SOUZA (IFRN)

COORDENAÇÃO DO COMITÊ DE PROGRAMA

FELIPE SAMPAIO DANTAS DA SILVA (IFRN)

COORDENAÇÃO DE DIVULGAÇÃO

DANIELLE GOMES DE FREITAS MEDEIROS (IFRN)

COORDENAÇÃO LOCAL

SARAH THOMAZ DE SA ROSSITER (IFRN)

MEMBROS DO COMITÊ DE PROGRAMA

ALEXANDRE GOMES DE LIMA (IFRN)

ANDERSON CAVALCANTI (UFRN)

ANDRECIA PEREIRA DA COSTA (UFPA)

CHARLES HALLAN FERNANDES DOS SANTOS (IFRN)

DIEGO DA SILVA PEREIRA (IFRN)

EDMAR JOSÉ DO NASCIMENTO (UNIVASF)

ELIEL POGGI DOS SANTOS (UFT)

EMIDIO DE PAIVA NETO (PICPAY)

FRANCISCO ARY ALVES DE SOUZA (IFRN)

HELBER WAGNER DA SILVA (IFRN)

JORGE FREDERICSON DE MACEDO COSTA DA SILVA (IFCE)

LEONARDO MARTINS DE LIMA (SODEXO)

LUIZ FELIPE DE QUEIROZ SILVEIRA (UFRN)

LUIZ GONZAGA DE QUEIROZ SILVEIRA JÚNIOR (UFRN)

MATHEWS PHILLIPP SANTOS DE LIMA (UFRN)
PEDRO THIAGO VALERIO DE SOUZA (UFERSA)
PLÁCIDO ANTONIO DE SOUZA NETO (IFRN)
RAFAEL JULLIAN OLIVEIRA DO NASCIMENTO (UFRN)
ROMÁRIO DA COSTA SILVA (UFPA)
TADEU FERREIRA OLIVEIRA (IFRN)
THIAGO HENRIQUE FREIRE DE OLIVEIRA (IFRN)

MEMBROS DA COORDENAÇÃO DE CAPTAÇÃO DE RECURSOS

ANDERSON LUIZ DE OLIVEIRA CAVALCANTI (UFRN)
FABRÍCIO GERÔNIMO SIMÕES SILVA (IFBA)
JOEL EUGÊNIO CORDEIRO JUNIOR
MARCELO SAMPAIO DE ALENCAR (IECOM)

MEMBROS DA COORDENAÇÃO DE DIVULGAÇÃO

ANDERSON LUIZ DE OLIVEIRA CAVALCANTI (UFRN)
AUGUSTO JOSÉ VENÂNCIO NETO (UFRN)
EVERSON MIZAEEL CORTEZ SILVA (IFRN)
JORGE FREDERICSON COSTA DA SILVA (IFCE)
LUIZ FELIPE DE QUEIROZ SILVEIRA (UFRN)

Horizon: Adicionando Live Analysis Forense ao IPED

João Pedro Barbosa Medeiros*, Galileu Batista de Sousa*, Francisco Sales de Lima Filho*, Carlos Gustavo Rocha*

*Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, Brasil

horizon.resolved973@passinbox.com

Resumo—Este artigo apresenta o Horizon, uma ferramenta que adiciona o recurso de live analysis ao sistema Indexador e Processador de Evidências Digitais (IPED) por meio de um conjunto de plugins de comunicação com o software Volatility. Os experimentos mostraram que a solução proposta reduz tempo da etapa de processamento de evidências, facilita o trabalho pericial e simplifica a integração entre ferramentas de análise forense.

Palavras-chave—Análise forense; Processamento de vestígios; Análise de memória; live analysis; IPED; Volatility.

I. INTRODUÇÃO

A investigação de crimes envolvendo computadores é denominada como perícia forense e requer a análise de dados ou metadados armazenados nas mídias desses dispositivos, tais como discos rígidos, *pendrives* e memória RAM [1]. Os peritos que realizam essa atividade usam ferramentas específicas e seguem um processo que, tipicamente, é composto por quatro etapas, sendo a primeira delas a coleta dos dados, seguida pelo processamento, análise e por fim a etapa de documentação de vestígios em relatórios específicos.

As principais ferramentas em uso atualmente para extração e processamento de vestígios operam sobre dados em memória não-volátil (*dead analysis*) [2], como o FTK¹, Encase², Autopsy³ e IPED⁴. Embora muito úteis para *dead analysis*, elas não oferecem oportunidades de obtenção de vestígios com características temporárias, como conexões de rede ativas, processos em execução ou mesmo senhas de aplicativos que ficam disponíveis em memória volátil, em alguns casos, por um curto período de tempo. Explorar esses vestígios em memória volátil (*live analysis*) é uma demanda importante na área pericial, pois permite, por exemplo, analisar dispositivos de memória permanente protegidas por chaves criptográficas eventualmente mantidas em memória.

Uma maneira de realizar *live analysis* é utilizar ferramentas multiplataforma especializadas, como o Volatility⁵, que permite extrair artefatos diversos a partir de imagens de dispositivos de memória RAM [3]. Essa ferramenta modular consegue fazer extração de dados em formato semi-estruturado, mas possui alto grau de complexidade de uso e baixa integração com outras ferramentas de perícia forense de uso mais extenso como o IPED.

Nesse sentido, incorporar a capacidade de realizar *live analysis* a uma ferramenta de perícia mais robusta e geral como o IPED, permite simplificar a etapa de processamento

das evidências e possibilita a criação de um *pipeline* que também contempla a obtenção de vestígios dinâmicos, resultando em um avanço relevante na área da computação forense.

A. Proposta

Com o intuito de incorporar a análise de memória volátil ao IPED, este trabalho apresenta o sistema Horizon na forma de uma extensão (*Parser*) do IPED para permitir o reconhecimento de *dumps* de memória por meio de uma integração com Volatility 3 usando um conjunto de *plugins*.

B. Contribuições

Em essência, esse estudo apresenta o (i) sistema Horizon como um (*Parser*) para o IPED; (ii) um conjunto de 25 *plugins* que permite realizar *live analysis* usando o IPED; (iii) e os resultados de validação da solução utilizando imagens de teste para os sistemas Windows, Linux e MacOS. Essa solução propõe facilitar o trabalho de peritos forenses, reduzindo erros e agilizando a etapa de processamento e análise de evidências digitais em memórias voláteis.

II. METODOLOGIA

A construção do sistema Horizon envolveu a modelagem e desenvolvimento de *plugins* para adicionar a capacidade de realizar *live analysis* disponível no Volatility 3 ao fluxo clássico de processamento de vestígios digitais existente no IPED. Composto por um *Parser* e um conjunto de *Plugins*, o Horizon funciona como uma ponte entre o IPED e o Volatility e atua na etapa de processamento, conforme ilustrado na Fig. 1.

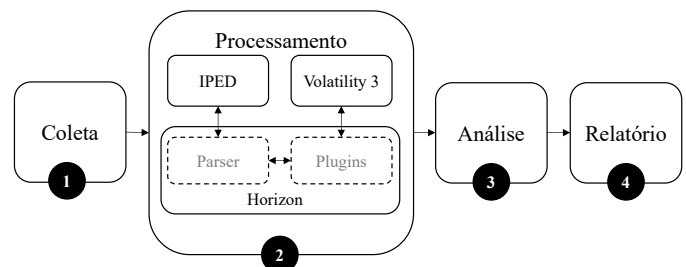


Figura 1. Etapas do processo pericial usando o Horizon.

Cada *Plugin* do Horizon interage com um módulo específico do Volatility e faz a tradução dos dados para apresentação no IPED usando o *Parser* de comunicação. O resultado final do processamento é organizado e visualizado diretamente na interface gráfica do IPED.

¹<https://www.exterro.com/ftk-imager>

²<https://www.opentext.com/products/encase-forensic>

³<https://www.autopsy.com/>

⁴<https://github.com/sepinf-inc/IPED>

⁵<https://github.com/volatilityfoundation/volatility3>

Os experimentos foram conduzidos em ambiente controlado usando computador com Windows 10, processador AMD Ryzen 7 Eight-Core de 3.70 GHz, 16GB de RAM e disco SSD 450GB. Foram processadas três imagens de sistemas operacionais diferentes (Windows, Linux e MacOS) disponibilizadas para estudo⁶, usando um total de 25 *plugins* de *live analysis*, sendo 11 para sistemas Windows, 7 para Linux e 7 para MacOS. Os resultados foram avaliados por meio de comparação com os dados contidos nas imagens de teste.

III. RESULTADOS

Trabalhos anteriores [4], [5] apresentaram os desafios da perícia forense, dentre eles o processamento e visualização de vestígios em memória volátil. Em nosso estudo foi desenvolvida a ferramenta Horizon, que facilita o processamento e visualização de vários tipos de evidências presentes na memória RAM. Os *plugins* que processam essas informações estão disponíveis online no repositório⁷ do projeto, assim como uma wiki com um vídeo demonstrativo e arquivos de desenvolvimento. Os plugins são resumidos por categoria na Tabela I.

Tabela I
PLUGINS HORIZON PARA LIVE ANALYSIS.

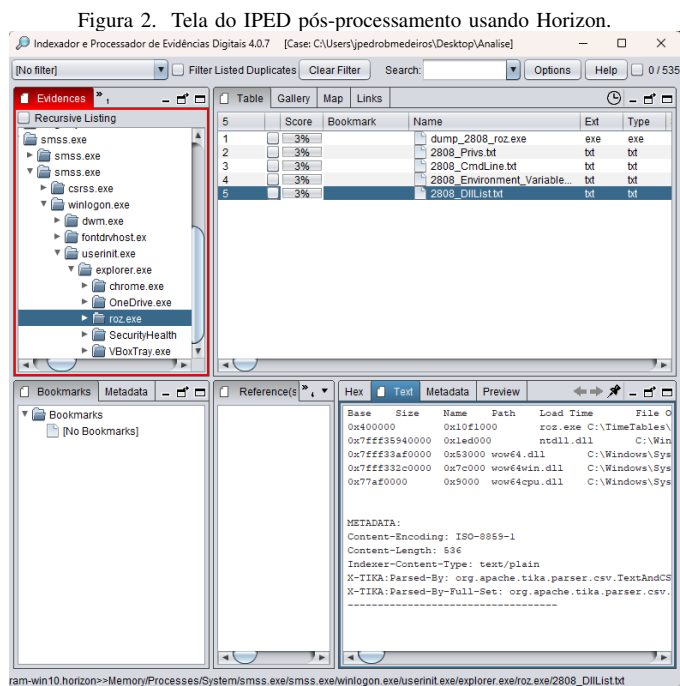
	Windows	Linux	MacOS
Processos	1	1	1
Terminal de comandos	1	2	0
Módulos/Bibliotecas	1	1	1
Pontos de montagem	1	0	1
Informações de rede	1	0	1
Outros	6	3	3

Essa categorização não é rígida e alguns plugins extraem mais de um tipo de informação, como o *LsofFilesExtractor* que trás várias informações como arquivos abertos, processos e conexões de rede. Foram processadas três imagens de testes, sendo uma de cada sistema operacional suportado. A imagem Windows com 2GB, a Linux com 2GB e a MacOS com 1GB. Os tempos de processamento foram, respectivamente, de 3m54s, 7m24s e 1m41s. A Fig. 2 ilustra o resultado do processamento da imagem Windows na tela do sistema IPED, destacando os dados do programa roz.exe.

Foram realizadas análises comparativas entre os dados brutos extraídos pelo Volatility com os dados importados e exibidos no IPED. De modo geral, os elementos foram corretamente extraídos pelo Horizon via Volatility e apresentadas na interface gráfica do IPED com baixo esforço de configuração. Embora essa análise tenha sido feita por amostragem e realizada pelo perito, não foram identificadas divergências nas informações. O maior diferencial do Horizon, no entanto, está na facilidade de processar as evidências e visualizá-las organizadas diretamente no IPED. Essa característica se converte em mais agilidade e assertividade no processamento e análise das evidências, métricas que fazem a diferença na atividade investigativa.

⁶<https://drive.google.com/drive/folders/1tBx-o5MydYWb0rlbnNxb7xnKDKp6JDvh?usp=sharing>

⁷Repositório: <https://github.com/j4xtr1x/Horizon>



IV. CONCLUSÕES

Estudos anteriores trouxeram os desafios da perícia forense computacional e realçaram a importância de ferramentas robustas como o IPED para o processo pericial.

A possibilidade processar e indexar mais rapidamente os dados de memória volátil traz benefícios substanciais ao processo forense. Neste trabalho mostramos que o Horizon adiciona essa facilidade ao IPED por meio de plugins que se comunicam com o Volatility 3. Esta solução supre algumas das dificuldades apresentadas por ambas as ferramentas, integrando-as e oferecendo uma maneira automatizada de processamento de vestígios em memória RAM com uma arquitetura customizável e extensível.

Embora o Horizon represente um avanço relevante, ainda há bastante espaço para melhorias nessa área em futuras investigações, como a (i) redução do nível de acoplamento entre IPED-Horizon; (ii) automatização da identificação dos arquivos de memória e checagem de suas respectivas assinaturas em bases de conhecimento existentes (iii); (iv) e implementação de mecanismos de pré-análise.

REFERÊNCIAS

- [1] A. Case and G. G. Richard, "Memory forensics: The path forward," *Digital Investigation*, vol. 20, pp. 23–33, 2017, special Issue on Volatile Memory Analysis. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1742287616301529>
- [2] M. Kolhe and P. Ahirao, "Live vs dead computer forensic image acquisition," *International Journal of Computer Science and Information Technologies*, vol. 8, pp. 455–457, 2017.
- [3] V. 3, "Volatility 3: The volatile memory extraction framework." Webpage, 2019. [Online]. Available: <https://github.com/volatilityfoundation/volatility3>
- [4] F. Casino, T. K. Dasaklis, G. P. Spathoulas, M. Anagnostopoulos, A. Ghosal, I. Borocz, A. Solanas, M. Conti, and C. Patsakis, "Research trends, challenges, and emerging topics in digital forensics: A review of reviews," *IEEE Access*, vol. 10, pp. 25 464–25 493, 2022.
- [5] IPED, "Iped digital forensic tool," Webpage, 2019. [Online]. Available: <https://github.com/sepinf-inc/IPED>

Segurança Cibernética em Redes 5G *Advanced*

Luiz Gonzaga de Queiroz Silveira Júnior
Departamento de Engenharia de Comunicações
Universidade Federal do Rio Grande do Norte
Campus de Lagoa Nova, Natal, Rio Grande do Norte
luiz.silveira@ufrn.br

Resumo—As redes 5G oferecem maior largura de banda, capacidade para conectar milhões de dispositivos e latência reduzida. No caso do 5G *Advanced*, a promoção de uma rede orientada para IA aumenta os riscos das ameaças à segurança cibernética. Este artigo discute a segurança cibernética das redes 5G *Advanced* e apresenta diversos exemplos de como a inteligência artificial pode ser empregada como ferramenta para o desenvolvimento de uma rede 5G *Advanced* segura.

Index Terms—5G *Advanced*, Segurança Cibernética, IA, ZTA.

I. INTRODUÇÃO

Com o congelamento da *Release* 18 pela 3GPP, em março de 2024, a marca “5G *Advanced*” passa a ser massivamente empregada por toda a indústria de equipamentos de redes sem fio para destacar as vantagens superiores das inovações tecnológicas introduzidas pela 3GPP após anos de discussões e muito trabalho [1].

As principais vantagens do 5G *Advanced* sobre o 5G são evidentes: maior eficiência dos recursos existentes para proporcionar redução do consumo de energia, expansão para novos casos de uso, como dispositivos *RedCap*¹ de baixo custo e rede orientada para ML/IA (ML, sigla inglesa para *Machine Learning*; e IA, sigla para Inteligência Artificial) para proporcionar uso mais flexível e eficiente do espectro, avanços no suporte de diversos dispositivos e topologia de rede em evolução para facilitar diferentes implantações [2]–[4].

Além das expectativas crescerem impulsionadas por um *Massive MIMO* aprimorado e por uma eficiência energética maior, a *Release* 18 da 3GPP fornece também indícios do que o 6G será, ao enfatizar o papel crescente da IA na segurança de redes 5G e além [5].

A próxima Seção discute como a IA pode contribuir para o aperfeiçoamento da segurança cibernética em redes 5G *Advanced*.

II. SEGURANÇA CIBERNÉTICA ORIENTADA POR ML/IA

Enquanto ferramenta, a IA pode ser empregada sob 2 vieses: (a) em ataques cibernéticos sofisticados e (b) na instanciação do Princípio de “Segurança por *Design*”, ao incorporar a IA nos mecanismos de segurança. Essa Seção discute ambas as vertentes e fornece exemplos.

¹A *Release* 17 da 3GPP introduziu 5G NR-Light, ou *Reduced Capability* (RedCap). A *Release* 18 introduziu dispositivos RedCap de nível inferior, buscando encontrar uma *trade-off* entre custo, desempenho e consumo de energia.

A. Contribuições inteligentes para o aperfeiçoamento da segurança cibernética das redes

A IA pode contribuir para melhorar a segurança cibernética de redes 5G de várias formas, destacando-se:

- 1) Detecção de ataques, mediante o aprendizado de padrões de comportamento legítimo através de grandes bancos de dados das operadoras, em que a IA classifica as anomalias como indícios de ataques, de acordo com padrões de interesse, como ataques de negação de serviços (DDoS) [6];
- 2) Previsão e Mitigação, cujo objetivo é usar uma IA para prever ataques com base na análise dinâmica do comportamento das tendências em termos dos padrões de predição, o que favorece o desenvolvimento de soluções sob demanda, na forma de um conjunto de contramedidas eficazes para aperfeiçoar o nível de segurança cibernética da rede [7];
- 3) Segurança aprimorada da camada física, quando a IA é responsável por gerenciar os padrões de *beamforming* nos sistemas com *Massive MIMO* visando otimizar o aproveitamento dos recursos de rádio, mas modifica o feixe do sinal radioelétrico do enlace para mitigar o risco de interceptação das comunicações e violação da privacidade, imediatamente após a detecção de uma invasão [8];
- 4) Gerenciamento de identidade e de acesso, que é realizada após avaliação de informações de contexto por uma IA, como localização do dispositivo e comportamento da rede alvo, para a tomada de decisão no provimento do acesso à rede por um dispositivo, cuja automatização é crucial, considerando-se a necessidade de autenticação e autorização de milhões de dispositivos IoT que pedem acesso à rede [9];
- 5) Automação de Segurança, que ocorre quando a IA automatiza tarefas críticas, assumindo o gerenciamento de tarefas anteriormente delegadas para os especialistas em segurança, que vão desde o provisionamento de políticas de segurança e o gerenciamento de vulnerabilidade, à definição das contramedidas que serão empregadas como respostas para os incidentes de segurança cibernética, incluindo-se, também, as ameaças de dia zero [10];
- 6) Computação de Borda Segura, quando a IA implementa mecanismos de detecção automática de anomalias e políticas de controle de acesso diretamente sobre a

borda da rede, para fins de redução da latência, o que é desejável à medida que a computação na borda se torna cada vez mais predominante na definição de uma arquitetura de serviços [11].

B. Ameaças inteligentes à segurança cibernética das redes

Entidades mal-intencionadas podem empregar uma IA visando ameaçar a segurança cibernética das redes ao promover [12]:

- 1) Ataques de *Spoofing* de Identidade, quando a IA aprende o comportamento e passa a imitar usuários legítimos sob alvo, como administradores da própria rede, com o objetivo de acessar seus dados confidenciais, garantindo o acesso de alto nível para pessoas mal-intencionadas, que podem aproveitar desse acesso para contornar/desligar todas os sistemas de defesa e demais proteções que constituem a arquitetura de segurança cibernética da própria rede;
- 2) Desenvolvimento de *Malware* de Próxima Geração, que podem ser *spywares*, *botnets* e *exploits*, dentre outros, os quais são desenvolvidos de forma automática pela IA, com os objetivos de (i) espionar o ambiente (*spywares*), (ii) selecionar alvos (*botnets*) e (iii) escolher a forma de ataque mais eficiente para contornar as medidas de segurança existentes (*exploits*), como *firewalls*, sistemas de gerenciamento de acesso, sistemas de detecção de intrusão, sistemas de monitoramento de ameaças, dentre outros;
- 3) Ataques de Reconhecimento, nos quais a IA é empregada no reconhecimento de uma rede alvo, identificando dispositivos, serviços, sistemas operacionais, portas abertas, usuários com privilégio de administrador de rede e o mapeamento de todas as vulnerabilidades. Todas essas informações podem servir para a IA explorar essas oportunidades para desencadear ataques de larga escala, mediante a orquestração, isto é, a configuração, o gerenciamento e a coordenação automatizada de ataques muito sofisticados;
- 4) Detecção e Exploração de Vulnerabilidades *Zero-Day*, nos quais a IA explora tais vulnerabilidades para promover ataques DDoS em dispositivos de IoT, cuja escala é capaz de produzir um tráfego colossal que irá consumir rapidamente todos os recursos da rede, causando a sobrecarga. Outro exemplo é um ataque de *botnet* baseado em IA, que pode identificar e explorar automaticamente as vulnerabilidades de dia zero.

III. “NUNCA CONFIE, SEMPRE VERIFIQUE”

A ZTA (ZTA, sigla inglesa para *Zero Trust Architecture*) é uma resposta ao emprego da IA como ferramenta de promoção de ataques cibernéticos sofisticados. Tal definição parte do pressuposto de que entidades mal-intencionadas já possuem acesso à rede e já estão usando desse acesso para alcançar seus objetivos [9].

Desta forma, a adoção da ZTA promove a máxima “nunca confie, sempre verifique” em relação aos dispositivos, usuários,

requisições e solicitações, independentemente do nível de acesso, da localização física do solicitante, do nível de complexidade do dispositivo e do mecanismo de autenticação empregado [13].

Dessa forma, ao promover uma postura proativa, combinada com modelos gerais de implantação e casos de uso da ZTA [13], é possível aperfeiçoar o emprego da IA como uma ferramenta mais eficaz para garantir a segurança cibernética das redes 5G *Advanced* e além [9].

IV. CONCLUSÕES

Nesse artigo, o emprego da IA no contexto da segurança cibernética de redes 5G *Advanced* foi discutido e o princípio da ZTA apresentado. A partir da discussão dos exemplos desse emprego, é possível concluir que a ZTA pode contribuir para aplicações novas da IA visando o desenvolvimento das redes 5G *Advanced* mais resilientes e seguras.

REFERÊNCIAS

- [1] W. Chen, J. Montojo, J. Lee, M. Shafi and Y. Kim, “The Standardization of 5G-Advanced in 3GPP,” in IEEE Communications Magazine, vol. 60, no. 11, pp. 98-104, November 2022, doi: 10.1109/MCOM.005.2200074.
- [2] X. Lin, “An Overview of 5G Advanced Evolution in 3GPP Release 18,” in IEEE Communications Standards Magazine, vol. 6, no. 3, pp. 77-83, September 2022, doi: 10.1109/MCOMSTD.0001.2200001.
- [3] W. Chen et al., “5G-Advanced Toward 6G: Past, Present, and Future,” in IEEE Journal on Selected Areas in Communications, vol. 41, no. 6, pp. 1592-1619, June 2023, doi: 10.1109/JSAC.2023.3274037.
- [4] O. Liberg et al., “Introducing 5G Advanced,” in IEEE Communications Standards Magazine, vol. 8, no. 1, pp. 52-57, March 2024, doi: 10.1109/MCOMSTD.0003.2200059.
- [5] X. Lin, L. Kundu, C. Dick and S. Velayutham, “Embracing AI in 5G-Advanced Toward 6G: A Joint 3GPP and O-RAN Perspective,” in IEEE Communications Standards Magazine, vol. 7, no. 4, pp. 76-83, December 2023, doi: 10.1109/MCOMSTD.0005.2200070.
- [6] B. S. Rawal, S. Patel and M. Sathiyarayanan, “Identifying DDoS Attack using Split-Machine Learning System in 5G and Beyond Networks,” in IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), New York, NY, USA, 2022, pp. 1-6, doi: 10.1109/INFOCOMWKSHPS54753.2022.9798301..
- [7] F. Salahdine, T. Han, N. Zhang, “Security in 5G and Beyond Recent Advances and Future Challenges,” in Security and Privacy. 2023; 6(1):e271. doi:10.1002/spy2.271.
- [8] M. S. J. Solaija, H. Salman and H. Arslan, “Towards a Unified Framework for Physical Layer Security in 5G and Beyond Networks,” in IEEE Open Journal of Vehicular Technology, vol. 3, pp. 321-343, 2022, doi: 10.1109/OJVT.2022.3183218.
- [9] P. Scalise, M. Boeding, M. Hempel, H. Sharif, J. Delloiacovo, J. Reed, “A Systematic Survey on 5G and 6G Security Considerations, Challenges, Trends, and Research Areas,” in Future Internet 2024, 16, 67. <https://doi.org/10.3390/fi16030067>.
- [10] 3GPP, “3GPP Highlights Newsletter: Standards for 5G,” in issue 3 (October), 2021.
- [11] J.A. Brito, J.I. Moreno, L.M. Contreras, M. AlvarezCampana, M. Blanco Caamaño, “Programmable Data Plane Applications in 5G and Beyond Architectures: A Systematic Review,” in , Sensors 2023, 23, 6955. <https://doi.org/10.3390/s23156955>.
- [12] C. Benzaïd and T. Taleb, “AI for Beyond 5G Networks: A Cyber-Security Defense or Offense Enabler?,” in IEEE Network, vol. 34, no. 6, pp. 140-147, November/December 2020, doi: 10.1109/MNET.011.2000088.
- [13] S. Rose, O. Borchert , S. Mitchell and S. Connelly, “Zero Trust Architecture,” in Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], <https://doi.org/10.6028/NIST.SP.800-207>, (Data de acesso: 10 de julho de 2024), 2020.

Sobre o Espaço de Chaves de Uma Técnica de Marca D'água Baseada em Quantização Vetorial

Marcos José Canêjo
Escola UNICAP ICAM-TECH
Universidade Católica de Pernambuco
Recife, Brasil
marcos.azevedo@unicap.br

Jair Galvão
Instituto Federal de Pernambuco
Belo Jardim, Brasil
jair.galvao@upe.br

Francisco Madeiro
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brasil
madeiro@poli.br

Resumo—A ocultação de informação tem sido abordada em vários estudos. Quando a informação é ocultada em imagens digitais, isso pode ser feito em três domínios: domínio espacial (espaço original dos pixels da imagem), domínio da frequência (ou domínio transformado, como o domínio da Transformada do Cosseno) e domínio da compressão. No último caso, pode-se citar a ocultação de dados em imagens comprimidas por quantização vetorial. Este artigo apresenta um cenário de fragilidade, referente ao espaço de chaves, de uma técnica de inserção de marca d'água em imagens comprimidas por quantização vetorial (QV).

Index Terms—ocultação de informação, quantização vetorial, marca d'água.

I. INTRODUÇÃO

Com o crescimento do uso da internet, a criação, replicação e transmissão de conteúdo digital tornaram-se bastante comuns [1]. A proteção de direitos autorais, identificação de propriedades de mídia digital e transmissão segura são cada vez mais importantes [2]. Neste contexto, desempenham papel importante os métodos de inserção de marca d'água visando, por exemplo, avaliar integridade ou assegurar direitos autorais [3].

A Quantização Vetorial (QV) [4] é uma técnica de compressão com perdas que utiliza blocos de amostras (no caso de sinais de voz) ou blocos de pixels (no caso de imagens). Seu desempenho depende dos dicionários projetados. Um algoritmo amplamente conhecido para o projeto de dicionários é o Linde-Buzo-Gray (LBG) [5]. Existe uma ampla variedade de técnicas de projetos de dicionários, e.g. [6]–[8].

Este artigo apresenta uma avaliação inédita da robustez da chave utilizada em um sistema de inserção de marca d'água em imagens comprimidas por QV. Aponta-se um cenário de fragilidade da técnica.

II. MARCA D'ÁGUA BASEADA EM QV

A marca d'água é uma técnica que oculta informações (como um logotipo ou identificador de código em uma sequência de bits) com o objetivo, por exemplo, de identificar o autor ou verificar a integridade do conteúdo digital. Ela pode ser classificada de acordo com sua robustez e visibilidade [1].

O método de marca d'água com uso de QV considerado neste trabalho foi introduzido por Wang et al. [9]. Consiste em dividir o dicionário W em dois sub-dicionários, G_0 e G_1 ,

utilizando uma chave $C = c_1, c_2, \dots, c_N$ e $c_i \in \{0, 1\}$, $1 \leq i \leq N$. Os sub-dicionários G_0 e G_1 ocultam os bits 0 e 1, respectivamente. O Algoritmo 1 mostra a sequência de etapas no processo de inserção da marca d'água baseada em QV. No Algoritmo 2 são apresentados os passos para a extração dos bits ocultos da marca d'água.

Algoritmo 1 Inserindo a marca d'água

1. Dividir o dicionário W em dois sub-dicionários, G_0 e G_1 , utilizando a chave C ;
 2. Dividir a imagem original $X = x_1, x_2, \dots, x_T$ em blocos de tamanho K pixels, onde T é o número de vetores;
 3. Para o vetor de entrada x_i e o bit y da marca d'água a ser inserida, $y \in \{0, 1\}$, obter o vetor-código no sub-dicionário G_0 ou G_1 , com a menor distância para o vetor de entrada. Precisamente, se $y = 0$, a busca pelo vetor-código com a menor distância é realizada em G_0 ; se $y = 1$, essa busca é realizada em G_1 .
 4. Obter o vetor marcado x'_i , que é o vetor-código com a menor distância de x_i ;
 5. Repetir os passos 3-4 até que todos os bits da marca d'água sejam inseridos;
 6. Obter a imagem marcada X' composta por todos os vetores de saída $x'_1, x'_2, \dots, x'_T$.
-

Algoritmo 2 Extraíndo a marca d'água

1. Decompor a imagem marcada X' em blocos não sobrepostos de tamanho K pixels;
 2. Para cada vetor x'_i , executar a busca pelo vetor-código mais próximo para obter um vetor-código mais próximo (w_j) no dicionário W ;
 3. O bit da marca d'água oculto é o valor contido na posição correspondente j da chave C ;
 4. Repetir os passos 2-3 para extrair cada bit da marca d'água.
-

III. METODOLOGIA

Foi considerada a técnica introduzida em [9], em que os índices dos vetores-código são usados para ocultar bits da marca d'água por meio de um dicionário previamente

dividido em dois sub-dicionários com base em uma chave. Se uma pessoa mal-intencionada tiver acesso ao dicionário, ela poderia, por exemplo, realizar um ataque de força bruta visando obter a chave e, com isso, a marca d'água. Esta seção do artigo examina a robustez da chave. Esta avaliação não foi realizada no artigo original.

As simulações foram realizadas utilizando quatro imagens de 512×512 pixels com 256 níveis de cinza (8 bpp): Lena, Elaine, Peppers e Man. A imagem Rose, com 128×128 pixels e 1 bit por pixel, foi utilizada como a marca d'água. A Figura 1 mostra a imagem Lena e a imagem Rose usadas nas simulações. Para avaliar o desempenho do ataque é utilizada a taxa de bits corretos (*Bit Correct Ratio* - BCR).

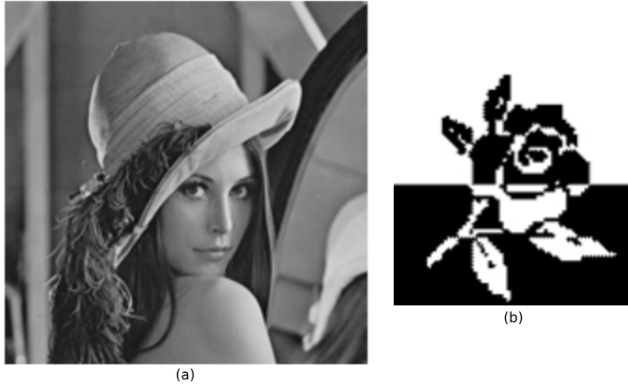


Figura 1. Imagem (a) Lena e (b) Marca D'água.

IV. RESULTADOS

O baixo valor de BCR (Tabela I) correspondente a 90% dos bits da chave alterados pode sugerir, à primeira vista, que a marca d'água extraída é completamente diferente da original. No entanto, a partir de uma análise visual, podemos observar uma semelhança da marca d'água extraída com a original, porém com os bits invertidos: aqueles que originalmente correspondiam a preto agora correspondem a branco e vice-versa, como pode ser visto na Figura 2, que mostra a marca d'água extraída com uma chave com 80%, 90% e 100% dos bits alterados. A Tabela I mostra que BCR diminui com o percentual de bits diferentes entre a chave do atacante e a chave verdadeira.

V. CONCLUSÃO

Este artigo apresenta resultados de ataque de força de bruta no espaço da chaves. Aproximadamente 90% dos bits incorretos são suficientes para exibir uma “marca d'água espelhada”, ou seja, uma marca d'água que corresponde à troca de pixels brancos por pretos e de pretos por brancos na marca d'água original.

REFERÊNCIAS

[1] O. Evsutin, A. Melman, and R. Meshcheryakov, “Digital steganography and watermarking for digital images: A review of current research directions,” *IEEE Access*, vol. 8, pp. 166589-166611, 2020.

Tabela I
VALORES DE BCR PARA ATAQUE DE FORÇA BRUTA NA CHAVE.

% de mudança	BCR (%)			
	Lena	Elaine	Man	Peppers
10 (25 bits)	91,21	89,20	91,28	90,48
20 (51 bits)	79,74	82,06	76,01	77,80
30 (76 bits)	68,48	68,30	68,27	71,95
40 (102 bits)	65,72	58,31	63,82	52,77
50 (128 bits)	46,35	53,24	47,94	54,14
60 (153 bits)	39,96	36,46	41,54	44,97
70 (179 bits)	28,92	29,82	30,28	34,63
80 (204 bits)	21,84	21,79	17,72	20,94
90 (230 bits)	11,70	9,71	12,08	6,70
100 (256 bits)	0	0	0	0

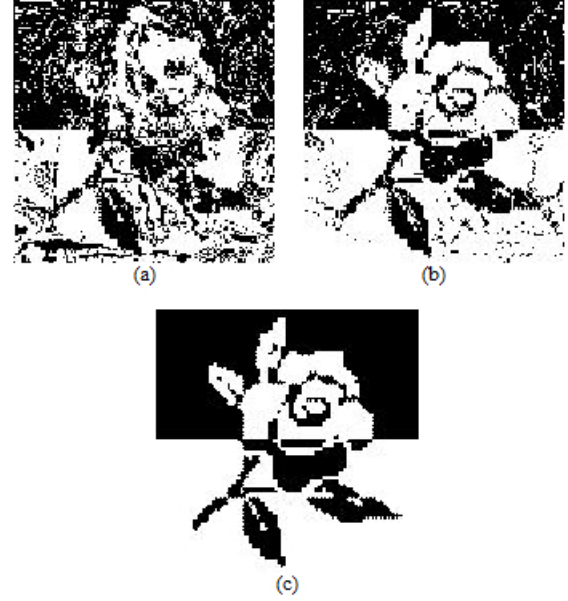


Figura 2. Marca d'água extraída da imagem Peppers de uma chave com (a) 80%, (b) 90% e (c) 100% dos bits alterados.

[2] J. Fei, Z. Xia, B. Tondi, and M. Barni, “Supervised GAN watermarking for intellectual property protection,” in *2022 IEEE International Workshop on Information Forensics and Security (WIFS)*, Shanghai, China, pp. 1-6, 2022.

[3] G.-D. Su, C.-C. Chang, and C.-C. Lin, “Effective self-recovery and tampering localization fragile watermarking for medical images,” *IEEE Access*, vol. 8, pp. 160840-160857, 2020.

[4] A. Gersho and R. M. Gray, “Vector quantization and signal compression,” *Kluwer Academic Publishers*, 1992.

[5] Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design,” *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–85, 1980.

[6] W. Jiang, P. Liu and F. Wen, “An improved vector quantization method using deep neural network,” *International Journal of Electronics and Communications*, vol. 72, pp 178-183, 2017.

[7] C.S. Fonseca, F.A.B.S. Ferreira, F. Madeiro, “Vector quantization codebook design based on Fish School Search algorithm,” *Applied Soft Computing*, vol. 73, pp 958-968, 2018.

[8] V. Severo, F. B. S. Ferreira, R. Spencer, A. Nascimento, and F. Madeiro, “On the Initialization of Swarm Intelligence Algorithms for Vector Quantization Codebook Design,” *Sensors*, vol. 24, no. 8, 2024.

[9] F.-H. Wang, L. C. Jain, and J.-S. Pan, “VQ-based watermarking scheme with genetic codebook partition,” *Journal of Network and Computer Applications*, vol. 30, no. 1, pp. 4–23, 2007.

Alocação do Fator de Espalhamento e Consumo de Bateria em Redes LoRaWAN

Lucas L. de Oliveira, Gabriel M. de Macedo, Vicente A. de Sousa Jr. e Álvaro A. M. de Medeiros

Resumo—A rede LoRaWAN (*Long Range Wide Area Network*) foi projetada para fornecer longo alcance e baixo consumo energético. Por se apoiar na modulação LoRa, o LoRaWAN faz uso de sinais com diferentes fatores de espalhamentos, portanto a escolha impacta na cobertura e no consumo energético da rede. Este artigo analisa o *trade-off* cobertura-consumo energético para redes LoRaWAN e compara o algoritmo de alocação de fator de espalhamento padrão com algoritmos sugeridos na literatura por meio de simulações no *software ns-3*.

Index Terms—LoRaWAN, Bateria, consumo energético, ADR

I. INTRODUÇÃO

Redes LoRaWAN são comumente encontradas em diversas aplicações, tais como *smart farming* [1] e *smart cities* [2]. O padrão LoRaWAN utiliza a camada física LoRa (*Long Range*), que define sinais com seis fatores de espalhamento (SFs) ortogonais [3] entre si, de forma a aumentar cobertura (SF maior) ou taxa de transmissão (SF menor), bem como reduzir a interferência entre os dispositivos de SFs diferentes. O uso de SFs maiores possibilita um maior alcance da transmissão, porém com menor *throughput* e maior tempo para a entrega de pacotes (*air time*) e, consequentemente, maior consumo energético pelos *End Devices* (EDs).

A especificação LoRaWAN determina um método de atribuição de SF para os EDs chamado de *Adaptive Data Rate* (ADR), com base no nível de potência recebida no *Gateway* (GW): quanto maior o nível de potência recebida, menor o valor de SF atribuído [2]. Por conseguinte, obtém-se uma maior taxa de transmissão com o menor consumo energético possível. Entretanto, para redes com muitos EDs, a alocação de vários em um mesmo SF aumenta a possibilidade de interferência e colisão entre pacotes [4].

Grande parte dos trabalhos referentes à alocação de SF focam em cobertura, capacidade, e escalabilidade [1], [4], [5]. No entanto, o consumo de energia, vital no contexto LoRaWAN já que EDs podem ser alimentados por bateria, é pouco explorado [6]. Diante disso, surge a necessidade de analisar o consumo energético dos EDs frente aos algoritmos de alocação de SF, de forma a otimizar o desempenho da rede e prolongar a vida útil das baterias. Este artigo tem o objetivo de analisar, por meio de simulações, o consumo energético de algumas propostas de alocação de SF discutidas em [4].

Lucas L. de Oliveira e Álvaro A. M. de Medeiros são do Programa de Pós Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora, Juiz de Fora-MG, Brasil, e-mail: {lima.lucas, alvaro}@engenharia.ufjf.br; Gabriel M de Macedo e Vicente A. de Sousa Jr. são do Programa de Pós Graduação em Engenharia Elétrica e de Computação da Universidade Federal do Rio Grande do Norte, Natal-RN, Brasil, email: {gabriel.maia.115,vicente.sousa}@ufrn.edu.br. Esse trabalho foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)- Código de Financiamento 001. As simulações fornecidas por esse trabalho foram realizadas com o apoio do Núcleo de Processamento de Alto Desempenho da UFRN (NPAD/UFRN). O Instituto de Pesquisas Eldorado, Manaus-AM, contribuiu com a presença de Gabriel M. de Macedo.

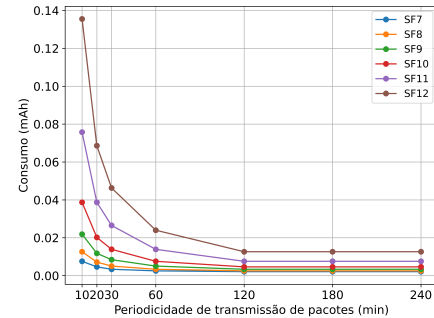


Figura 1. Consumo energético para cada SF.

II. CONSUMO ENERGÉTICO EM REDES LORAWAN

A tecnologia LoRaWAN se tornou o padrão mais utilizado em aplicações de baixo *throughput* que demandam alta cobertura em virtude da sua eficiência energética. A modulação LoRa permite um maior alcance com baixa potência de transmissão sem uso de retransmissões. Os EDs têm a capacidade de operar nas classes A, B e C, sendo a classe A a mais comum, na qual os EDs abrem pequenas janelas de recepção após a transmissão, minimizando o tempo ativo e consumo de energia. Com isso, os EDs não precisam permanecer no estado de transmissão, com corrente de operação de 52 ± 0.216 mA, transitando facilmente para os estados *standby* (10^{-3} mA) ou *sleep* (1.6 ± 0.064 mA) [7].

A alta eficiência energética também é explicada pela simplicidade do ED, sendo o processamento concentrado no *Network Server* (NS) e pelo mecanismo ADR, que ajusta o SF de acordo com a sensibilidade do receptor, garantindo que os EDs consumam o mínimo de energia que viabiliza a comunicação com o GW. Os SFs variam entre 7 e 12 e possuem sensibilidade conforme apresentado na Tabela I de [4]. Por meio de simulações, é possível avaliar o consumo energético para cada SF. A Fig. 1 apresenta o comportamento de um ED, situado a 2 km do GW, com potência de transmissão igual a 14 dBm e largura de banda de 125 kHz, para diferentes periodicidades de transmissão de pacotes. É possível observar a diferença em consumo para SF maiores, principalmente em situações em que as transmissões são mais frequentes.

Para análise do consumo energético do ADR, compara-se o seu desempenho com a transmissão utilizando unicamente o menor e o maior SFs (7 e 12), bem como uma atribuição de SF de forma proporcional a todos SFs, chamada de *Equal Split* [4]. Nessa política de alocação de SFs, os EDs são ordenados em nível de potência recebida e divididos em 6 grupos, de forma que o primeiro grupo é alocado ao SF 7, o segundo ao SF 8, e assim por diante. A simulação contempla apenas um GW no centro de um círculo de raio variável, no qual os EDs

Tabela I
CONFIGURAÇÕES DA SIMULAÇÃO.

Parâmetro	Valor
Tamanho do pacote	20 Bytes
Raios	2,4,6,8,10,12,15 km
Número de EDs	1000
Tráfego	Uplink
Potência de transmissão	14 dBm
Largura de banda	125 kHz
Índices <i>Path Loss</i>	$d_0 = 1m; PL(d_0) = 7.7; n = 3.7$
Periodicidade de envio	600 s

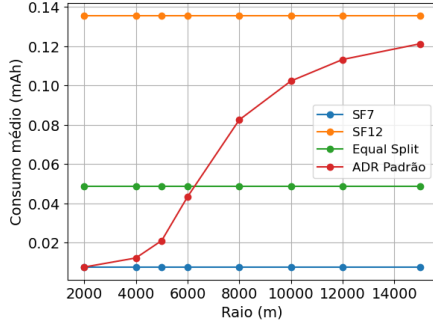


Figura 2. Consumo energético por algoritmo.

são posicionados de forma uniformemente aleatória. Assim, à medida que esse raio aumenta, os EDs ficam mais espalhados na área, reduzindo a densidade de nós, mas aumentando a distância média entre ED e GW. Cada ED envia um pacote em um tempo aleatório dentro de um intervalo de 10 minutos. A Tabela I apresenta os parâmetros da simulação. O modelo de propagação de perda de caminho log-distância foi adotado para focar no *trade-off* cobertura-consumo energético sem aleatoriedade imposta por modelos de canal mais complexos. Os resultados apresentados são considerados para uma rede em estado estacionário e as colisões entre SF são tratadas de forma quase ortogonal, ou seja, podem ocorrer mesmo com diferença de potência recebida entre sinais simultâneos.

III. RESULTADOS

A Fig. 2 apresenta o consumo médio por ED em função do raio de ocupação dos EDs, na qual é possível observar a adaptabilidade do ADR em função do aumento de raio. Visando um maior alcance, SFs maiores são alocados, aumentando o consumo energético. Como os outros algoritmos alocam SFs de forma estática, o consumo médio é sempre o mesmo, tendo os algoritmos que usam apenas o SF 7 e o SF 12 como limitantes inferior e superior, respectivamente. Para raios maiores que 6 km, o algoritmo *Equal Split* apresenta menor consumo que o ADR por manter a alocação em SFs menores, o que deve gerar um aumento na perda de pacotes recebidos.

A Fig. 3 mostra a porcentagem de pacotes pedidos ou *Packet Loss Rate* (PLR), em linha cheia, e a porcentagem de pacotes perdidos por interferência ou *Packet Interference Rate* (PIR), em linha tracejada. Nota-se que a interferência entre EDs de mesmo SF é a principal causa da perda de pacotes até o raio de 8 km. Após esse raio, há uma redução da PIR, indicando uma perda maior por falta de cobertura, pois a PLR aumenta mesmo para o SF 12. Em termos de desempenho, o ADR prevalece em relação as outras propostas, com maior porcentagem de acerto de pacotes para quase todos os raios, exceto para o intervalo

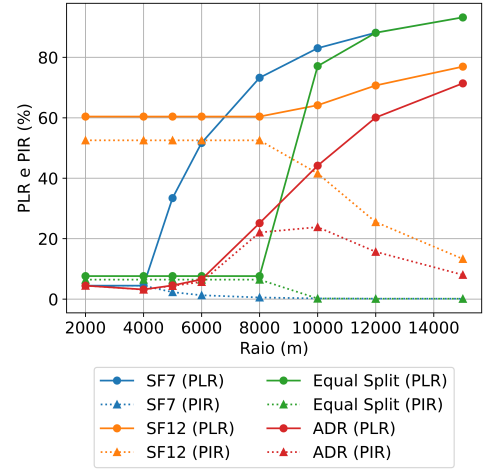


Figura 3. PLR e PIR para cada algoritmo.

entre 6 e 8 km. Nessa região, o *Equal Split* apresenta tanto consumo quanto PLR menores que o ADR.

IV. CONSIDERAÇÕES FINAIS

Neste trabalho, foi avaliado o desempenho do ADR frente ao *tradeoff* cobertura-consumo energético. Ao verificar a PIR na Fig. 3, nota-se que há um aumento da interferência no ADR na região de final de cobertura, enquanto o *Equal Split* mantém o nível de interferência, visto que sua alocação de SF é a mesma para qualquer densidade de nós. Como o ADR é baseado na sensibilidade, o algoritmo tende a atribuir mais EDs ao SF 12 para raios maiores. Com mais nós no mesmo SF, há um aumento da da colisão entre pacotes e, por conseguinte, da interferência. Já o *Equal Split* apresenta menores consumo e interferência nessa região de borda. Isso mostra que há oportunidades para melhorias do ADR para EDs na borda da área de cobertura do GW, e que propostas de novos algoritmos podem superá-lo também em termos de consumo energético.

Para trabalho futuros, novos algoritmos baseados em consumo energético podem maximizar a vida útil de bateria, especialmente para aplicações de difícil carregamento ou troca da bateria. Algoritmos baseados em inteligência artificial podem encontrar padrões que otimizem a rede, em particular para aplicações tolerantes à perda de pacotes.

REFERÊNCIAS

- [1] Gabriel M. de Macedo *et al.*, "LoRaWAN em Pecuária Inteligente: Análise de Desempenho de Algoritmos de ADR," *XLI Simp. Brasileiro de Telecomunicações e Processamento de Sinais (SBrT2023)*, 2023.
- [2] D. Magrin *et al.*, "Performance Evaluation of LoRa Networks in a Smart City Scenario," *IEEE Int. Conf. on Communications (ICC)*, 2017.
- [3] A. Mahmood *et al.*, "Scalability Analysis of a LoRa Network Under Imperfect Orthogonality," *IEEE Trans. on Industrial Informatics*, vol. 15, no. 3, pp. 1425, 2019.
- [4] Luiz Q. R. da C Filho *et al.*, "Spreading Factor Assignment Strategy for Coverage and Capacity Flexible Tradeoff," *Journal of Communication and Information Systems*, vol. 37, pp. 47-51, 2022.
- [5] K. Mikhaylov *et al.*, "Analysis of Capacity and Scalability of the LoRa Low Power Wide Area Network Technology," *22th European Wireless Conference*, 2016.
- [6] L. Casals *et al.*, "Modeling the Energy Performance of LoRaWAN," *Sensors*, vol. 17, no. 10 (2364), 2017.
- [7] J. Finnegan *et al.*, "Modeling the Energy Consumption of LoRaWAN in ns-3 Based on Real World Measurements," *Global Information Infrastructure and Networking Symposium (GIIS)*, 2018.

Determinação Ótima de Spread Factor (SF) em Redes IoT-Lora

Ivan Igor Antunes Neves, Filipe Cavalcanti Fernandes, Caio Mateus Machado Cardoso, Fabrício Barros e Jasmine P. L. Araújo

Abstract — As redes LoRaWAN estão sendo muito utilizadas em aplicações de Internet das Coisas (IoT), a rede fornece comunicação de longo alcance e baixo consumo energético e utiliza a modulação Chirp Spread Spectrum (CSS), que possui um parâmetro ajustável, o Fator de Espalhamento (SF). Diante desse cenário, este artigo realiza uma análise comparativa de otimizadores para redes IoT-Lora visando a determinação do SF ótimo para maximizar a vazão considerando diferentes números de nós sensores. Os resultados obtidos mostram que o otimizador tradicional, solver que resolve problemas de convexidade (CVX), encontra resultados levemente melhores que a metaheurística inspirada no processo natural de polinização de flores (FPA), no entanto a metaheurística é, aproximadamente, 5 vezes mais rápida para determinar o resultado.

Index Terms — IoT-Lora, LoRaWAN, Vazão, Algoritmo Bioinspirado, Solver e Otimização.

I. INTRODUÇÃO

O Long-Range(LoRa) tem se mostrado uma solução eficaz para conectividade de Internet das Coisas(IoT) em cenários que exigem longo alcance e baixo consumo de energia [1]. Essas características o tornam eficiente para aplicações industriais, agrícolas e de segurança. O Lora opera na banda de frequência ISM, utilizando uma modulação de espectro espalhado.

A principal vantagem dessa tecnologia é a capacidade de penetração de obstáculos, como paredes e vegetação, mantendo a confiabilidade em ambientes difíceis, o tornando útil em áreas rurais ou em infraestrutura de difícil acesso, onde outros tipos de comunicações sem fio possam falhar. Além disso, o protocolo Long Range Wide Area Network (LoRaWAN) complementa a tecnologia LoRa disponibilizando abertamente a arquitetura e a camada de rede que gerencia a conectividade entre os dispositivos. Esta rede possui estrutura hierárquica, tendo gateways, sensores, servidores de rede e servidores de aplicação como é mostrado em [2].

Ainda sim, a implementação de redes LoRaWAN contém desafios no gerenciamento de rede e na otimização do uso do espectro. A seleção eficiente do SF é essencial para maximizar a capacidade e a confiabilidade da rede. Várias técnicas de otimização e algoritmos foram apresentados para melhorar o desempenho da rede LoRaWAN. Neste trabalho, é realizada a comparação entre duas técnicas de otimização, uma tradicional e a metaheurística da polinização de flores (FPA).

II. MODELO DE SISTEMA

O LoRaWAN é baseado no protocolo aloha, onde cada nó transmite em um canal de frequência aleatório. A razão dos nós que usam SF específico pertence $\{7, 8, 9, 10, 11 \text{ e } 12\}$ é dita como $p(s)$ e o número de nós cobertos por N_c e a taxa média de chegada (λ).

A carga de tráfego do canal $G(s)$, ou seja, o número médio de pacotes transmitidos por tempo de transmissão no ar $ToA(s)$ no SF, é definido pela Eq. (1):

$$G(s) = \lambda \cdot p(s) \cdot N_c \cdot ToA(s) \quad (1)$$

A taxa de transferência total (S) é definida pela Eq. (2), onde (b) é o número de bits por pacote para cada SF :

$$S = \sum_{s=Smin}^{S=Smax} \lambda \cdot p(s) \cdot N_c \cdot b \cdot \exp(-2G(s)) \quad (2)$$

A função utilidade(R) usada para obter uma taxa de transferência justa e proporcional entre diferentes SFs é definida pela Eq. (3):

$$R = \sum_{s=Smin}^{S=Smax} \log(\lambda \cdot p(s) \cdot N_c \cdot b) - 2 \cdot \sum_{s=Smin}^{S=Smax} G(s) \quad (3)$$

III. SIMULAÇÃO

O modelo de sistema foi proposto em [3] que foi considerado uma rede LoRaWAN com 4 gateways (r) e 4000 nós sensores (N_t) sendo eles os mesmos transceptores (LoRa SX1272) em uma área de 7km usando a largura de banda(BW) de 125kHz, o comprimento de carga útil (k) igual a 40 bytes, o ciclo de trabalho(d) de 1%.

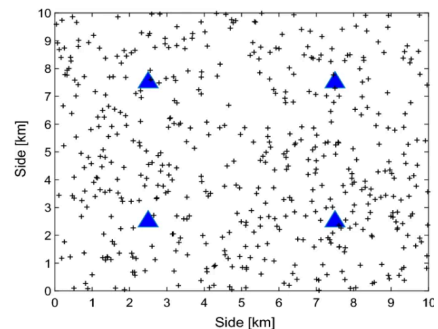


Figura 1: Rede simulado em [3].

A transmissão ocorre de acordo com processo de poisson e o modelo de perda de caminho Okumura-hata assim obtendo λ , ToA e b para cada SF.

	SF	b	ToA	Lambda
1	7	50	0.1048	343.32
2	8	75	0.1802	199.75
3	9	100	0.3211	112.1
4	10	125	0.5636	63.87
5	11	150	1.0485	34.33
6	12	175	1.9398	18.55

figura 2 : Resultados da simulação de [3]

IV. MODELOS DE OTIMIZAÇÃO

O Solver CVX [4] é um conjunto de softwares, integrados no ambiente MATLAB, construídos para resolver problemas de otimização convexas. Esse solver possibilita que os problemas de otimização sejam definidos de forma clara e concisa. Isso inclui problemas como programação linear, programação quadrática, programação cônica de segunda ordem e programação semidefinida.

O algoritmo Flower Pollination (FPA) [5] é um meta-heurística inspirada na natureza desenvolvido por Xin-She Yang em 2012. Essa otimização bio-inspirada imita comportamentos de polinização de flores para encontrar soluções ótimas que se caracterizam em dois tipos: polinização global e polinização local onde é modelada pela polinização cruzada e pela polinização autogamia respectivamente. Isso permite sua capacidade de exploração eficiente.

V. RESULTADOS

Levando em conta Eq.(1) como a função objetivo, $p(s)$ a variável de decisão e $\sum p(s) = I$; $p(s) >= 0$ como restrições, duas técnicas de otimização (CVX e o FPA) foram aplicadas para mostrar a seleção ótima de SF para cada nó em uma rede LoRaWAN. A vazão total da rede foi calculada e a Figura 3 ilustra a vazão ótima para cada otimizador considerando a presença de diferentes números de nós na rede.

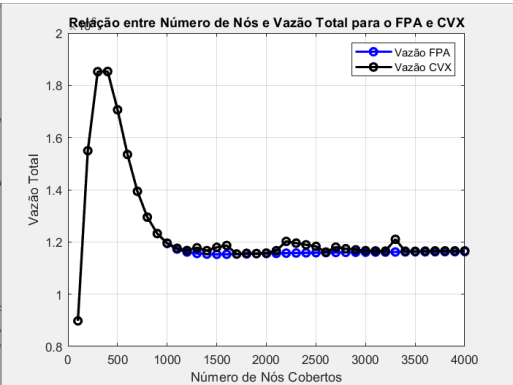


Figura 3: Gráfico de Comparação do CVX e FPA

Nota-se que a vazão para o CVX e o FPA se assemelham até aproximadamente 1100 nós cobertos, depois há uma leve diferença em alguns casos. Percebemos o pico de vazão na faixa dos 300 a 400 nós e também foi notado que o tempo de processamento do FPA foi 5 vezes mais rápido que o CVX.

	FPA	CVX
Tempo de processamento	22.430631 sec.	113.834692 sec.

Especificação da máquina usada	Intel(R) Core(TM) i5-1035G1 CPU @ 1.00GHz 1.19 GHz
--------------------------------	--

VI. CONCLUSÃO

A partir dos resultados entendemos que o solver tem um melhor rendimento de vazão se comparado ao algoritmo bioinspirado a partir de 1100 nós. No entanto, essa diferença não exclui o FPA devido ao tempo de processamento, para determinadas aplicações como: estratégias de posicionamento de veículos aéreos não tripulados (UAV) em redes LpWANs [6].

Observamos também que para trabalhos futuros espera-se determinar a eficiência energética como objetivo da otimização, uma vez que sabemos a quantidade de números de nós cobertos que é necessário para ter a maior vazão.

REFERÊNCIAS

[1] Jayavignesh Thyagarajan, Aaron Alex Abraham, Aakash R et al. Performance Evaluation and Comprehensive Analysis of Lora Network Planning for Iot Deployment Scenarios, 14 May 2024, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-4319621/v1].

[2] L. Alliance, “Lorawan® regional parameters rp002-1.0.4” Tech .Rep., 2022, disponível em: https://resources.lora-alliance.org/technicalspecifications/rp002-1-0-4-regional-parameters (acessado em: 24-05-2024).

[3] Loubany, A., Lahoud, S., Samhat, A.E. et al. Joint throughput-energy optimization in multi-gateway LoRaWAN networks. Telecommun Syst 84, 271–283 (2023). https://doi.org/10.1007/s11235-023-01048-8

[4] M. Grant; S. Boyd.; CVX: Software Matlab para programação convexa disciplinada, versão 2.0 beta. 2013, disponível em: https://cvxr.com/cvx (acessado em: 20/03/2024)

[5] M. -u. -N. Khursheed et al., "Review of Flower Pollination Algorithm: Applications and Variants", 2020 International Conference on Engineering and Emerging Technologies (ICEET), Lahore, Paquistão, 2020, pp. 1-6, doi: 10.1109/ICEET48479.2020.9048215.

[6] Tarab, H. "Real time performance testing of LoRa-LPWAN based environmental monitoring UAV system. University of Windsor, Electronic Theses and Dissertations. 7578 (2018)."

Novas Implementações Tecnológicas no Setor de Segurança do IFRN Campus SGA

Alisson G. D. Ribeiro, Carlos V. G. Cunha, Erick R. M. Balbino, Josué L. S. da Cruz, Lucas L. dos Santos, Samuel V. G. Moraes, Leonardo G. de P. Amorim.

Resumo—O presente projeto tem como finalidade promover a detecção de intrusões em locais estratégicos da instituição por meio da instalação de tecnologias avançadas de controle de acesso, tais como sensores, atuadores e microcontroladores. Com isso, espera-se fornecer um suporte significativo ao setor de segurança (COSGEM) da instituição. Essas tecnologias serão integradas com microcontroladores altamente eficientes, capazes de se conectar à rede Wi-Fi, proporcionando eficiência e comodidade.

Palavras-Chave—IoT, Monitoramento, Segurança, Automação.

I. INTRODUÇÃO

É amplamente conhecido que a criminalidade e a violência representam problemas sociais profundamente enraizados na sociedade brasileira. Como consequência, a população vive em constante medo, com privação do bem-estar social e fortalecimento do sentimento de insegurança. Esse cenário reforça a necessidade de implementar sistemas de segurança em diversos âmbitos sociais, principalmente o educacional [1].

Devido ao rápido avanço tecnológico das últimas décadas, muitos sistemas de segurança se tornaram ineficientes. Tecnologias convencionais, como câmeras de segurança e alarmes básicos, não garantem um monitoramento eficaz. No Campus SGA, pesquisas conduzidas com membros do Setor de Segurança de Segurança (COSGEM) apresentaram dificuldades no monitoramento de locais de difícil acesso e visibilidade, especialmente no período noturno.

Para abordar essas questões, este artigo apresenta a implementação de sistemas de segurança eficientes que contam com a tecnologia Internet das Coisas (IoT) [2], como: Sensores infravermelhos, microcontroladores e atuadores. Os sensores infravermelhos são capazes de detectar intrusões independentemente da visibilidade do local. Em caso de detecção, os sensores transmitem para os microcontroladores, que processam os dados e acionam os atuadores rapidamente.

A implementação destes sistemas não visa apenas aumentar a eficiência do monitoramento, mas também reduzir o tempo de resposta a intrusões, proporcionando melhorias significativas na segurança da instituição [1].

Alisson G. D. Ribeiro, Carlos V. G. Cunha, Erick R. M. Balbino, Josué L. S. da Cruz, Lucas L. dos Santos, Samuel V. G. Moraes, Leonardo G. de P. Amorim, Grupo de Estudos Aplicados à Tecnologia da Informação e Comunicação (GEATIC). Instituto Federal do Rio Grande do Norte, São Gonçalo do Amarante-RN, Brasil. E-mails: alisson.gabriel@escolar.ifrn.edu.br, cunha.gomes@escolar.ifrn.edu.br, e.renne@escolar.ifrn.edu.br, josue.luis@escolar.ifrn.edu.br, lourenco.s@escolar.ifrn.edu.br, victor.samuel@escolar.ifrn.edu.br, leonardo.amorim@escolar.ifrn.edu.br

II. METODOLOGIA

Adotou-se, para a realização deste estudo, uma metodologia de pesquisa experimental, que consiste [4] em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto. Os sistemas de segurança, como dito anteriormente, fazem uso de dispositivos IoT [3]. Utilizou-se o software Tasmota, um firmware open-source para dispositivos baseados no chip de microcontroladores da série ESP8266. Esses dispositivos se conectam ao servidor Home Assistant, software open-source, por meio de sua integração com o Tasmota via protocolo MQTT. O Home Assistant é utilizado para controlar e automatizar os dispositivos. O artigo apresenta dois tipos de sistemas de segurança: um para uso interno e outro para externo.

A. Sistema de Segurança Interno

Dois módulos foram desenvolvidos compostos por um sensor infravermelho passivo (PIR) de modelo D-SUN HC-SR501, que detecta movimento por variação de calor emitido por corpos, e o Sonoff Mini R2, um relé WiFi baseado no chip ESP8266. Inicialmente, o sinal de entrada do Sonoff Mini R2 estava configurado como ativo baixo. Foi realizada uma modificação de forma que o dispositivo passasse a ser ativo alto. Dessa forma, o relé WiFi e sensor PIR puderam operar juntos. Os dispositivos foram conectados por três condutores: alimentação, sinal de saída e ground. Após isso, os dois módulos foram instalados na Sala de Aula 08 para acompanhamento dos resultados.

No Home Assistant, foram criadas duas automações para esses módulos. A primeira automação aciona um contador regressivo de 10 minutos quando qualquer um dos módulos detecta movimento e registra no servidor que há presença no local. A segunda desativa o registro de presença no servidor quando o contador chega a zero, indicando que não houve detecção de movimento no período definido.

B. Sistema de Segurança Externo

Foi desenvolvido uma nova Central de Alarmes para o sistema de segurança dos portões de acesso ao campus, que é composto por sensores de barreira infravermelhos ativos modelo IRA-315, que detectam presença quando o duplo feixe é interrompido, e sirene. A central, composta por um gabinete, microcontrolador ESP 32, duas chaves gangorras e quatro LEDs, monitora dois acessos: zona 1 (portão principal) e zona 2 (portão lateral).

No gabinete, foram instaladas na parte de baixo as chaves e dois leds e, na tampa, os leds restantes. As chaves têm a função de ativar/desativar o monitoramento das duas zonas, com os leds indicando o status de ativação. Os leds da tampa indicam a detecção de intrusão nos dois acessos, sendo um para a zona 1 e o outro para a zona 2.

Os sensores, as chaves, os leds e a sirene são conectados ao ESP 32, que executa as automações desenvolvidas no firmware Tasmota. Duas regras de automação foram criadas:

- Se os sensores detectarem presença e o monitoramento estiver ativo (chave ativa), a sirene é acionada e inicia-se um contador regressivo.
- Se o monitoramento estiver desativado (chaves desativadas) ou o contador chegar a zero, a sirene é desligada.

III. RESULTADOS

Esta seção é destinada à discussão dos resultados obtidos da metodologia experimental apresentada na seção anterior.

A. Sistema de Segurança Interno

Após os módulos terem sido finalizados, vistos nas figuras 1 e 2, ambos foram instalados na Sala de Aula 08, o primeiro na parte da frente e o segundo na de trás. Na figura 3, é visto o registro de monitoramento dos módulos na interface do Home Assistant. O primeiro gráfico demonstra a detecção do primeiro módulo, o segundo mostra a detecção do segundo módulo e o terceiro apresenta o tempo total que o servidor registrou presença no local.



Figura 1. Primeiro Módulo Finalizado



Figura 2. Segundo Módulo Finalizado

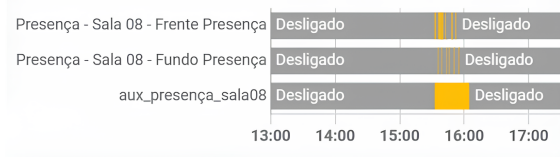


Figura 3. Registro de Monitoramento

A análise do primeiro e segundo gráfico mostram a sinergia entre os módulos. Nos testes foi observado que ambos os módulos possuem um bom tempo de resposta ao movimento. Conclui-se que os dois dispositivos são capazes de monitorar uniformemente o ambiente.

B. Sistema de Segurança Externa

Na figura 4, é vista a central de alarmes finalizada. Em seguida, iniciaram-se os testes com as automações desenvolvidas por meio das Rules do firmware Tasmota. É válido ressaltar que até o momento da escrita, os testes foram limitados à Zona 1.

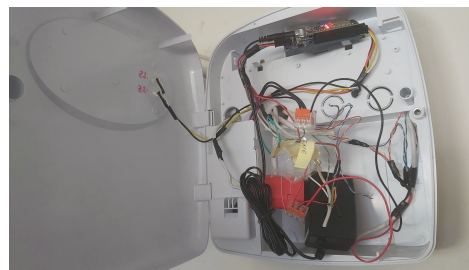


Figura 4. Central de Alarmes Finalizada

A primeira automação funciona da seguinte forma: se o monitoramento estiver ativo e os sensores detectarem intrusão, a sirene é acionada por tempo definido. A segunda automação desativa a sirene quando o período de tempo chega a zero ou quando ambos os monitoramentos são desativados (chaves desativadas). Durante os testes, as duas regras de automação funcionaram conforme o esperado.

IV. CONCLUSÃO

Com base nos testes realizados, observou-se que ambos os sistemas demonstraram grande potencial de contribuir significativamente para a segurança do ambiente escolar. Além disso, a combinação dos monitoramentos internos e externos oferece uma cobertura abrangente de proteção ao espaço educacional.

REFERÊNCIAS

- [1] E. R. Cordeiro, "Gestão da segurança das instituições federais de ensino superior do Brasil," Anais do 14º Colóquio Internacional de Gestão Universitária, Florianópolis, Brasil, 2014, vol. 1, no. 1, pp. 1-14.
- [2] R. A. Mouha, "Internet of things (IoT)," Journal of Data Analysis and Information Processing, vol. 9, no. 2, pp. 77, May 2021.
- [3] S. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of Things (IoT): A literature review," Journal of Computer and Communications, vol. 3, no. 5, pp. 164-173, May 2015.
- [4] A. C. Gil, Como elaborar projetos de pesquisa, 4rd ed.. São Paulo, Brasil: Atlas, 2002.

Um Hub e Gateway MQTT de baixo custo para Sensores IoT com interface USB

Cláudio Rodrigues Muniz da Silva
Departamento de Engenharia de
Comunicações (DCO)
Universidade Federal do Rio Grande do
Norte (UFRN)
Natal, Brasil
claudio.rmsilva@gmail.com

João Paulo dos Santos Nascimento
Curso de Engenharia de
Telecomunicações (CETEL)
Universidade Federal do Rio Grande do
Norte (UFRN)
Natal, Brasil
joao.paulonsascimento.047@ufrn.edu.br

Felipe Augusto Cortez M. da Silva
Laboratório de Convergência de Redes
Universidade Federal do Rio Grande do
Norte (UFRN)
Natal, Brasil
felipe1197@hotmail.com

Abstract—It is undeniable that Internet of Things has revolutionized the way control systems can be developed and used. An IoT control system is a technological innovation that allows the transformation of common objects into smart devices capable of collecting and exchanging data. Sensors play an essential role in this context, providing the information necessary to make informed and automatic decisions. However, integrating sensors, especially those with a USB interface, into IoT solutions presents significant challenges related to connectivity, limited range and interoperability issues. This paper proposes the development of a low-cost MQTT hub and gateway for IoT sensors with USB interface. The methodology adopted for the development of the gateway includes selection of suitable hardware, development of MQTT data reading and publishing software, and performance and reliability tests. Temperature and humidity sensors were chosen as proof-of-concept devices due to their wide use in IoT applications. The results are promising and include validation of the gateway's effectiveness in terms of latency, data transmission rates and resource consumption within a real scenario created in the laboratory.

Keywords—Internet of things, sensors, gateways, usb interface

I. INTRODUÇÃO

A Internet das Coisas (IoT) tem possibilitado o desenvolvimento de soluções eficientes que permitem a interconexão de dispositivos e a coleta contínua de dados. Uma solução IoT envolve diversos componentes que trabalham juntos para coletar esses dados, processá-los e fornecer ações dentro de uma arquitetura típica para os sistemas IoT. Entre esses componentes estão: os sensores, os gateways, as plataformas e as aplicações de acesso IoT [1].

Os sensores desempenham um papel essencial nesse contexto, fornecendo as informações necessárias para a tomada de decisões informadas e automáticas. Eles utilizam uma variedade de interfaces de acesso para se comunicar com os microcontroladores.

Os sensores com interface USB possuem diversas vantagens em comparação com os sensores que usam outras formas de interfaceamento: a) facilidade de uso (os sensores USB são dispositivos plug-and-play); b) facilidade de alimentação; c) altas taxas de transferência de dados; e d) compatibilidade ampla. Apesar destas vantagens, a integração deste tipo de sensor, em soluções IoT apresenta desafios relacionados à conectividade externa, alcance limitado e interoperabilidade [2][3].

Uma abordagem que tem sido muito utilizada para esses desafios tem sido a utilização de Hubs de sensores USB, principalmente os que possuem também a função de gateway para algum tipo de protocolo usado nas soluções IoT, tal como o protocolo Message Queuing Telemetry Transport (MQTT)

[4][5]. Um hub de sensores IoT é um dispositivo que atua como um ponto central de coleta e distribuição de dados provenientes de vários sensores IoT. Ele serve como um intermediário entre os sensores e outros componentes da rede, como servidores, gateways ou sistemas de nuvem [6].

O protocolo MQTT tem se revelado uma solução eficiente para comunicação em redes IoT devido à sua simplicidade, leveza e suporte para arquiteturas de publicação/subscrição [7]. Ele é muito indicado para ambientes de baixa largura de banda e alta latência necessárias em muitas aplicações IoT [8].

A utilização de gateways MQTT pode simplificar sua integração em sistemas IoT, oferecendo um meio de comunicação padronizado e flexível [9][10]. Alguns fabricantes de soluções para sistemas IoT oferecem comercialmente soluções de hub com funcionalidade de gateway para MQTT, mas possuem custos relativamente altos o que dificulta o acesso a esse tipo de dispositivo.

Este artigo propõe o desenvolvimento de um protótipo de hub e gateway MQTT de baixo custo para sensores IoT com interface USB. O objetivo é fornecer uma solução acessível e eficiente para a coleta e transmissão de dados de sensores USB em ambientes IoT.

A estrutura do artigo está organizada da seguinte forma: na Seção II, detalha a metodologia para o desenvolvimento do gateway. A Seção III apresenta os resultados dos testes de desempenho e estudos de caso, bem como discute as implicações para futuras pesquisas. Por fim, a Seção IV oferece conclusões e sugestões para trabalhos futuros.

II. METODOLOGIA

A metodologia adotada é a prototipagem com as seguintes etapas: a) definição dos cenários de uso e seu requisitos; b) definição do hardware que será usado; c) desenvolvimento do software de controle do hub/gateway; e d) definição de testes e validação do protótipo.

Dois cenários de uso foram idealizados para a utilização do protótipo criado (Ver fig. 1): a) Acesso ao protótipo via Dashboard Web e b) Acesso às informações do protótipo e seus sensores através da intermediação de um Broker MQTT. Para o primeiro cenário, o hub deve ser capaz de exibir os dados dos sensores em tempo real através de um dashboard acessado pela web. Para tanto, as seguintes atividades foram executadas: a) Desenvolvimento do Dashboard, Integração com Frontend, desenvolvimento de um Servidor Web. Para o segundo cenário, o hub deve ser capaz de publicar informações dos sensores em um broker MQTT, permitindo que aplicações clientes MQTT acessem esses dados posteriormente. Para tanto, as seguintes atividades foram executadas: Configuração do Broker MQTT, Publicação dos Dados MQTT, Acesso pelos Clientes.

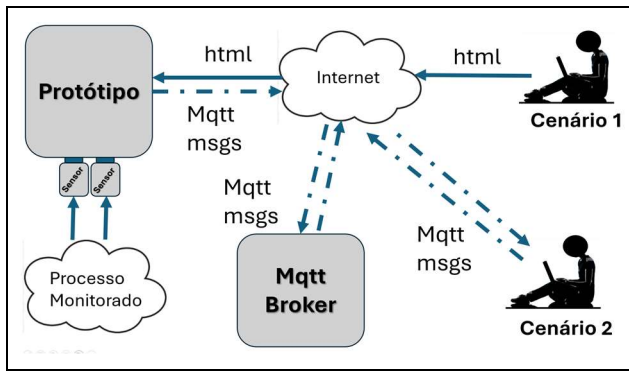


Fig. 1. Cenários de uso para o protótipo de hub/gateway MQTT

A. Seleção de Hardware

Para a construção do hub e gateway MQTT, foram usadas placas Raspberry PI 3 devido ao baixo custo de aquisição das mesmas e sua reconhecida facilidade de uso. Os principais componentes de hardware incluem: Raspberry PI 3.0, Sensores USB, Adaptadores e Cabos.

B. Desenvolvimento do Software

O software do hub e gateway MQTT foi desenvolvido em Python, utilizando bibliotecas amplamente reconhecidas na comunidade IoT. As etapas principais incluem: Configuração do Ambiente, Leitura de Dados dos Sensores, Conversão e Publicação de Dados.

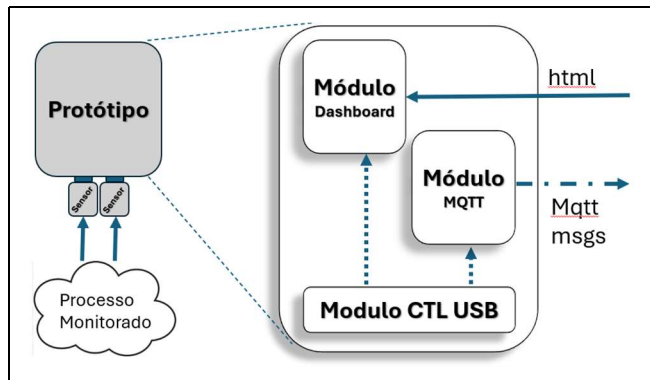


Fig. 2. Detalhamento dos módulos internos do protótipo

A Fig 2. Mostra como o projeto do protótipo considerou a implementação de três módulos distintos: a) O Módulo de gerência do dashboard; b) O Módulo de atualização da informação fornecida ao Broker MQTT externo; e c) O Módulo de coleta dos dados dos sensores usb conectados nas portas da placa Raspberry PI.

III. RESULTADOS E DISCUSSÃO

A implementação do protótipo demonstrou ser eficiente e eficaz em ambos os cenários de uso propostos. No primeiro cenário, o protótipo foi capaz de capturar dados de sensores USB e exibi-los em um dashboard acessível pela web.. Durante os testes, a latência média para a atualização dos dados no dashboard foi de aproximadamente 1 segundo o que é aceitável para a maioria das aplicações IoT que não requerem resposta instantânea. Os dados exibidos no dashboard corresponderam exatamente às leituras dos sensores, confirmando a precisão do sistema. No segundo cenário, o protótipo foi configurado para publicar dados de sensores em um broker MQTT, permitindo que aplicativos clientes MQTT acessassem essas informações. Utilizando a

biblioteca Paho-MQTT, os dados foram enviados periodicamente para o broker. Não foram observadas perdas de dados durante a transmissão para o broker MQTT, demonstrando a confiabilidade do sistema. Diversos clientes MQTT, incluindo dispositivos móveis e aplicativos de desktop, conseguiram subscrever e receber os dados publicados sem problemas. A utilização do protótipo como hardware demonstrou ser eficiente, com consumo de energia e recursos do sistema dentro dos limites esperados.

Os resultados obtidos demonstram que a solução proposta é viável para aplicações IoT de baixo custo, além de oferecer uma plataforma acessível e funcional para o desenvolvimento de hubs e gateways MQTT. A escolha do hardware permitiu reduzir os custos de implementação, tornando a solução atraente para pequenos e médios projetos de IoT.

IV. CONCLUSÕES

Este artigo apresentou o projeto de um protótipo de hub para sensores usb com função de gateway MQTT. A implementação do protótipo mostrou-se uma solução promissora por ser bastante flexível e econômica, particularmente para cenários que exigem monitoramento em tempo real e transmissão eficiente de dados. Os resultados positivos obtidos em ambos os cenários de uso validam a eficácia da abordagem proposta, apesar de algumas limitações e desafios identificados. Futuras melhorias podem ser feitas para otimizar aspectos de escalabilidade e robustez do sistema, bem como na exploração de novos tipos de sensores e interfaces para expandir as capacidades da solução.

REFERENCIAS

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, & M. Ayyash. "Internet of Things: A survey on enabling technologies, protocols, and applications". IEEE Communications Surveys & Tutorials, 17(4), 2015, pp. 2347-2376.
- [2] E. Qosja and A. Wang, "Message Queuing Telemetry Transport (MQTT) Protocol for Wireless Sensor Networks (WSNs) and Internet of Things (IoT)," in *Proceedings of the 2014 6th International Conference on Computational Intelligence, Communication Systems and Networks*, IEEE, 2014, pp. 230-234.
- [3] Y.K. Lam, A.K. Lee, and V.C.M. Leung, "MQTT-S: A Publish/Subscribe Protocol For Wireless Sensor Networks," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2013, pp. 3627-3632.
- [4] Y. Zhang, Z. Wang, & A. V. Vasilakos. "Comprehensive review on communication protocols for IoT". Journal of Network and Computer Applications, 102473, 2020.
- [5] I. Lee, & K. Lee. The Internet of Things (IoT): Applications, investments, and challenges for enterprises. Business Horizons, 2019, 58(4), pp. 431-440.
- [6] K. S. Kim, H. Y. Jeong, S. Y. Son, H. Kim, and H. K. Kim, "A Sensor-Hub System for the Internet of Things with Intelligence," *IEEE Sensors Journal*, vol. 17, no. 15, pp. 4883-4893, August 2017.
- [7] D. Thangavel, V. Ramnan, B. Nagarajan, and D.A. Ahmed, "A Comparative Study of MQTT and HTTP Protocols for Internet of Things," in *Proceedings of the 2014 IEEE International Conference on Emerging Technological Trends*, IEEE, 2014, pp. 1-5.
- [8] X. Su, L. Cheng, and H. Li, "Implementation and Performance Evaluation of MQTT Protocol for Internet of Things," *IEEE Sensors Journal*, vol. 20, no. 18, pp. 10878-10885, 2020.
- [9] H. Liu and X. Jiang, "A Comparative Study of MQTT and WebSocket for IoT Applications," *Future Internet*, vol. 12, no. 8, pp. 1-14, 2020.
- [10] C. R. M. Silva and F. A. C. M. Silva, "An IoT Gateway for Modbus and MQTT Integration," *2019 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, Aveiro, Portugal, 2019, pp. 1-3.

Análise Espectral de Demanda de Energia Utilizando a Transformada de Fourier

1st Etienne Cristina Pereira da Silva
Departamento de Engenharia Elétrica
Universidade Federal da Paraíba
João Pessoa-PB, Brasil
etiene.silva@estudante.cear.ufpb.br

2nd Waslon Terllizie Araújo Lopes
Departamento de Engenharia Elétrica
Universidade Federal da Paraíba
João Pessoa-PB, Brasil
waslon@cear.ufpb.br

Resumo—Este artigo apresenta uma análise espectral das demandas de energia das subestações de uma concessionária de distribuição, utilizando a Transformada de Fourier. Através dessa análise de frequência, os resultados permitiram a identificação de padrões cíclicos e sazonais presentes nos dados.

Palavras-chave—Análise espectral, Demanda, Transformada de Fourier.

I. INTRODUÇÃO

A demanda por energia apresenta variações temporais devido a diversos fatores, incluindo sazonalidade, eventos climáticos e comportamentos dos consumidores. Adicionalmente, é importante destacar as flutuações associadas ao cenário de transição energética atualmente em evidência no país, caracterizado pela descentralização da geração. Assim, torna-se cada vez mais imperativo gerir de forma eficiente a distribuição e a geração de energia, bem como compreender os padrões subjacentes a essas variações para planejamento e operação da rede [1].

A análise espectral da demanda elétrica [2] permite identificar frequências dominantes que correspondem a ciclos naturais de consumo, como variações diárias, semanais, mensais e sazonais [3]. Essas informações contribuem para a otimização da operação das subestações, permitindo a identificação de padrões cíclicos e anomalias que, de outra forma, não seriam facilmente detectáveis no domínio do tempo [4], [5].

Este artigo explora a aplicação da Transformada de Fourier em dados de demanda de energia de subestações para a análise dessas variações e está organizado da seguinte forma: Seção II apresenta a formulação matemática e metodologia de aplicação da Transformada de Fourier nos dados de demanda. A Seção III apresenta os resultados obtidos. Por fim, a Seção IV apresenta as conclusões e propostas para trabalhos futuros.

II. ANÁLISE ESPECTRAL

A Transformada de Fourier é uma operação matemática que transforma uma função do domínio do tempo para o domínio da frequência. A fórmula básica para a Transformada de Fourier contínua é dada por:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (1)$$

Em virtude do tipo de dado disponível para esta análise, faz-se necessário utilizar uma variação da Transformada de Fourier aplicada a casos discretos. Esta variação é chamada de Transformada Rápida de Fourier e é dada pela equação abaixo:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i \frac{2\pi k n}{N}} \quad (2)$$

Os dados utilizados neste estudo foram extraídos de um banco de dados de uma concessionária de energia do estado de São Paulo e Paraná, contendo registros mensais de demanda de energia de 86 subestações, abrangendo o período de fevereiro de 2022 a janeiro de 2024. Cada subestação possui uma série temporal de valores de demanda reais, em MW, obtidas através do software de medição *Hemera*. A Tabela 1 apresenta um exemplo dos dados utilizados.

Tabela I: Exemplo de Dados de Demanda Mensal (MW)

Subestação	fev-22	mar-22	abr-22	mai-22	jun-22
Adamantina	16.32	16.64	15.08	14.68	14.27
Adolfo	3.61	4.34	3.38	4.94	4.34
Água do Almoço	2.14	1.75	1.34	2.51	2.01

A Figura 1 apresenta a área de concessão onde as subestações estão englobadas.

A análise espectral dos dados foi realizada utilizando o ambiente de desenvolvimento *Google Colab*, através da linguagem *Python* e suas bibliotecas *Pandas*, para leitura dos dados, *Numpy*, para aplicação das funções referentes a Transformada de Fourier e *Matplotlib*, para visualização dos resultados em forma gráfica.

Os dados foram normalizados para elaboração da análise, com metodologia de captura da média de cada série temporal e divisão por seu desvio padrão. Esta normalização é necessária para remoção de outliers e possíveis distorções nos resultados da aplicação da Transformada de Fourier, obtendo foco apenas nas variações cíclicas [4].

III. RESULTADOS

A partir da análise dos dados de demanda registrados em anos completos, foi possível identificar padrões sazonais. Nos anos de 2022 e 2023, os picos de demanda foram medidos

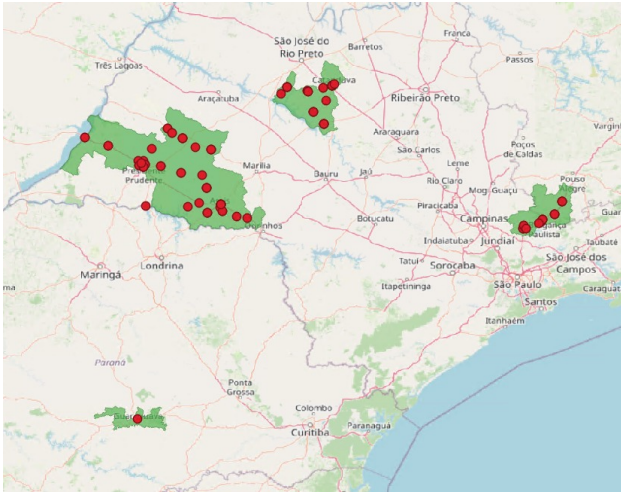


Figura 1: Localização das Subestações

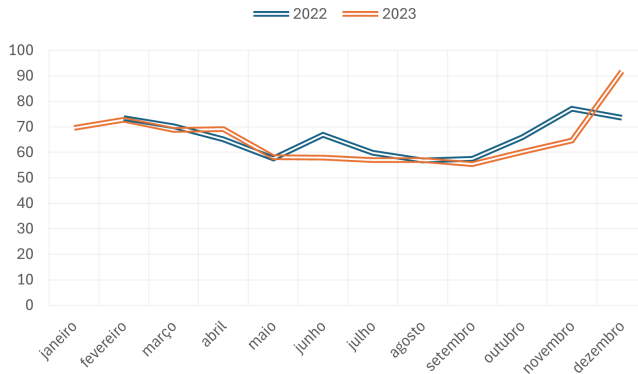


Figura 2: Curva de Demanda 2022 e 2023

nos meses de outubro a dezembro. A Figura 2 mostra a curva de demanda por ano.

A Transformada de Fourier foi aplicada individualmente para análise da resposta em frequência de cada subestação e em seguida, foi realizada a aplicação nos valores normalizados das 86 subestações do banco de dados.

A Tabela 2 apresenta as 5 subestações analisadas que apresentaram maior variação de demanda.

Tabela II: Subestações com as Variações mais Significativas nas Demandas Elétricas

Subestação	Desvio Padrão da Demanda
Catanduva I	12.34
Presidente Prudente I	6.64
Presidente Prudente IV	6.05
Assis I	5.29
Presidente Venceslau	4.73

A Figura 3 representa os resultados no domínio do tempo e a resposta em frequência após aplicação da Transformada de Fourier.

Os resultados individuais indicam que a maioria das subestações apresenta um pico significativo na frequência

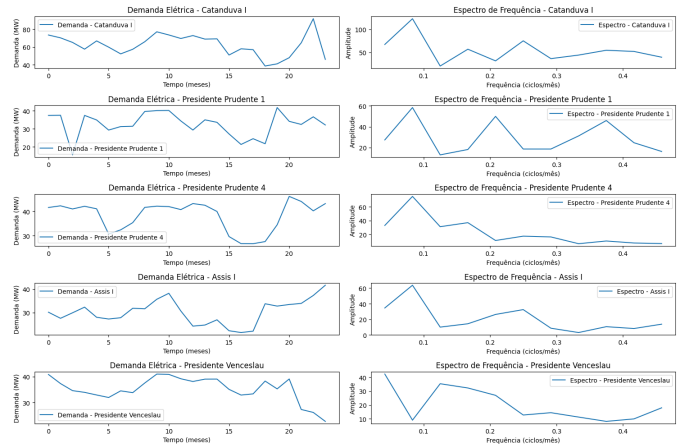


Figura 3: Resultados Domínio do Tempo e Frequência

correspondente a um ciclo anual, refletindo variações sazonais. Além disso, alguns espectros apresentam picos em frequências menores, sugerindo ciclos mensais ou bimestrais no consumo de energia. A resultante do padrão anual está em linha com as condições climáticas da região noroeste do estado de São Paulo e sudoeste do Paraná, onde a concessão está situada, que experimenta período de seca e temperaturas mais elevadas no início do verão.

IV. CONCLUSÃO

A análise espectral da demanda de energia utilizando a Transformada de Fourier fornece uma visão detalhada dos padrões de consumo nas subestações. Este estudo destaca a importância de considerar componentes de frequência ao planejar a operação e manutenção das subestações, de modo a melhorar a eficiência e gestão de recursos no setor elétrico. Como trabalhos futuros, propõe-se a extensão da análise considerando parâmetros de distúrbio da rede elétrica para comparação, como penetração de energia reativa consequente de geração distribuída e fatores climáticos adversos como o El Niño.

REFERÊNCIAS

- [1] Cristiane A. J. Schmidt e Marcos A. M. Lima. "A demanda por energia elétrica no Brasil". *SciELO Brazil*, 2020. Disponível em: <https://www.scielo.br/j/article> "Signal Compression: Technology Targets and Research Directions". *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 5, pp. 796–818, June 1992.
- [2] V. Ferreira. "PSI3214-2020: Análise Espectral da Tensão da Rede Elétrica". *e-Disciplinas, Universidade de São Paulo (USP)*, 2020. Disponível em: <https://edisciplinas.usp.br/mod/resource/view.php?id=3162573>
- [3] C. H. Lee e Y. J. Kim. "Seasonal Variations in Electric Demand and Fourier Analysis". *Energy Reports*, vol. 8, pp. 1220–1234, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352484722001234>
- [4] J. A. Smith e L. Brown. "Fourier Transform Applications in Power Systems Analysis". *Journal of Electrical Engineering*, vol. 45, no. 3, pp. 345–356, 2021. Disponível em: <https://www.mdpi.com/1424-8220/20/9/2671>
- [5] Silvio A. de Souza. "Utilização da Transformada de Fourier Janelada para caracterização de distúrbios na qualidade da energia elétrica". *Universidade de São Paulo (USP)*, 2022. Disponível em: <https://www.teses.usp.br/teses/disponiveis/18/18133/tde-12052017-104820/en.php>

Aplicação da FFT em Sinais de Sensores de Nível em Reservatórios Apoiados de Água

1st Márcio Miranda Córdula

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba
João Pessoa, Brasil
marcio.cordula@estudante.cear.ufpb.br

2nd Diego Perazzo Creazzola Campos

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba
João Pessoa, Brasil
diego.campos@estudante.cear.ufpb.br

3rd Laryssa de Lima Sousa

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba
João Pessoa, Brasil
laryssa.sousa@estudante.cear.ufpb.br

4nd Cícero da Rocha Souto

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba
João Pessoa, Brasil
cicerosouto@cear.ufpb.br

5nd Waslon Terllizie A. Lopes

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba
João Pessoa, Brasil
waslon@cear.ufpb.br

Resumo—Este artigo propõe um sistema de monitoramento remoto dos mananciais da grande João Pessoa, operados pela CAGEPA. O sistema coleta dados dos níveis das barragens de Marés e Gramame e enviando-os ao centro de operação da Cagepa. A implementação oferece informações em tempo real e ajuda na construção de um histórico de dados, permitindo análises e correções rápidas de problemas, reduzindo perdas hídricas.

Palavras-chave—processamento digital de sinais, SCADA-LTS, ESP8266, medição de nível, transformada rápida de Fourier.

I. INTRODUÇÃO

A implementação de um sistema de monitoramento melhora a eficiência dos operadores, fornecendo informações em tempo real e criando um histórico detalhado dos dados das barragens, facilitando a análise. Esse sistema identifica e corrige rapidamente falhas nas estações de captação e tratamento de água, reduzindo perdas hídricas e ajudando a alcançar as metas do Marco Legal do Saneamento no Brasil. O objetivo deste artigo é analisar o sinal do sensor de nível no domínio da frequência para identificar ruídos que interferem na medição. Além disso, aplica-se um filtro para minimizar esses ruídos, melhorando a precisão do sistema de monitoramento.

II. MONITORAMENTO DO NÍVEL DOS MANANCIAIS

Para monitorar o nível dos mananciais, um sensor hidrostático com corrente de saída proporcional à coluna de fluido acima dele está sendo utilizado (na faixa de 4 mA a 20 mA), que corresponde respectivamente à uma coluna de fluido de 0 a 10 metros de profundidade. Como a placa ESP8266 trabalha com níveis de tensão em suas entradas, será utilizado um resistor de $3,3 \text{ V}/20 \text{ mA} = 165 \Omega$ em série com o sensor, possibilitando assim a medição do nível dos mananciais. O sistema está representado pela Figura 1.

O conversor analógico-digital do ESP8266 é de 10 bits, o microcontrolador é capaz de distinguir até 1024 níveis discretos de tensão entre 0 e 3,3 V. O comportamento da

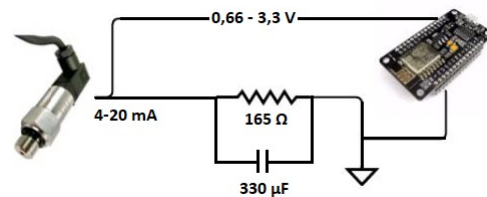


Figura 1. Sistema utilizado para medições de nível, respectivamente.

corrente de saída do sensor em função do nível do poço é linear, tem-se como consequência uma relação também linear entre nível do manancial e leitura convertida pelo conversor digital-analógico. Nesse caso será necessário discretizar o sinal de entrada e levantar a curva de calibração do conversor AD do ESP8266.

O comportamento da corrente de saída do sensor em função do nível do poço é linear, tem-se como consequência uma relação também linear entre nível do poço e a leitura convertida pelo conversor analógico-digital. A expressão que relaciona o nível do poço N em metros e a leitura L obtida pelo conversor analógico-digital é tal que $N = \frac{10-L}{819}$. Aplicando-se a equação obtida anteriormente para o cálculo do nível do poço ao ensaio de calibração da entrada analógica do ESP8266 temos:

$$y = N' = \frac{0,938(10L - 2050)}{819} + 0,003. \quad (1)$$

Com isso, observou-se diminuição considerável no erro de medição, bem como das oscilações entre medições sucessivas.

III. PROBLEMÁTICA

Devido à presença de inversores de potência no local onde o sistema está instalado, há interferência eletromagnética, resultando em ruídos no sinal do sensor de nível. Para minimizar

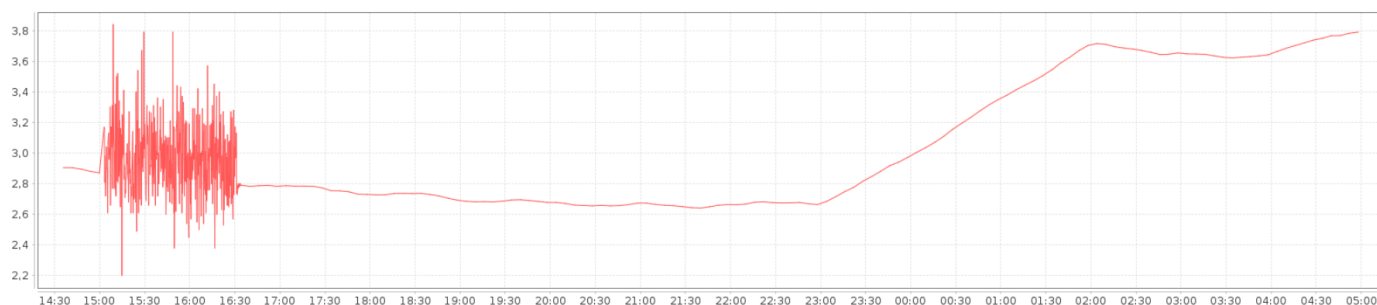


Figura 2. Resultado do nível recebido pelo sistema SCADA.

esse ruído, foi aplicado um filtro passa-baixas, implementado por meio do capacitor mostrado na Figura 1.

IV. RESULTADOS

Os dados foram coletados utilizando a placa de aquisição NI USB-6212, com 16 bits de resolução, 6 mil amostras e frequência de amostragem de 6 kHz, totalizando um tempo de janelamento de 1 segundo. A Figura 3 representa a análise do sinal no domínio do tempo, (Figura 3(a)) com filtro e (Figura 3(b)) sem filtro, respectivamente. É possível verificar que o sinal é bastante ruidoso sem a presença do filtro no sistema. Na Figura 4 representa o sinal no domínio da frequência da Transformada Rápida de Fourier. As frequências de maior amplitude são encontradas em 165 Hz, 2665 Hz e 3000 Hz. As frequências de 2665 Hz e 3000 Hz são provenientes do chaveamento dos IGBTs dos inversores de frequência presentes no ambiente, enquanto a frequência de 165 Hz é provocada pela corrente de saída dos motores que operam com rotação variável.

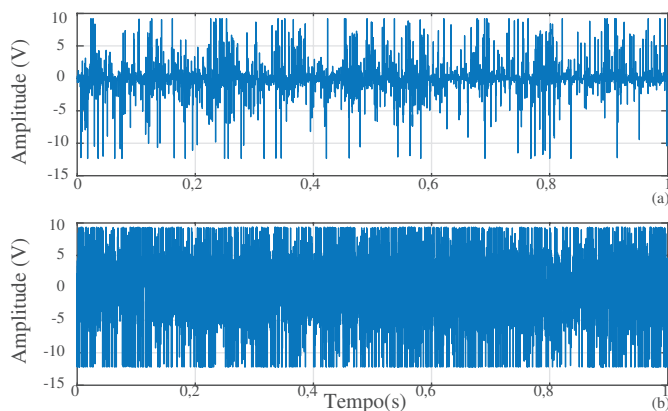


Figura 3. Sinal no domínio do tempo sem e com filtro, respectivamente.

A Figura 2 foi retirada do sistema supervisor Scada-LTS em operação após toda a sua implementação e representa o sinal com e sem filtro. O sinal coletado das 15h às 16h30 apresenta interferência provocada principalmente por motores e inversores de frequência presentes no ambiente em que o sensor está inserido, além de toda a turbulência promovida pelo próprio fluido, provocando uma variação de 80 cm em torno do valor médio. O filtro foi inserido após as 16h30,

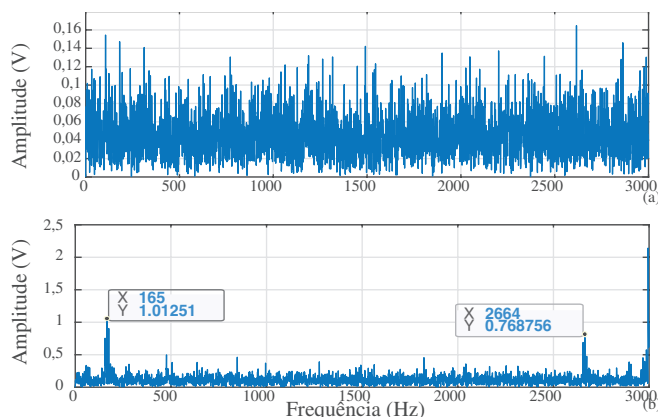


Figura 4. Sinal no domínio da frequência com e sem filtro.

suavizando a leitura do sensor conforme apresentado na Figura 2.

V. CONCLUSÃO

A Transformada Rápida de Fourier demonstrou ser uma ferramenta muito eficaz para analisar o sinal de nível do sensor piezoresistivo, contribuindo na etapa do projeto do filtro que atenda às necessidades do sistema de aquisição de dados. O filtro, por sua vez, atendeu aos requisitos do projeto, filtrando todas as interferências presentes no processo de monitoramento remoto.

AGRADECIMENTOS

Márcio Miranda Córdula agradece a CAGEPA, ao Prof. Dr. Waslon Terlizie A. Lopes, ao Prof. Dr. Cícero da Rocha Souto (LaSea-UFPB), a Laryssa de Lima Sousa (UFPB) e Diego P. C. Campos (CAGEPA) pelo suporte dado a este projeto.

REFERÊNCIAS

- [1] M. M. Córdula. "Relatório de Aproveitamento de Atividade Profissional Realizada na Cagepa – Companhia de Água e Esgotos da Paraíba". 2022.
- [2] S. de T. M. Bezerra et al. "Controle automático de pressão em sistemas de distribuição de água". *Sistemas de Saneamento – Eficiência Energética*, João Pessoa, 1ª Edição, pp. 9–29, 2010.
- [3] J. M. M. Guidi. "Análise da Variação do Consumo de Água Utilizando Dados Obtidos por Sistemas Supervisores Remotos: Estudo de Caso: Município de Franca-SP". *FENASAN*, 2017.
- [4] B. P. Lathi. "Sinais e Sistemas". Porto Alegre: Bookman, 2001. 662p.

Técnicas de Processamento de Dados Para Análise de Vibração em Redes de Distribuição de Água

1st Diego Perazzo Creazzola Campos

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba

João Pessoa, Brasil

diego.campos@estudante.cear.ufpb.br

2nd Laryssa de Lima Sousa

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba

João Pessoa, Brasil

laryssa.sousa@estudante.cear.ufpb.br

3rd Cícero da Rocha Souto

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba

João Pessoa, Brasil

cicerosouto@cear.ufpb.br

4nd Márcio Miranda Córdula

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba

João Pessoa, Brasil

marcio.cordula@estudante.cear.ufpb.br

5nd Waslon Terllizie A. Lopes

Departamento de Engenharia Elétrica
Universidade Federal da Paraíba

João Pessoa, Brasil

waslon@cear.ufpb.br

Resumo—O processamento de dados de vibração em tubulações de redes de distribuição de água possui grande potencial para ser uma ferramenta no combate às perdas. Vibrações anormais podem indicar anomalias em redes de distribuição de água, como vazamentos (perceptíveis ou não) bem como a presença de ar. Monitorar esses sinais permite uma análise contínua e não invasiva, possibilitando intervenções preventivas, minimizando o impacto das perdas físicas. Para esse processamento, foram utilizadas a Transformada Rápida de Fourier (FFT), reconhecida por sua eficiência na análise espectral e a Transformada Wavelet Contínua (TWC) que oferece uma melhor resolução temporal e frequência. Na conclusão o artigo é apresentado um estudo comparativo de um evento de vazamento induzido, onde foram aplicadas as duas transformadas, cuja combinação se mostrou eficaz para a caracterização de vazamentos.

Palavras-chave—processamento de sinais, transformada rápida de Fourier, transformada Wavelet, vazamentos em redes de distribuição de água.

I. INTRODUÇÃO

A análise de dados de vibração em redes de distribuição de água pode se tornar uma ferramenta eficaz na detecção de vazamentos. Estudos apontam que 2/3 dos vazamentos são imperceptíveis [4]. A busca por métodos mais avançados para detecção dessa tipologia de vazamento impacta diretamente na viabilidade econômica das empresas de saneamento. Abordagens tradicionais (inspeções visuais e hastes de escuta) são limitadas. Este estudo compara a transformada Rápida de Fourier (FFT) e a Transformada Wavelet para determinar qual técnica se mostra mais eficaz na caracterização e identificação de vazamentos.

II. TRANSFORMADA RÁPIDA DE FOURIER (FFT)

A Transformada Rápida de Fourier (FFT), uma implementação matemática da Transformada Discreta de Fourier (DFT), é uma ferramenta amplamente utilizada na

análise de sinais no domínio da frequência. O objetivo da utilização nesse estudo foi o de verificar as componentes frequenciais dos sinais de vibração antes e durante o vazamento induzido para caracterização de padrões. Neste trabalho foi utilizada a função `fft()` do Matlab para calcular a Transformada de Fourier, que é baseada nas bibliotecas apresentadas em [1], [2].

III. TRANSFORMADA WAVELET CONTÍNUA (TWC)

Transformada Wavelet oferece uma análise em tempo-frequência, sendo capaz de lidar com sinais não estacionários. Ela permite a decomposição do sinal em várias escalas, capturando tanto as características de alta frequência, associadas a ruídos e pequenas alterações, quanto as de baixa frequência, que podem indicar variações mais significativas no sistema.

A utilização dessa transformada no trabalho foi por meio da função do Matlab `ctw()` que se baseia na equação de wavelet para calcular de forma discreta a transformada de wavelet.

IV. SISTEMA DE AQUISIÇÃO E PROCESSAMENTO

Na figura 5, está ilustrado um esquema que mostra como foi realizada a montagem do sistema de aquisição utilizando a placa de aquisição NI USB-6212.

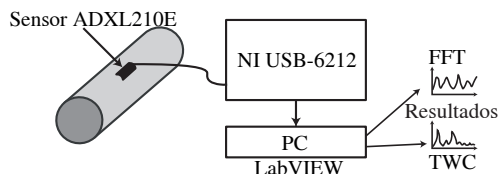


Figura 1. Sistema de aquisição.

O sistema é composto de um sensor de vibração (ADXL210E), que é um acelerômetro, uma placa de aquisição da National Instruments, e uma interface elaborada no Lab-View, que possibilita salvar as informações do sensor. As

informações salvas são processadas por um script de Matlab que resultam nas duas transformadas do sinal (Fourier e Wavelet).

V. RESULTADOS

Os resultados apresentados nesta seção foram obtidos ao medir a vibração de uma tubulação de água DN25 utilizando o sistema de aquisição descrito. Dois cenários diferentes foram obtidos, a partir dos resultados: no primeiro cenário não há vazamentos na tubulação, e no segundo cenário, um vazamento foi induzido, danificando a tubulação para que o sinal pudesse ser captado durante o vazamento, possibilitando a realização do estudo comparativo.

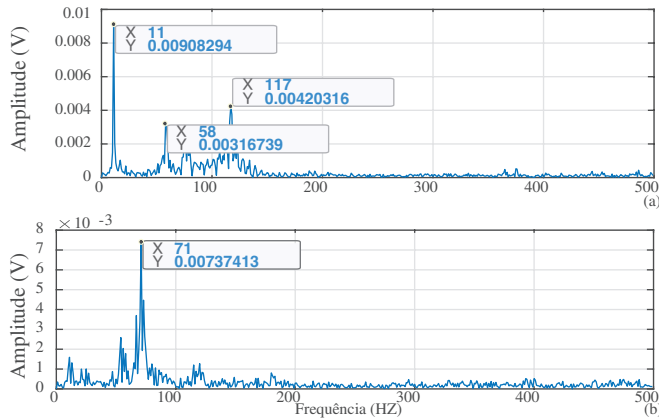


Figura 2. FFT do sinal antes e durante o vazamento.

Nas figuras 2 (a) e 2 (b) são apresentados os gráficos da FFT dos cenários antes e depois do vazamento, respectivamente. Na comparação entre os dois gráficos, é possível observar que a frequência fundamental do sinal mudou de 11 Hz para 71 Hz. Além disso, no gráfico da FFT durante o vazamento é possível notar que a amplitudes de outras componentes de frequência mudaram em comparação ao gráfico da FFT antes do vazamento. Essas diferenças que podem ser utilizadas para a identificação de vazamentos.

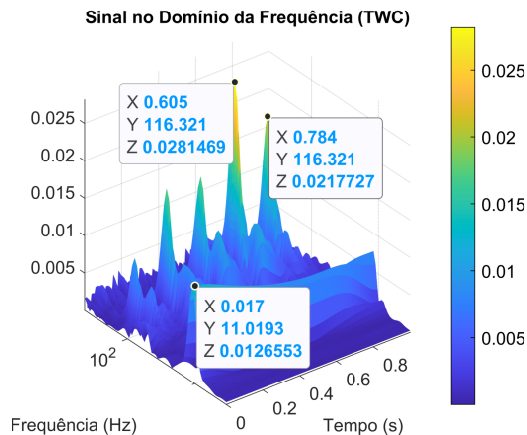


Figura 3. TWC do sinal antes do vazamento.

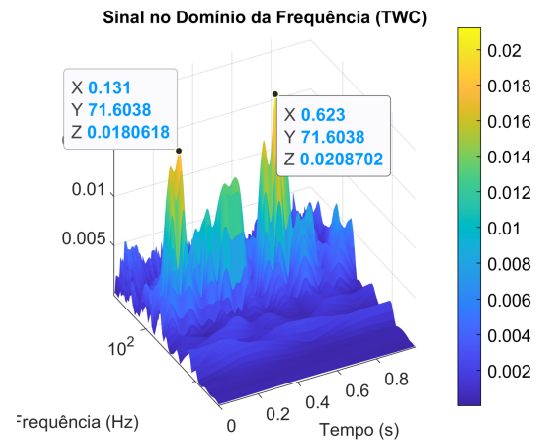


Figura 4. TWC do sinal durante o vazamento.

Nas figuras 3 e 4 são apresentados os gráficos da TWC dos cenários antes e depois do vazamento, respectivamente. A transformada de Wavelet revelou outra característica do sinal que pode ser utilizada para a detecção de vazamentos. No gráfico da figura 3, a componente principal de frequência apresenta variações suaves ao longo do tempo, sendo esta uma característica de um cenário sem vazamentos. Enquanto isso, no gráfico da figura 4, a frequência principal apresentou variações mais bruscas ao longo do tempo, caracterizando desta forma um vazamento.

VI. CONCLUSÕES

O artigo apresentou um sistema de aquisição de dados para processar sinais de vibrações em tubulações, utilizando a FFT e a TWC. A FFT é eficiente na identificação da frequência dominante associada ao vazamento, enquanto a TWC fornece informações sobre o momento do vazamento, permitindo respostas mais eficazes. Essas características podem ser usadas para criar algoritmos de detecção de vazamentos e sistemas de monitoramento para reduzir perdas. A elaboração de algoritmos baseados nas transformadas de Fourier e Wavelet é uma perspectiva para trabalhos futuros.

AGRADECIMENTOS

Diego Perazzo Creazzola Campos agradece ao Laboratório de Sistemas e Estruturas Ativas (LaSEA) juntamente ao prof. Dr. Cícero da Rocha Souto, Laryssa de Lima Sousa, Márcio Miranda Córdula, pelo suporte a esta pesquisa e ao Prof. Dr. Waslon Terllizie A. Lopes a orientação durante a disciplina.

REFERÊNCIAS

- [1] M. Frigo and S. G. Johnson. "FFTW: Fastest Fourier Transform in the West". <http://www.fftw.org>. Accessed: June 24, 2024.
- [2] M. Frigo and S. G. Johnson. "FFTW: An Adaptive Software Architecture for the FFT". *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1998, pp. 1381–1384.
- [3] M. V. de Siqueira and L. F. A. Campos. "Transformada Wavelets". *Semana Acadêmica*, Accessed: June 24, 2024. https://semanaacademica.org.br/system/files/artigos/transformada_wavelets.pdf
- [4] E. G. de Melo and R. R. Machado. "Curso Básico de Controle e Redução de Perdas". *Curso EAD*, São Paulo, SP, 2017.

5G SA Throughput Performance Evaluation for Different MIMO Configuration

1st Marcos A. A. Pontes ^{*}, 2nd Isys G. S. De Souza [†], 3th Felipe R. M. De Andrade [‡], 4nd Yuri H. S. Barbosa [§]

Sidia Institute of Science and Technology - Manaus, Brazil

Emails: (marcos.pontes-e ^{*}, isys.souza [†], felipe.andrade [‡], yuri.barbosa [§])@sidia.com

Abstract—Modern users demand data services to be fast and reliable, and because of that, one of the primary purposes of radio technology has been to increase data rates to offer better services. The antenna configuration used is critical to determining service quality in relation to data rate. This article describes the findings of laboratory tests performed on 5G smartphones to assess the effect of different Multiple-Input Multiple-Output and Single-Input Single-Output configurations on data rate over a 5G network. According to the data, more transmitters and receiver schemes dramatically improve throughput and decrease the transfer time needed to complete a download. The MIMO-4X4 scenario, in particular, increased transfer rates by up to 281% concerning the SISO scenario while reducing transfer times from 3.32 minutes to 56 seconds for a 10Gb file.

Index Terms—5G, SA, SISO, MIMO

I. INTRODUCTION

The fifth generation of mobile networks (5G) represents a significant advancement over previous technologies. The next generation of mobile communication introduces new features such as ultra-reliable low-latency communications (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (MTC). To offer these advancements, 5G continues to use some technologies from the previous generation. Techniques like multi-input multi-output (MIMO) are widely applied to enhance the throughput for user equipment (UE) [1] [2].

MIMO for mobile networks refers to the fact that when a packet enters the channel, it is transmitted via multiple antennas and received by multiple antennas upon exit. In contrast, a Single-Input Single-Output (SISO) system has one antenna at each end of the link. This technique enables the transfer of more bits per second, resulting in improved performance for User Equipment (UE). As defined in [4], many Transmission Modes (TM) can be applied in a 5G network; MIMO transmission may be performed in a variety of modalities, such as diversity transmission, in which the same signal is delivered to several antennas, boosting reliability in diverse propagation paths. Another method for transmitting data via MIMO is spatial multiplexing, which sends separate data streams to each transmitting antenna, boosting data rate [3] [4]. This researcher used diversity transmission for the MIMO scenario to keep the focus on the best throughput for the UE.

This work is structured as follows: Section II discusses the essential elements of the 5G and MIMO. Section III

presents the methodology used to assess the performance of all simulated scenarios. Section IV shows the results of all measurements. Section V presents the conclusion of this work.

II. KEY CONCEPTS

A. 5G New Radio

5G technology seeks to deliver more reliable communication for humans, smart devices, and systems. The 5G system integrates innovative technologies like intelligent antennas, adaptive modulation, and network virtualization to give more bandwidth, lower latency, and enhanced connection capacity. 5G networks can accommodate a variety of wireless technologies. MIMO is a crucial technology in 5G, enabling the network to modify the number of antennas available to send data according to network conditions to ensure the Quality of Service (QoS) and throughput necessities of the final user [5] [6].

B. MIMO for Mobile Network

MIMO for mobile communication is an antenna configuration paradigm designed to increase data flow by utilizing various configurations of antennas to send and receive packets. MIMO can provide users with improved quality signals transmitted from mobile networks and high throughput performance in UE services [7].

The technical specifications for the 3GPP include details on how to enhance communication between base stations (eNBs) and user equipment. This improved transmission method aims to enhance throughput, reliability, and user management. One example is Single-Input Single-Output (SISO), which uses a single antenna on both sides for essential communication. Multiple-Input Multiple-Output (MIMO) enhances reception by employing multiple receive antennas at the UE. Different scenarios leverage MIMO technology with multiple transmit antennas at the eNB, incorporating feedback from the UE for improved performance. Later versions introduced even more complex MIMO configurations for higher capacity. Release 10 introduced techniques for reducing interference and improving stability by separating data and reference signals. This work focuses explicitly on MIMO diversity transmission, aiming to achieve and compare high data rates in multiple antenna configuration scenarios [3] [8].

III. METHODOLOGY

The test configuration used a high-end smartphone with a Snapdragon 8 Gen 3 chipset and a Snapdragon X75 modem with 5G SA capabilities. The Anritsu Signaling Tester

This work was presented as part of the result of Project AMAN, executed by Sidia Institute of Science and Technology in partnership with Samsung Eletrônica da Amazônia LTDA, according to Informatics Law n.8387/91 and Art.39 of Decree 10.521/2020.

MT8000A was employed to simulate an ideal measurement environment, which can emulate 5G base station functions. The Anritsu MA8161A shield box could be employed to prevent external radio interference. Three measurement scenarios were created with the following parameters: band NR n78 with a bandwidth of 100 MHz, Sub-Carrier Spacing (SCS) of 30 kHz, and Modulation and Coding Scheme (MCS) of 256-QAM and 27, respectively. All data collected during the test was for a 90-second duration, and the throughput values were calculated for each successful frame of data received by the device, with an average of 0.5 seconds.

IV. RESULTS

The results were achieved with the premise that SISO or MIMO configuration would be one of the most significant modifications in all three scenarios. To ensure that the best throughput valuer was collected, the SNR was set to the best scenario channel, making it possible to ensure the ideal network condition in all cases of the test, and it was possible to avoid interference or noise signal in the results. The signal level remains consistent for all transmission modes. Based on the above description, 1 displays the throughput profile for three transmission scenarios: SISO, MIMO - 2X2, and MIMO - 4X4. The results reveal that throughput (kbps) increases with more antennas being used to transmit data, i.e. when the scenario uses MIMO - 4X4, the data rate is significantly greater than with SISO and MIMO - 2X2 as expected. The transmission mode with MIMO - 4X4 went up to 1544.81 Mbps, MIMO - 2X2 reached 862.03 Mbps, and SISO achieved 405.63 Mbps.

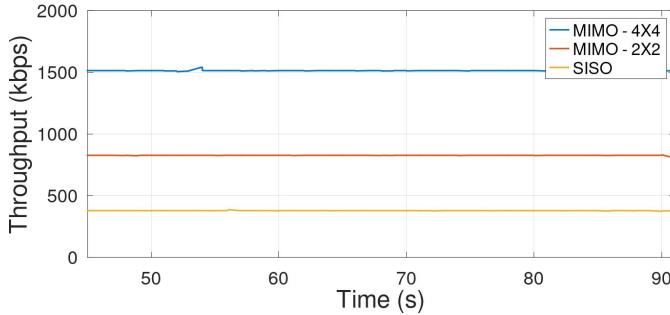


Fig. 1: Measurement data of all MIMO and SISO scenario

Using the data shown in the figure 1 and presented in the results section, it is possible to calculate the time cost in terms of the download time effectiveness of a file, where the best MIMO configuration, MIMO-4X4, must conduct the downloads faster than in other scenarios. Comparing MIMO-4X4 to MIMO-2X2 and SISO in terms of sample measurement will help illustrate this. Figure 2 indicates that the MIMO-2X2 can increase about 113% of the data rate concerning SISO configuration by more than twice the throughput. Looking for MIMO - 4X4, it is possible to find the most significant variation of the tests, where SISO could reach an increase of 281% , almost four times more data rate. From another viewer's point of time-saving, it is possible to go from 3 minutes and 32 seconds of download time of a 10 Gigabyte file to only 56 seconds. In addition, with MIMO-2X2 configuration, it is also

possible to find significant results, enabling to reach 113% more data rate than SISO, which means downloading the same file of 10Gb in 1 minute and 40 seconds, a difference of 79% of MIMO-4X4.

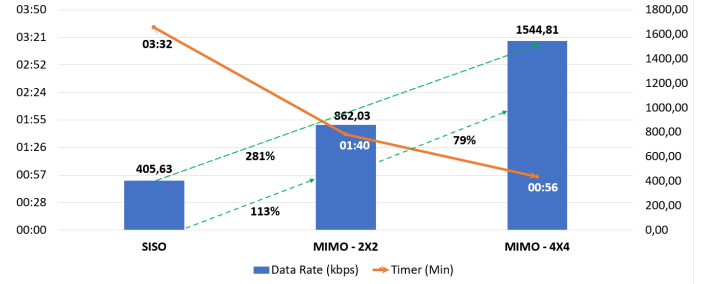


Fig. 2: Comparison between scenarios of MIMO and SISO

V. CONCLUSION

This work evaluated the impact of different transmission modes in 5G SA, considering three scenarios (SISO, MIMO-2X2, and MIMO-4X4). With the result of measurements in the simulated environment, as well as expected, the scenario with the highest transmission mode obtained the best results. In terms of the throughput achieved from each scenario, the MIMO-4X4 obtained values that were up to 281% greater than the SISO and 79% higher than the MIMO-2X2. Comparing the MIMO-2X2 scenario with that of SISO obtained a data rate of approximately 113% higher. From a download time perspective, the scenarios must download a 10Gb file in 0.56 seconds, 1.40, and 3.32 minutes for MIMO-4X4, MIMO-2X2, and SISO configuration, respectively. In conclusion, it is feasible to conclude that the higher transmission mode scenario outperforms the others, indicating that a set between device and network that provides MIMO-4X4 could improve the quality of throughput services for users.

REFERENCES

- [1] G. D. Surabhi and A. Chockalingam, "Low-complexity linear equalization for 2x2 mimo-otfs signals," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.
- [2] X. Lin and N. Lee, "5g and beyond: Fundamentals and standards," pp. 1–551, 2021.
- [3] 3rd Generation Partnership Project, "Lte; evolved universal terrestrial radio access (e-utra); physical layer procedures," *ETSI TS 136 213*, vol. 10, 2018.
- [4] D. A. Joe, K. Madhumitha, B. R. Prabhaa, and P. Dharshanya, "2 x 2 mimo antenna design for 5g applications," in *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, 2023, pp. 1–6.
- [5] M. Gundall, M. Strufe, H. D. Schotten, P. Rost, C. Markwart, R. Blunk, A. Neumann, J. Griebbach, M. Aleksy, and D. Wübben, "Introduction of a 5g-enabled architecture for the realization of industry 4.0 use cases," *IEEE Access*, vol. 9, pp. 25 508–25 521, 2021.
- [6] F. Bouhaf, A. Raschella, M. Mackay, and F. den Hartog, "A spectrum management platform architecture to enable a sharing economy in 6g," *Future Internet*, vol. 14, no. 11, 2022.
- [7] M. J. Riaz, A. Sultan, M. Zahid, A. Javed, Y. Amin, and J. Loo, "Mimo antennas for future 5g communications," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 2020, pp. 1–4.
- [8] X. Z. Xincheng Zhang, "Lte-advanced air interface technology," *CRC Press*, vol. 18, pp. 243–244, 2012.

A Estrutura do 6G, Casos de Uso e Capacidades

Luiz Gonzaga de Queiroz Silveira Júnior
Departamento de Engenharia de Comunicações
Universidade Federal do Rio Grande do Norte
Campus de Lagoa Nova, Natal, Rio Grande do Norte
luiz.silveira@ufrn.br

Abstract—Esse trabalho apresenta os casos de uso e as capacidades definidos pela estrutura do IMT-2030 da ITU-R, cujas diretrizes orientam os esforços em termos de pesquisa, desenvolvimento e padronização global para o futuro das comunicações em Terahertz. Além disso, discute alguns dos esforços necessários para a superação dos desafios que se impõem ao projeto da interface aérea para as redes 6G e o papel que a IA terá como tecnologia habilitadora capaz de tornar o 6G uma realidade. Dessa forma, a discussão realizada nesse trabalho contribui para desmistificar as características principais, os requisitos de desempenho e os casos de uso potenciais da próxima geração das redes de comunicações móveis, 6G.

Index Terms—ITU-R, IMT-2030, 6G, Casos de Uso, Capacidades.

I. INTRODUÇÃO

A Seção de Radiocomunicações da União Internacional de Telecomunicações (ITU-R) desempenha um papel fundamental na definição da visão, dos requisitos técnicos e das metodologias de avaliação de desempenho para a sexta geração das redes de comunicações móveis (6G), que é denominada de “IMT-2030” pela ITU-R, e que deve se tornar uma realidade apenas em 2030 [1].

Até o momento, somente a estrutura do IMT-2030 está definida pela ITU-R.

Pode parecer pouco, mas tal definição tem o condão de orientar os esforços para a pesquisa, o desenvolvimento, a inovação e a padronização que serão realizados até 2029, por organizações governamentais, academia, institutos de pesquisa, associações da indústria de equipamentos e de serviços de telecomunicações e demais entidades com alcance global interessadas no estabelecimento, operação e evolução da próxima geração das redes de comunicações celulares, 6G.

Este trabalho apresenta a estrutura do IMT-2030. Além disso, discute alguns dos esforços necessários para superar os desafios do projeto da interface aérea para redes 6G, destaca o papel da IA como uma tecnologia habilitadora para tornar o 6G uma realidade e desmistifica as principais características, requisitos de desempenho e casos de uso potenciais da próxima geração de redes de comunicações móveis.

II. CASOS DE USO

No último ano, o grupo de trabalho “WP5D” da ITU-R, que é formado por especialistas de todo o mundo nos sistemas de comunicações móveis, concluiu seu trabalho de recomendação da estrutura do IMT-2030, a qual define casos

de uso e capacidades do 6G [2]. Os 3 casos de usos que foram definidos como principais são¹ [2]–[4]:

- Comunicação Imersiva, necessário para prover a oferta de experiências envolventes e realistas, incluindo desde aplicações de realidade estendida, onde há a renderização de objetos, até aplicações com holografia, onde a presença holográfica em ambientes remotos poderia incluir a telecirurgia e o entretenimento. Trata-se de uma evolução do eMBB;
- Conectividade Ubíqua, que integrará todas as infraestruturas das redes de comunicação, desde as existentes no assoalho dos oceanos, ao espaço, de forma a prover conectividade ininterrupta, tanto em áreas escassamente povoadas, quanto para a conectividade massiva das redes densas de sensores de diversos setores da economia, como a Indústria Inteligente Conectada à Nuvem; nesse caso, trata-se de uma evolução do mMTC.
- Sensoriamento e Comunicações Integradas, que prevê a integração entre os sistemas de percepção e os sistemas de comunicação. Dessa forma, as redes de comunicações evoluiriam para além da transmissão de dados, e seriam utilizadas na percepção e interação com o meio ambiente onde estão inseridas. Dessa forma, o 6G tornará realidade o sensoriamento como um serviço de rede, que pode impulsionar verticais importantes.

Uma das maiores expectativas que surge com o advento das redes 6G é a aplicação de Gêmeos Digitais no contexto da Saúde 4.0, principalmente para o advento do gêmeo digital humano, o que favorece a habilitação de serviços de saúde personalizados e orientados por IA [5], [6].

III. CAPACIDADES DO 6G

Para satisfazer as demandas dos casos de uso elencados acima e de forma a suportar a ampla gama de aplicações em 2030 e além, a sexta geração das redes de comunicações móveis precisará suportar requisitos de desempenho ainda maiores. Para isso, será necessário atingir as seguintes capacidades [7]–[11]:

- Aumentar em 50 vezes a taxa de pico, de forma a alcançar 1 Tbps;
- Aumentar a taxa de transmissão para o usuário para 10 Gbps;

¹Os casos de usos do 5G, eMBB, mMTC e uRLLC, continuarão em evolução no 6G e coexistirão com os casos de usos novos.

- Reduzir a latência para 0,1 ms;
- Aumentar a capacidade para suportar cerca de 10^8 dispositivos por km^2 ;
- Aumentar a mobilidade para cerca de 1000 km/h;
- Aumentar a eficiência energética em 100 vezes com relação ao 5G.

Para o atendimento dos casos de uso e das capacidades exigidas será necessário vencer muitos desafios.

O FG NET-2030 (sigla inglesa para *Focus Group Technologies for Network 2030*) foi estabelecido em julho de 2018 pela ITU-T para identificar as lacunas e desafios tecnológicos para o advento das redes 6G [10].

Desde então, o FG NET-2030 revisou padrões e tecnologias maduras, avaliou inovações tecnológicas, estabeleceu metodologias e publicou relatórios apontando os maiores desafios em relação às capacidades das redes para 2030 e além.

Ao seguir o roteiro para o desenvolvimento do 6G, definido pela ITU-R, esperam-se pelas propostas de tecnologias para o IMT-2030 até o final de 2028, garantindo-se que as tecnologias que serão padronizadas em 2029 pela ITU-R satisfaçam padrões internacionais, de forma a garantir a interoperabilidade global das redes 6G.

Assim, através de várias frentes, a ITU fornece a orientação para outras organizações de padronização, como o 3GPP² e o ETSI³, no desenvolvimento das tecnologias necessárias para o desenvolvimento das redes 6G.

IV. EXEMPLO DE TECNOLOGIA HABILITADORA

Um dos exemplos de tecnologias que impulsionará o desenvolvimento do 6G é a exploração de faixas de frequências sub-THz, isto é, de 92-300 GHz, com capacidade de alcançar picos de taxas de transmissão maiores do que 100 Gbps, o que é crucial para casos de uso de nicho. Isso pavimentará o desenvolvimento da tecnologia necessária para vencer os desafios do alcance da cobertura e das dificuldades na miniaturização dos componentes de *hardware* para as comunicações em Terahertz (THz). Além disso, é necessário promover campanhas de medições para a caracterização dos canais e sua modelagem diante de um ambiente de transmissão em THz, cujos efeitos ainda não são totalmente compreendidos, o que é fundamental para o desenvolvimento de novos modelos de propagação de ondas para frequências em THz [7]–[10], [12].

O desenvolvimento das redes 6G pode explorar o potencial da Inteligência Artificial (IA) no projeto da interface aérea. Um exemplo de aplicação é a modelagem dos canais de comunicações em THz, onde os vários parâmetros de propagação podem ser aprendidos pela IA e utilizados na predição da propagação, fornecendo subsídios para alocação sub-ótima de antenas, estações rádio-base, largura de banda, potência, dentre outros [13].

Dessa forma, por meio da estrutura do IMT-2030, a ITU-R está orientando a pesquisa e o desenvolvimento da próxima geração das redes de comunicações móveis, promovendo o

atendimento das demandas futuras da sociedade, através da construção e evolução de um ecossistema unificado de serviços avançados com alcance global.

V. CONCLUSÕES

Esse artigo apresentou os casos de uso e as capacidades das redes 6G, de acordo com a estrutura definida pela ITU-R. Também discutiu alguns dos desafios no desenvolvimento da interface aérea para as redes 6G e elencou exemplos de tecnologias habilitadoras necessárias, destacando o papel que a IA pode ter para tornar as redes 6G uma realidade. Dessa forma, esse trabalho contribui para desmistificar as características principais da próxima geração das redes de comunicações móveis, 6G.

REFERENCES

- [1] P. Ahokangas, A. Aagaard “The Changing World of Mobile Communications: 5G, 6G and the Future of Digital Services,” Palgrave Macmillan, 2024, doi.org/10.1007/978-3-031-33191-6.
- [2] R. Liu, H. Lin, H. Lee, F. Chaves, H. Lim and J. Sködl, “Beginning of the Journey Toward 6G: Vision and Framework,” in IEEE Communications Magazine, vol. 61, no. 10, pp. 8-9, October 2023, doi: 10.1109/MCOM.2023.10298069.
- [3] J.T.J. Penttinen, “On 6G Visions and Requirements,” in Journal of ICT Standardization , vol. 9, no. 3, pp. 311-325, 2021, doi: 10.13052/jicts2245-800X.931.
- [4] R. Bassoli, F.H.P. Fitzek, E. C. Strinati , “Why Do We Need 6G?,” in ITU Journal on Future and Evolving Technologies, v. 2, n. 6, p. 1-31, 2021.
- [5] M. Alazab et al., “Digital Twins for Healthcare 4.0—Recent Advances, Architecture, and Open Challenges,” in IEEE Consumer Electronics Magazine, vol. 12, no. 6, pp. 29-37, 1 Nov. 2023, doi: 10.1109/MCE.2022.3208986.
- [6] S. D. Okegbile, J. Cai, D. Niyato , C. Yi , “Human Digital Twin for Personalized Healthcare: Vision, Architecture and Future Directions,” in IEEE Network, vol. 37, no. 2, pp. 262-269, March/April 2023, doi: 10.1109/MNET.118.2200071.
- [7] J.R. Bhat e S.A. Alqahtani, “6G Ecosystem: Current Status and Future Perspective,” in IEEE Access, vol. 9, pp. 43134-43167, 2021, doi: 10.1109/ACCESS.2021.3054833.
- [8] W. Jiang, B. Han, M. A. Habibi, H. D. Schotten , “The Road Towards 6G: A Comprehensive Survey,” in IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366, 2021, doi: 10.1109/OJCOMS.2021.3057679.
- [9] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, F. Tufvesson , “6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities,” in Proceedings of the IEEE, vol. 109, no. 7, pp. 1166-1199, July 2021, doi: 10.1109/JPROC.2021.3061701.
- [10] W. Jiang , H. D. Schotten, “The KICK-OFF of 6G Research Worldwide: An Overview,” in 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 2021, pp. 2274-2279, doi: 10.1109/ICCC54389.2021.9674614.
- [11] S. Yrjölä, P. Ahokangas, M. Matinmikko-Blue , “Visions for 6G Futures: a Causal Layered Analysis,” in 2022 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Grenoble, France, 2022, pp. 535-540, doi: 10.1109/Eu-CNC/6GSummit54941.2022.9815809.
- [12] E. Semaan, e. Tejedor, R. K. Kochhar, S. Magnusson, S. Parkvall, “6G Spectrum — Enabling the Future Mobile Life Beyond 2030,” in Ericsson White Paper, GFTL-23:000243 Uen, 2024.
- [13] F. Qamar, M. U. A. Siddiqui, MHD N. Hindia, R. Hassan, Q. N. Nguyen , “Issues, Challenges, and Research Trends in Spectrum Management: A Comprehensive Overview and New Vision for Designing 6G Networks,” in Electronics (Switzerland). 2020 Sept;9(9):1-39. 1416. doi: 10.3390/electronics9091416.

²3rd Generation Partnership Project.

³European Telecommunications Standards Institute.

Análise da Proteção de Serviços URLLC através de RL em Diferentes Cenários da Indústria 4.0

Weskley Mauricio, Cleverson Nahum, Francisco Parente, Maykon Silva e Aldebaro Klautau

Resumo—Neste trabalho, propomos um escalonador de recursos de rádio baseado em aprendizado por reforço multi-agente com as funções de escalonador *inter-slice* e *intra-slice* para atender aos *slices* de eMBB, mMTC e URLLC. Demonstramos que o escalonador proposto obtém melhor desempenho na redução de violações e proteção do *slice* prioritário em relação ao *baseline* em diferentes cenários.

Palavras-Chave—RL, RRS, Indústria 4.0, Network Slicing.

I. INTRODUÇÃO

A fim de contornar os desafios encontrados em *Network Slicing* (NS) e melhorar a *Quality of Service* (QoS), alguns trabalhos têm usado aprendizagem de máquina com agentes de *Reinforcement Learning* (RL). Em [1], por exemplo, é proposto um problema de *Multi-Agent Reinforcement Learning* (MARL) com o objetivo de gerenciar os recursos entre *slices* em tempo real. Apesar de o uso de agentes de RL escalonar os recursos de rede para atender a requisitos rigorosos de variados casos de uso em quinta geração (5G), muitas abordagens propostas sofrem com o provisionamento excessivo ou violações frequentes do *Service Level Agreements* (SLA), pois não garantem a proteção de *slices* de rede que são essenciais para *Radio Resource Scheduling* (RRS) em NS da *Radio Access Network* (RAN) [2]. Mais recentemente, Nahum *et al.* [3] introduzem um agente RL de RRS para NS que orquestra e gera menos violações de *slice*, protegendo o *slice Ultra Reliable Low Latency Communications* (URLLC). Essa abordagem em cenários da Indústria 4.0 é uma demanda essencial para o RRS devido as intensas variações do canal e ao número excessivo de dispositivos conectados à rede.

II. MODELO DE SISTEMA E MÉTODO PROPOSTO

Neste trabalho, consideramos o enlace *downlink* de um sistema Múltipla Entrada Múltipla Saída (MIMO) industrial *indoor* com tecnologia *Frequency Division Duplex* (FDD) *Orthogonal Frequency Division Multiple Access* (OFDMA), como em um galpão de fábrica. O sistema é composto por

uma *Base Station* (BS) equipada com um modelo de antena 3D $n_v \times n_h$, servindo U *User Equipments* (UEs), cada um dos quais equipado com antenas únicas omnidirecionais, onde n_v e n_h são o número de antenas verticais e horizontais, respectivamente. Identificamos três tipos de *slices* distintos, cada um adaptado para casos de uso específicos, com requisitos únicos de atraso e taxas de dados alcançáveis. Assumimos também que cada UE pertence exclusivamente a um desses tipos de *slices* e consideramos R *Resource Blocks* (RBs) disponíveis para cada *Transmission Time Interval* (TTI), em que um RB representa a menor unidade de recursos que pode ser alocada para uma UE. Cada RB é subdividida em N_{sc} subportadoras OFDMA adjacentes e N_{symb} símbolos *Orthogonal Frequency Division Multiplexing* (OFDM) consecutivos.

Utilizamos um cenário especificamente feito para aplicações industriais *indoor*, tais como galpões de fábricas, no contexto de ambientes da Indústria 4.0. Para isso, consideramos o modelo de canal *Quasi Deterministic Radio Channel Generator* (QuaDRiGa) com *Line Of Sight* (LOS) e *Non-Line Of Sight* (NLOS), conforme documentado na referência [4].

O escalonador MARL ora proposto utiliza o método *Soft Actor-Critic* (SAC) para desempenhar as funções de escalonamento *inter-slice* e *intra-slice*. Para dar suporte à utilização de RL nessas duas funções, nós estendemos o agente introduzido em [3], melhorando a proteção aos *slices* críticos em cenários industriais, via priorização de alocação, quando a quantidade de recursos não é suficiente para atender a todos os requisitos.

Os espaços de observação representam as entradas dos métodos para os agentes de *inter-slice* e *intra-slice* com composição diferente para cada tipo de agente de escalonamento. O espaço de observação do agente de *inter-slice* no passo de simulação n é definido como $O_n^{\text{inter}} = [q_{\text{embb}}, q_{\text{urllc}}, q_{\text{mmtc}}, s_{\text{embb}}, s_{\text{urllc}}, s_{\text{mmtc}}]$, em que $q_s = [r_s^{\text{req}}, d_s^{\text{req}}, U_s]$ representa os requisitos do *slice* para a taxa de transferência, a latência e a quantidade de UEs associadas ao *slice*. As métricas de taxa de transferência, ocupação de *buffer* e atraso no *buffer* para cada *slice* são representadas em $s_s = [r_s(n), b_s^{\text{occ}}(n), d_s(n)]$. Essas métricas de *slice* são os valores médios dos valores obtidos dos UEs associados ao *slice* s . O espaço de observação para o agente de *intra-slice* no passo de simulação n é $O_{n,s}^{\text{intra}} = [R_s(n), q_s, s_s]$, em que, além dos requisitos do *slice* analisado e das suas métricas, também é incluído a quantidade de RBs alocados pelo agente de *inter-slice* para o agente de *intra-slice* considerado.

O espaço de ação do agente de *inter-slice* no passo de simulação n é definido de maneira similar a [3] como $A_n^{\text{inter}} = [a_{\text{embb}}^{\text{inter}}, a_{\text{urllc}}^{\text{inter}}, a_{\text{mmtc}}^{\text{inter}}]$, em que a_s representa um fator de ação

Weskley Mauricio, Francisco Parente e Maykon Silva fazem parte da Fundação Centro de Pesquisa e Desenvolvimento em Telecomunicações (CPQD), Campinas-SP, Brasil. Cleverson Nahum e Aldebaro Klautau fazem parte da Universidade Federal do Pará (UFPA). E-mail: {wmauricio, fparente, mpsilva}@cpqd.com.br, {cleverson, aldebaro}@ufpa.br. Este trabalho foi financiado parcialmente pela RNP, com recursos do MCTIC, Outorga n° 01245.010604/2020-14, no âmbito do Projeto Brasil 6G do Instituto Nacional de Telecomunicações (Inatel), Brasil; Projeto Universal (CNPq grant 405111/2021-5) e Projeto Smart 5G Core And MultiRAN Integration (SAMURAI) (MCTIC/CGI.br/FAPESP under Grant 2020/05127-2) e, pelo projeto ForELERAN apoiado pelo Centro de Excelência em Redes Abertas EMBRAPII/CPQD, com recursos financeiros do PPI IoT/Manufatura 4.0/PPI HardwareBR do edital MCTI número 51/2023, firmado com a EMBRAPII.

para o *slice* s com valor no intervalo $[-1, 1]$. A ação escolhida pelo agente A_n é mapeada para o número de RBs a serem alocados para cada *slice* usando [3, Eq. (11)]. A ação gerada pelo agente de *inter-slice* define quantos RBs serão utilizados por cada *slice*. O agente de *intra-slice* tem o seu espaço de ação representado por uma única variável a_s^{inter} com valores inteiros no intervalo de $[0, 2]$, onde cada valor representa uma escolha de algoritmo dentre as opções *round-robin*, *maximum-throughput* e *proportional-fair* para o *intra-slice*.

A função de recompensa $W(n)$ considera a distância para cumprir os requisitos de cada *slice* na sua construção em conjunto com o uso de mecanismo de proteção ao *slice* prioritário. As recompensas dos *slices* de *enhanced Mobile Broadband* (eMBB) e URLLC são contabilizados apenas quando os requisitos do *slice* de URLLC são cumpridos. A função de recompensa do agente de *inter-slice* é [3, Eq. (12)]

$$W^{\text{inter}}(n) = \begin{cases} W_{\text{embb}}(n) + W_{\text{mmtc}}(n), & \text{se } W_{\text{urllc}}(n) = 0 \\ W_{\text{urllc}}(n), & \text{caso contrário} \end{cases}, \quad (1)$$

em que $W_{\text{embb}}(n)$, $W_{\text{urllc}}(n)$ e $W_{\text{mmtc}}(n)$ são dados em [3, Eqs. (13)–(15)] e representam a função de recompensa para os *slices* de eMBB, URLLC e *Massive Machine Type Communication* (mMTC) no passo de simulação n . Dessa forma, o agente de *inter-slice* têm como prioridade atender os requisitos do *slice* prioritário de URLLC. Já a função de recompensa do agente de *intra-slice* é

$$W_s^{\text{intra}}(n) = W_s^r(n) + W_s^d(n), \quad (2)$$

em que $W_s^r(n)$ e $W_s^d(n)$ são dadas em [3, Eqs. (16) e (17)]. As distâncias para atingir os requisitos da taxa de transferência e o atraso do *buffer* são contabilizados pelo agente que busca minimizar essas distâncias através da seleção do melhor algoritmo entre as opções *round-robin*, *maximum-throughput*, e *proportional-fair* para cada passo de simulação n . No caso do *slice* de mMTC, o valor da contribuição da taxa de transferência no $W_s^r(n)$ é sempre considerado zero dado que o *slice* de mMTC não possui requisitos de taxa de transferência.

III. RESULTADOS

Neste artigo, consideramos uma BS com uma antena no canto superior de uma fábrica, atendendo a UEs em três casos de uso: eMBB, URLLC e mMTC. As medidas foram extraídas dos Halls 5.3 e 5.6 da fábrica de Nuremberg (Alemanha) conforme [4]. A BS atende 100 UEs distribuídos uniformemente, sendo 30 para eMBB, 30 para URLLC e 40 para mMTC. O modelo de canal industrial possui 100 RBs, distribuídos entre fatias e UEs usando escalonamento inter e intra-fatias. O método de baseline SLA *Satisfaction Rate* (SSR) para comparação está descrito em [3]. Os parâmetros da simulação do canal são: largura de banda de 100 MHz, TTI de 1 ms, 1000 TTIs, potência de transmissão de 35 dBm, UEs a 3 km/h, 3 fatias de rede, vazão requerida de $r_{\text{embb}}=20$ Mbps e $r_{\text{embb}}=5$ Mbps, e latência requerida de $d_{\text{embb}}=30=30$, $d_{\text{urllc}}=1$ e $d_{\text{mmtc}}=50$ ms. Realizamos 100 rodadas de simulação com o QuaDRiGa, divididas em 70 para treinamento e 30 para teste, com alocação de potência igualitária entre os RBs.

A Figura 1 mostra as médias de violações de requisitos totais e para o *slice* prioritário URLLC em 30 episódios

de teste para os Halls 5.3 e 5.6. No Hall 5.3, o método proposto apresenta menos violações tanto em todos os *slices* quanto no *slice* URLLC. No Hall 5.6, há mais violações em todos os *slices*, mas menos violações no *slice* URLLC. Isso ocorre porque o Hall 5.6 é mais desafiador, levando o método proposto a sacrificar o desempenho dos outros *slices* para melhorar o desempenho do URLLC. Por fim, os resultados mostram que a solução proposta consegue garantir uma baixa taxa de violação para o serviço URLLC em ambos os cenários, mostrando-se uma solução robusta e independente do cenário.

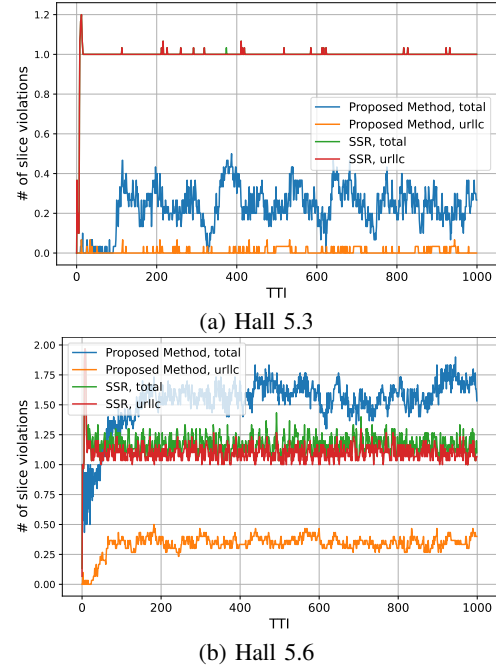


Figura 1: Violações de SLA nos cenários Hall 5.3 e 5.6 [4].

IV. CONCLUSÃO

Apresentamos um método de alocação de recursos de rádio usando MARL para escalonamento inter-slice e intra-slice em cenários de Indústria 4.0 em dois Halls distintos. O agente proposto atende aos requisitos dos *slices* combinando ações de escalonamento e priorizando os *slices* URLLC quando os recursos são insuficientes. Resultados preliminares mostram que o método proposto tem significativamente menos violações em comparação com o método SSR nos *slices* URLLC.

REFERÊNCIAS

- [1] Y. Shao *et al.*, “Graph attention network-based multi-agent reinforcement learning for slicing resource management in dense cellular network,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10792–10803, 2021.
- [2] M. Zangoeei, N. Saha, M. Golkarifard, and R. Boutaba, “Reinforcement learning for radio resource management in ran slicing: A survey,” *IEEE Communications Magazine*, vol. 61, no. 2, pp. 118–124, Feb. 2023.
- [3] C. Nahum, W. Mauricio, M. Silva, M. Takeda, and A. Klautau, “Safeguard urllc services in industry 4.0 through reinforcement learning scheduling,” in *2023 Workshop on Communication Networks and Power Systems (WCNPS)*. IEEE, 2023, pp. 1–7.
- [4] S. Jaeckel *et al.*, “Industrial indoor measurements from 2-6 ghz for the 3gpp-nr and quadriga channel model,” in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–7.

Análise de Desempenho dos Protocolos de Transporte em Redes 5G Open RAN

Weskley Mauricio, Francisco Hugo Costa Neto, Eduardo Melão e Maykon Silva

Resumo—Este artigo investiga o impacto dos modos de Controle de Enlace de Rádio UM e AM no desempenho dos protocolos de transporte TCP e UDP em redes de comunicação sem fio 5G que utilizam a arquitetura Open RAN. Por meio de testes realizados em *testbed* constituído por equipamentos comerciais, verifica-se que o modo UM apresenta os melhores níveis de *throughput* para os protocolos de transporte considerados, chegando a ganhos de até 20%.

Index Terms—Protocolos de Transporte, RLC, Open RAN.

I. INTRODUÇÃO

O *Open Radio Access Network* (RAN), ou Rede de Acesso de Rádio Aberta, é um paradigma de arquitetura de redes de comunicação baseado em componentes desagregados, virtualizados e baseados em software, conectados por meio de interfaces abertas e interoperáveis entre diferentes fornecedores [1].

A desagregação da RAN divide as funcionalidades da estação base de nó de próxima geração, *Next Generation Node Base* (gNB), em diferentes unidades denominadas *Centralized Unit* (CU), *Distributed Unit* (DU), e *Radio Unit* (RU). Essa divisão lógica permite a implantação de diferentes funcionalidades em diversos locais da rede, bem como em diferentes plataformas de hardware.

As camadas superiores do protocolo de rádio, isto é *Radio Resource Control* (RRC), *Packet Data Convergence Protocol* (PDCP) e o *Service Data Adaption Protocol* (SDAP), estão sob responsabilidade da CU [2]. A DU controla funções de controle de enlace do rádio, tais como detecção e correção de erros, retransmissão de pacotes perdidos e sequenciação de pacotes. Ela gerencia o protocolo *Radio Link Control* (RLC), sendo responsável pela segmentação de Unidades de Dados de Serviço, *Service Data Units* (SDUs), vindos da camada PDCP em Unidades de Dados de Protocolo, *Protocol Data Units* (PDUs), de tamanho adequado. Ela também lida com a retransmissão de PDUs recebidas erroneamente e com a remoção de PDUs duplicadas [3].

Dependendo do tipo de serviço, o RLC pode ser configurado em um dos três modos: *Transparent Mode* (TM), *Unacknowledged Mode* (UM) e *Acknowledged Mode* (AM). O RLC é completamente transparente no modo TM, isto é, ele é essencialmente ignorado, pois não há retransmissões,

nenhuma detecção de duplicatas e nenhuma segmentação. Esta configuração é usada para canais de transmissão do plano de controle. O modo UM suporta segmentação e detecção de duplicatas. Este modo é usado quando a entrega sem erros não é necessária, como por exemplo no serviço *Voice Over IP* (VoIP). Já o modo AM suporta segmentação, remoção de duplicatas e retransmissões de dados errôneos, sendo o principal modo de operação para os canais *Downlink Shared Channel* (DL-SCH) e *Uplink Shared Channel* (UL-SCH) [3].

Técnicas de retransmissão são necessárias para recuperar as PDUs perdidas. A retransmissão é feita em várias camadas da pilha de protocolo. Na camada *Medium Access Control* (MAC) é realizada por meio do *Hybrid Automatic Repeat Request* (HARQ). Na camada RLC, esse controle é realizado pelo *Automatic Repeat Request* (ARQ) e pelo protocolo RLC, com seus diferentes modos.

Este artigo analisa o impacto das diferentes configurações dos modos RLC no desempenho dos protocolos de transporte *Transmission Control Protocol* (TCP) e *User Datagram Protocol* (UDP) por meio dos mecanismos de retransmissão. Tal compreensão é vital para otimizar o desempenho das redes em termos de *throughput* e latência. A literatura apresenta vários estudos desse impacto em redes *Long Term Evolution* (LTE), como realizado em [4]. Entretanto, ainda são limitadas as avaliações dessa natureza em redes 5G implementadas considerando a arquitetura Open RAN.

II. CONFIGURAÇÃO DO TESTBED 5G Open RAN

O *testbed* utiliza uma versão *Commercial Off-The-Shelf* (COTS) de componentes Open RAN (DUs, CUs e RUs). As DUs, CUs utilizam a pilha comercial Open RAN Radisys que oferece diversas vantagens para modelar condições complexas do ecossistema [5]. Em primeiro lugar, permite aos investigadores replicar cenários realistas com testes que se assemelham às condições reais da rede, incluindo canais reais entre o *User Equipment* (UE) e a rede. Este foco em componentes comerciais aumenta a validade e a relevância dos testes, fornecendo informações valiosas sobre o desempenho do Open RAN e a otimização para implantações práticas.

O *testbed* possui as seguintes configurações: dois servidores estão interligados entre si, neste caso, a CU com a DU através da interface *midhaul*. Além disso, a DU também está ligada ao *switch fronthaul*, que atua como *Precision Time Protocol* (PTP) *Grandmaster*, com o objetivo de permitir a sincronização dos relógios entre a DU e RU, facilitando a comutação entre o plano de controle e plano de usuário da interface *Open Fronthaul*. Além disso, vale ressaltar que o

Weskley Mauricio, Francisco Hugo Costa Neto, Eduardo Melão e Maykon Silva fazem parte da Fundação Centro de Pesquisa e Desenvolvimento em Telecomunicações (CPQD), Campinas-SP, Brasil. e-mail: {wmauricio, fhugo, emelao, mpsilva}@cpqd.com.br. Este trabalho foi parcialmente financiado pelo projeto ForELERAN apoiado pelo Centro de Excelência em Redes Abertas EMBRAPII/CPQD, com recursos financeiros do PPI IoT/Manufatura 4.0/PPI HardwareBR do edital MCTI número 51/2023, firmado com a EMBRAPII.

testbed possui uma configuração de *Splits* funcionais seguindo o que está padronizado pela O-RAN Alliance, ou seja: o *Split 2* para o *midhaul*, neste caso, os protocolos RRC e PDCP ficam localizados na CU, enquanto que na DU estão os protocolos RLC, MAC e por fim, a *high* PHY; e o *Split 7.2x* para o *fronthaul* com o objetivo principal de dividir os blocos da camada PHY, onde a *low* PHY fica localizada na RU [2], como mostra a Figura 1.

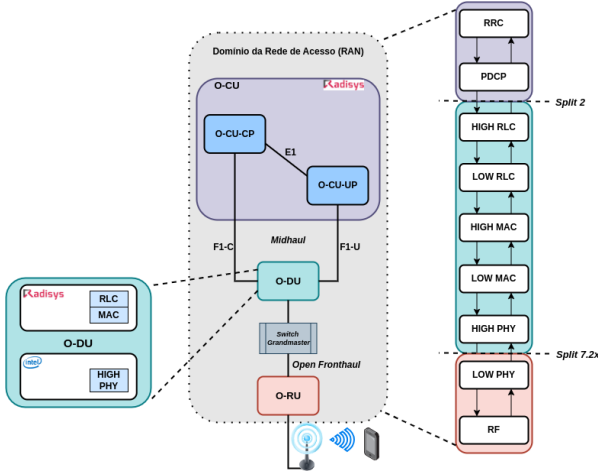


Figura 1. Arquitetura do *testbed* Open RAN.

III. RESULTADOS

Neste artigo, utiliza-se o *iperf3* para medir a taxa de dados de *downlink* entre o UE 5G NR comercial e o Core Network (CN) com duas condições de *Reference Signal Receive Power* (RSRP) distintas, uma condição boa com média de RSRP igual a -76 dBm e uma condição de RSRP ruim com média de RSRP igual a -115 dBm. O RSRP estimado entre o UE e a RU foram coletados através da análise dos *logs* de *dumpstate* do UE, obtido através da ferramenta Logcat. Para significância estatística, realiza-se em cada ponto 5 medidas independentes com duração de 40 segundos cada e os valores de média e variância foram calculados considerando um intervalo de confiança de 95%. Utiliza-se uma RU Foxconn RPQN 7801e operando em *Time Division Duplex* (TDD) com banda *n78* (3, 75 GHz). A largura de banda da portadora ocupa 100 MHz, usando 30 kHz de espaçamento de subportadora e com uma potência de transmissão de 24 dBm.

A Fig. 2 mostra o *throughput* médio para os protocolos TCP e UDP com os modos RLC AM e UM considerando diferentes níveis de RSRP. À medida que os níveis de RSRP são reduzidos, tanto TCP quanto UDP sofrem uma redução significativa no *throughput* em ambos os modos RLC. No cenário de RSRP ruim, consegue-se verificar que o RLC UM alcança um desempenho superior ao RLC AM em ambos os protocolos de transporte TCP e UDP, chegando a ganhos de até 20% em termos de *throughput*. Enquanto no cenário de RSRP bom, ambos o RLC AM e UM atingem níveis de *throughput* semelhantes quando considera-se os mesmos modos de transmissão. Este comportamento acontece devido a quantidades de retransmissões necessárias no modo RLC AM que

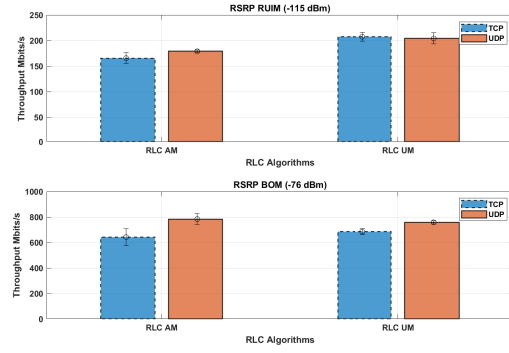


Figura 2. Throughput médio para diferentes configurações de RLC.

é inversamente proporcional ao nível de RSRP, degradando assim o *throughput* alcançado. Comparando os protocolos TCP e UDP em ambos os níveis de RSRP, consegue-se verificar um ganho de desempenho de *throughput* quando o UDP é utilizado, como observado anteriormente, também observa-se ganhos de até 20%. O maior ganho observado acontece quando se compara TCP e UDP no cenário utilizando RLC AM, isto acontece devido a utilização de duas camadas de retransmissão (RLC AM e TCP) degradando assim consideravelmente o *throughput*. Portanto, a partir deste conjunto de resultados, consegue-se observar que as escolhas dos modos RLC e protocolos de transporte impactam diretamente o *throughput* alcançado. Pois eles sobrecarregam o sistema através da inclusão de níveis retransmissão nas diferentes camadas, trazendo assim uma troca entre confiabilidade e desempenho. Além disso, o conjunto que alcançou melhor desempenho de *throughput* foi o RLC UM utilizando UDP, pois obteve o melhor desempenho independente do RSRP considerado.

IV. CONCLUSÃO

Este artigo investigou por meio de testes em equipamentos comerciais o impacto dos modos RLC no desempenho dos protocolos de transporte TCP e UDP em diferentes condições de enlace. Os resultados indicam que o modo UM apresenta os maiores níveis de *throughput*, uma vez que não há retransmissão. Ao reduzir os mecanismos de recuperação de dados, compromete-se a confiabilidade na transmissão de pacotes, o que pode prejudicar a experiência do usuário.

REFERÊNCIAS

- [1] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward Next Generation Open Radio Access Networks: What O-RAN Can and Cannot Do!" *IEEE Network*, vol. 36, no. 6, pp. 206–213, 2022.
- [2] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.
- [3] 3GPP, "TS 38.322 v17.4.0," *Radio Link Control (RLC) Protocol Specification*, 2024.
- [4] R. Li, M. Shariat, and M. Nekovee, "Transport Protocols Behaviour Study in Evolving Mobile Networks," in *International Symposium on Wireless Communication Systems*, 2016, pp. 456–460.
- [5] Radisys, "Radisys connect: Open and disaggregated access products," 2024. [Online]. Available: <https://www.radisys.com/connect>

Comparativo de Eficiência Energética de Plataformas x86 e ARM para Estações Rádio-Base 5G

Mateus Mattos, Elielder Melo, Luiz Henrique Silva e Eduardo Melão

Abstract—Este artigo compara o consumo energético de processadores x86 e ARM executando uma pilha de protocolos de uma estação rádio-base 5G Open RAN de código aberto com um terminal real (*smartphone*) conectado à rede. Por meio dos testes realizados utilizando duas plataformas de processamento de alto desempenho com arquiteturas x86 e ARM rodando o software 5G, verifica-se uma melhor eficiência energética associada à plataforma ARM, com quase o dobro de Mbps por Watt no cenário UL.

Keywords—5G, ARM, x86, GPP, Eficiência Energética

I. INTRODUÇÃO

A complexidade da implementação de pilhas de protocolos 5G Open RAN exige arquiteturas de processadores de uso genérico (*General Purpose Processors* - GPPs) capazes de atender a requisitos rigorosos e otimizar recursos. Duas plataformas dominantes, x86 e ARM (*Advanced RISC Machine*), oferecem desempenho para executar tarefas computacionais complexas e se colocam em contraste, dada a crescente demanda por eficiência energética em função do desempenho e do consumo durante a execução de sistemas de alta capacidade computacional, chamados HPC (*High Performance Computing*).

Em sistemas HPC, a plataforma x86 é dominante, pela preparação de sua arquitetura para executar tarefas complexas. Com a evolução da arquitetura em plataformas ARM, ela também oferece bom desempenho, se destaca pelo baixo custo agregado de operação e hoje se coloca como uma opção viável para sistemas HPC e, consequentemente, para redes 5G [1].

A distinção entre as plataformas x86 e ARM inicia-se com suas arquiteturas, a plataforma x86 adota a arquitetura CISC (*Complex Instruction Set Computing*), caracterizada pela execução de instruções complexas que exigem mais recursos de processamento. Em contraste, a plataforma ARM utiliza a arquitetura RISC (*Reduced Instruction Set Computing*), que é definida por instruções mais simples e pela menor quantidade de ciclos de máquina necessários para execução de cada instrução. Em sistemas computacionais complexos plataformas construídas com base em arquitetura CISC tendem a ser mais efetivas ao executar tarefas com menor consumo de memória,

por outro lado plataformas baseadas em arquitetura RISC exigem um gerenciamento de memória de maior robustez. Em termos de eficiência energética, por sua vez, a plataforma ARM destaca-se pela construção agregada de baixa dissipação térmica e menor número de operações para execução de uma tarefa, tendo um melhor resultado energético na relação desempenho por consumo de energia. [2]

Algumas soluções de código aberto implementam a pilha de protocolos 5G e podem ser executadas em GPPs x86 e ARM, como o OpenAirInterface [3] e o srsRAN Project [4]. Ambas podem ser implantadas com o *split* 8 utilizando SDRs (*Software Defined Radios*) ou por meio de RUs (*Radio Units*) com o *split* 7.2x Open RAN, também chamado de O-FH (*Open Fronthaul*).

Analisar a eficiência energética de uma implantação de redes móveis como o 5G é importante, principalmente, em cenários de conectividade rural, muito comuns no Brasil. Nesses cenários, o fornecimento de energia elétrica é deficitário e a solução por vezes é a implantação de fontes renováveis de fornecimento de energia, como painéis solares [5]. Economizar energia em tais implementações não apenas reduz os custos operacionais, mas também melhora a sustentabilidade e viabilidade de projetos nessas regiões.

Este artigo compara o consumo energético de uma pilha de protocolos 5G de código aberto rodando em GPPs x86 e ARM. Com isso, o artigo contribui para identificar opções que possam otimizar o uso de energia, reduzindo a dependência de fontes intermitentes e tornando viável o uso de fontes renováveis. A literatura apresenta vários estudos comparativos dessas arquiteturas de processador, mas ainda são poucos os trabalhos que avaliam os impactos na execução de tarefas relacionadas a protocolos de redes móveis.

II. DESCRIÇÃO DO AMBIENTE DE TESTES

O ambiente de testes montado conta com dois servidores: o primeiro possui um processador Intel® Xeon® Gold 5320 (x86) e 128GB de memória RAM; já o outro possui um processador Ampere® Altra® Q80-30 (ARM) e 256GB de memória RAM. Maiores detalhes sobre cada processador podem ser vistos na Tabela I.

Foi utilizada a pilha de protocolos de código aberto srsRAN como CU (*Centralized Unit*) e DU (*Distributed Unit*), na versão de junho de 2024 (*commit* a804c5), configurada em TDD, padrão 7DS2U, na banda 78, com 40MHz de largura de

TABELA I
DADOS DOS PROCESSADORES

	ARM	x86
Fabricante	Ampere	Intel
Modelo	Altra Q80-30	Xeon Gold 5320
Frequência de clock (GHz)	3	2,2
Cores Threads	80 80	26 52
Sockets	1	1
Cache L1 / Core (KiB)	5000	832
Cache L2 (MiB)	80	32,5
Cache L3 (MiB)	0	39

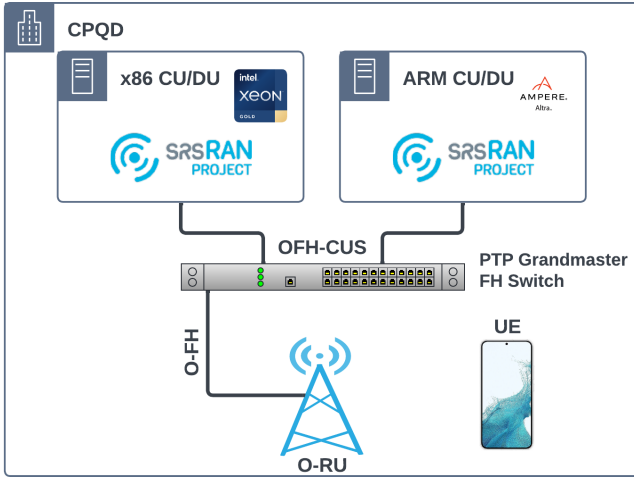


Fig. 1

DIAGRAMA DA CONFIGURAÇÃO MONTADA PARA OS TESTES.

banda e MIMO 4x1. Nas duas plataformas de processamento, a aplicação da rede 5G é executada sobre o sistema operacional Ubuntu 22.04.4 LTS, kernel 5.15.0-113-generic, integrada a uma RU Open RAN Foxconn RPQN 7801E conectada aos servidores por um switch de fronthaul pela interface O-FH CUS (Control, User and Synchronization). Para conexão móvel, foi utilizado um terminal (User Equipment) um Samsung Galaxy S23 Ultra. Um esquemático do ambiente de testes pode ser visto na Figura 1.

III. RESULTADOS

Com o ambiente de testes mencionado na seção II, foram realizados os seguintes testes:

- 1) Processadores em *idle* (ou seja, somente com a execução do sistema operacional);
- 2) Pilha de protocolos 5G sendo executada, porém sem nenhum terminal conectado;
- 3) Pilha de protocolos 5G sendo executada com um terminal conectado e tráfego de 300 Mbps UDP Downlink (DL);
- 4) Pilha de protocolos 5G sendo executada com um terminal conectado e tráfego de 30 Mbps UDP Uplink (UL);

Os resultados podem ser observados na Figura 2. Além disso, pode-se também comparar a eficiência energética (Mbps/W) para ambos processadores, apresentada na Tabela II (sendo o teste DL referente ao teste de número 3 e UL o teste de número 4).

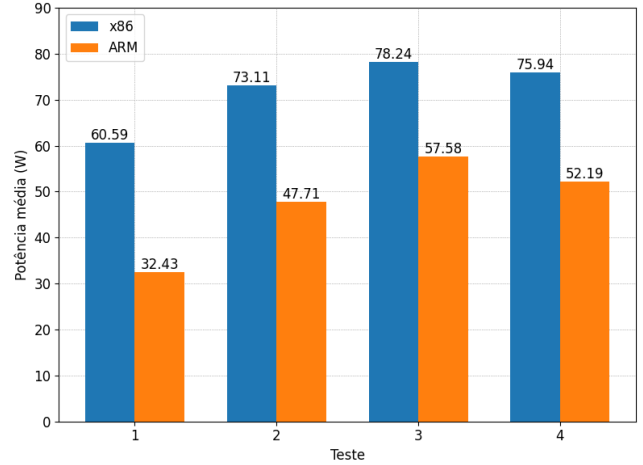


Fig. 2

COMPARAÇÃO DO CONSUMO ENERGÉTICO ENTRE GPPs x86 E ARM.

TABELA II
Throughput MÉDIO PARA CADA TESTE

Teste	Throughput Médio (Mbps)		Eficiência Energética (Mbps/W)	
	ARM	x86	ARM	x86
DL	285,7	282,4	4,96	3,61
UL	25,32	20,17	0,49	0,27

IV. CONCLUSÃO

Conforme os resultados apresentados na seção III, nota-se que o processador ARM possui um consumo energético inferior que o x86 mesmo possuindo mais núcleos de processamento e um clock maior. Com isso, sua eficiência energética, quando levado em consideração o tráfego de dados por Watt, consegue ser quase o dobro no caso do UL, como mostrado na Tabela II. É importante salientar, no entanto, que o aumento do consumo (com relação ao estado *idle*) ao executar testes com tráfego DL nos processadores ARM chega a ser quase o dobro quando comparado ao x86 (9,87 e 5,13 Watts, respectivamente), mas mantendo basicamente o mesmo desempenho. O mesmo ocorre no UL (4,48 Watts para o ARM e 2,83 Watts para o x86). Esses fatos abrem novas lacunas que serão abordadas em futuros estudos.

REFERÊNCIAS

- [1] D. Kumar, M. A. Rajput, P. Kumar, S. Ahmed, M. S. Bhatti, and A. Tsetse, "Performance evaluation of ARM-based versus x86-based processors in high performance computing clusters," *JISR-C*, vol. 21, pp. 32–41, 2023.
- [2] T. S. Khushi Gupta, "Changing trends in computer architecture : A comprehensive analysis of ARM and x86 processors," *IJS CSEIT*, vol. 7, pp. 619–631, 2021.
- [3] O. S. Alliance, "The OpenAirInterface Initiatives," 2014. [Online]. Available: <https://openairinterface.org>
- [4] S. R. Systems, "srsRAN project: Open source 5G software suite from software radio systems," 2023. [Online]. Available: <https://www.srsran.com>
- [5] A. M. Cavalcante, M. V. Marquezini, L. Mendes, and C. S. Moreno, "5g for remote areas: Challenges, opportunities and business modeling for brazil," *IEEE Access*, vol. 9, pp. 10829–10843, 2021.

Decodificação de Códigos Polares Usando Cancelamento Sucessivo: Abordagem Matricial versus Abordagem Vetorial

1st João Carlos Pinheiro Cavalcante
Centro de Energias Alternativas e Renováveis (CEAR)
Universidade Federal da Paraíba (UFPB)
João Pessoa, Brasil
joao.cavalcante@estudante.cear.ufpb.br

2nd Ana Clara Vilar Bandeira
Centro de Energias Alternativas e Renováveis (CEAR)
Universidade Federal da Paraíba (UFPB)
João Pessoa, Brasil
ana.bandeira@estudante.cear.ufpb.br

3rd Gláuber Vítor da Cruz Ribeiro
Centro de Energias Alternativas e Renováveis (CEAR)
Universidade Federal da Paraíba (UFPB)
João Pessoa, Brasil
glauber.ribeiro@estudante.cear.ufpb.br

4th Waslon Terllizzie Araújo Lopes
Centro de Energias Alternativas e Renováveis (CEAR)
Universidade Federal da Paraíba (UFPB)
João Pessoa, Brasil
waslon@cear.ufpb.br

Resumo—Os códigos polares, propostos por Arikan em 2009, utiliza o conceito de polarização de canal vetorial para gerar canais muito confiáveis para transmissão da informação e são usados com técnica de codificação para o canal de controle do padrão de telefonia móvel celular de quinta geração (5G). A forma tradicional de decodificação destes códigos é o cancelamento sucessivo de interferência em que os bits são decodificados de forma recursiva através do uso de uma árvore binária. Este artigo compara duas estratégias de implementação do processo de decodificação para dois tipos de representação da árvore binária: abordagem vetorial e matricial. Resultados mostram que a implementação baseada em vetores apresenta um tempo de execução menor, com eficiência de memória comparável à implementação com matrizes. Essa vantagem é atribuída à estrutura simplificada e à menor sobrecarga computacional dos vetores em comparação com as matrizes bidimensionais.

Palavras-chave—Códigos Polares, 5G, Decodificação por Cancelamento Sucessivo, Complexidade.

I. INTRODUÇÃO

A evolução das redes sem fio destaca avanços significativos na capacidade de transmissão, redução da latência e aumento da eficiência energética. As redes de quinta geração (5G) surgem como uma solução promissora, atendendo às crescentes demandas por taxa de dados e conectividade mais robusta [1]. Contudo, um dos principais desafios na implementação do 5G é garantir a confiabilidade e a eficiência da transmissão de dados em ambientes com alta interferência. Por isso, a utilização de códigos de correção de erros eficientes é crucial para assegurar a integridade dos dados transmitidos [2].

Entre os diversos esquemas de correção de erros, os códigos polares se destacam por seu potencial de alcançar a capacidade do canal com complexidade gerenciável [3]. Assim, este trabalho é voltado à implementação dos códigos polares em um canal de comunicação, comparando a eficiência e

complexidade computacional de duas abordagens: vetores e matrizes.

O restante deste trabalho encontra-se organizado da seguinte forma: os fundamentos de códigos polares são apresentados na Seção II. A implementação da abordagem vetorial a partir da abordagem matricial da decodificação é discutida na Seção III. Os resultados de simulação são apresentados na Seção IV enquanto que a Seção V é voltada às conclusões do trabalho.

II. CÓDIGOS POLARES

Os códigos polares são uma inovação significativa no campo da correção de erros de transmissão e baseiam-se no conceito de polarização de canais, em que um conjunto de canais de comunicação é transformado em um novo conjunto com canais completamente confiáveis e outros completamente não confiáveis. A informação é transmitida pelos canais confiáveis enquanto que os canais não confiáveis são usados como redundância. Neste contexto, o processo de decodificação é realizado por meio de algoritmos de decodificação com cancelamento sucessivo de interferência (SC – *Successive Cancellation* [4].

O algoritmo SC funciona processando os bits um a um, utilizando as probabilidades condicionais calculadas a partir dos bits já decodificados e da estrutura polarizada dos canais [5].

III. IMPLEMENTAÇÕES MATRICIAL E VETORIAL

A análise baseada nas relações recursivas durante a decodificação SC pode ser representada em uma árvore binária, conforme ilustra a Figura 1. Essa árvore binária contém importantes posições utilizadas no processo de decodificação sinal recebido. Para abranger todas as posições da árvore binária, foram analisados duas implementações: uma baseada em matrizes e outra baseada em vetores. Na conversão

entre elementos entre as duas abordagens, foi utilizada a seguinte expressão

$$\text{Pos} = 2^{\text{depth}} + \text{node} - 1, \quad (1)$$

em que Pos é a posição que está sendo manipulada no vetor; depth é a profundidade da árvore binária e node é posição do nó desejado presentes na profundidade analisada.

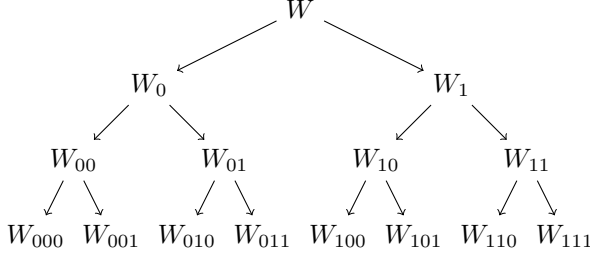


Figura 1: Árvore binária de decodificação dos códigos polares.

IV. RESULTADOS

Para a obtenção dos resultados, foram implementadas rotinas em C++ para decodificação de códigos polares baseada em SC. Duas abordagens (matrizes e vetores) foram usadas na transmissão de sinais através de canais contaminados com ruído aditivo gaussiano e relação sinal-ruído (SNR – *Signal-to-Noise Ratio*) variando de zero a 4 dB. A análise foi feita considerando tempo de execução de, no mínimo, 1000 transmissões para cada valor de SNR. Uma maior complexidade computacional tende a possuir tempo de execuções maiores devido à quantidade de operações a serem realizadas, o que leva a maiores tempos de execução. A Figura 2 mostra o tempo de execução em função do tamanho do código. Observa-se que, quanto maior o tamanho do código, maior a diferença em tempo de execução da abordagem matricial em relação à abordagem vetorial. Por exemplo, considerando um tamanho de código de 1024 bits, a diferença entre as duas abordagens é de aproximadamente 40 segundos.

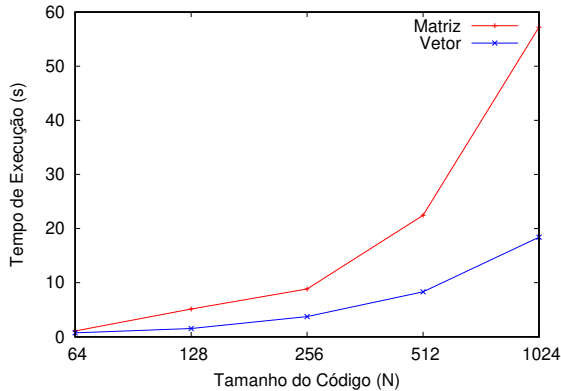


Figura 2: Tempo de execução em função do tamanho do código polar.

Outro fator crucial analisado foi a quantidade de memória alocada durante as operações dos códigos polares. A eficiência

no uso da memória é fundamental, uma vez que, em sistemas de comunicação e processamento de dados, os recursos são limitados e otimizar o desempenho é essencial, uma vez que é necessário reduzir os custos operacionais como também melhorar a confiabilidade do sistema [6]. A Figura 3 mostra que a quantidade de memória alocada nas duas abordagens é praticamente igual, conforme esperado.

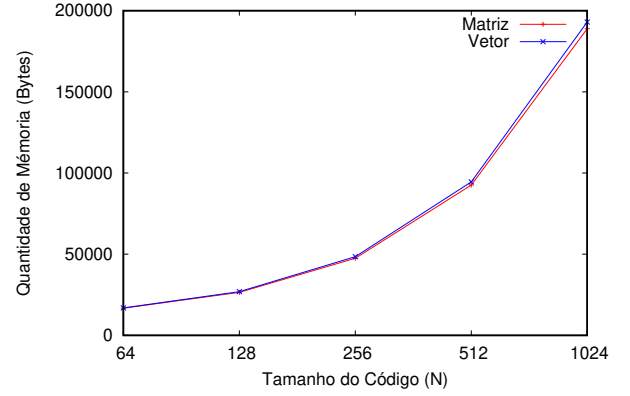


Figura 3: Quantidade de memória alocada em função do tamanho do código.

V. CONCLUSÃO

Os resultados obtidos mostram que a abordagem vetorial apresentou um tempo de execução significativamente menor em comparação à abordagem matricial, especialmente para códigos de maiores tamanhos. Esse ganho de desempenho é atribuído à estrutura simplificada e à menor sobrecarga computacional associada ao uso de vetores.

Ademais, ambas as abordagens se mostraram comparáveis em termos de eficiência de memória, pois a abordagem vetorial proporciona ganhos em tempos de execução, mantendo a alocação de memória dentro dos limites esperados. Esse fator é crucial nos sistemas de comunicação, aos quais a otimização dos recursos de memória é vital para a redução de custos operacionais e para a melhoria da confiabilidade do sistema.

Portanto, a implementação vetorial se apresenta como uma alternativa eficiente e prática para a decodificação de códigos polares, de modo que oferece uma solução balanceada entre tempo de execução e uso de memória.

REFERÊNCIAS

- [1] J. Kim, J. Park, J. Kang and J. Kang. "5G Vision: Requirements, Enabling Technologies, and Research Activities". *IEEE Communications Magazine*, vol. 54, no. 5, pp. 24–30, May 2016.
- [2] 3rd Generation Partnership Project and T. S. G. R. A. Network. *NR; Multiplexing and channel coding*. 3GPP, 2020. 3GPP TS 38.212 Version 15.6.0.
- [3] C. E. Shannon. "A Mathematical Theory of Communication". *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423 and 623–635, July 1948.
- [4] E. Arıkan. "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels". *IEEE Transactions on Information Theory*, vol. 55, no. 7, July 2009.
- [5] I. Tal and A. Vardy. "List Decoding of Polar Codes". *IEEE Transactions on Information Theory*, vol. 61, no. 5, May 2015.
- [6] A. Tanenbaum and H. Bos. *Modern Operating Systems*. Pearson Education, fourth edition, 2014.

Novo Algoritmo de Handover Baseado em Medições de Timing Advance para Redes 5G

Lidiano A. N. Oliveira, Francisco Madeiro, Waslon T. A. Lopes e Marcelo S. Alencar

Resumo—A rede 5G oferece maior capacidade, velocidade mais rápida e menor latência do que as redes predecessoras. No entanto, a alta densidade de gNodeBs e a exigência de atender a qualidade do serviço impõem desafios significativos na otimização do *handover*, essencial para garantir uma transição contínua e suave da conexão do usuário entre diferentes gNodeBs. Este artigo apresenta um novo algoritmo para cenários de *handover* em redes 5G utilizando o *Timing Advance* (TA), parâmetro empregado para ajustar o tempo de transmissão entre o usuário e a gNodeB. Combinado com medições de nível e qualidade de sinais, essa estratégia possibilitou decisões de *handovers* mais eficientes.

Palavras-Chave—Redes 5G, Handovers, Timing Advance.

I. INTRODUÇÃO

O *handover* é um procedimento das redes móveis fundamental para garantir a continuidade da conectividade do equipamento móvel (EM) com a rede à medida que ele se desloca entre diferentes áreas de cobertura. A rede 5G apresenta diversos desafios complexos devido à densidade das redes e ao número massivo de EMs [1].

A escolha adequada dos parâmetros de rede determina a sensibilidade e a eficiência das decisões de *handover* [2]. Diversos estudos buscam otimizar o processo de *handover* em redes móveis utilizando parâmetros adicionais às especificações. Apesar de potencialmente eficazes, esses parâmetros não estão disponíveis em tempo real ou não são facilmente acessíveis pela rede móvel, como a velocidade ou direção dos EMs [3]. Esses parâmetros, embora possam oferecer informações valiosas para prever e gerenciar melhor os *handovers*, enfrentam limitações práticas devido à dificuldade ou inexistência de possibilidade de coleta dessas informações.

Este estudo propõe a utilização do parâmetro de *Timing Advance* (TA) como critério para execução do *handover*. Os resultados obtidos por simulação indicam que o algoritmo proposto diminui o número de *handovers* ping-pong e as falhas de *handovers*, melhorando o desempenho da rede e a experiência do usuário.

Este artigo está organizado da seguinte forma: na Seção II é apresentada uma visão geral do procedimento de *handover*. Na Seção III é descrito o algoritmo que utiliza o parâmetro de *timing advance*. Na Seção IV são apresentados os resultados obtidos e a Seção V contém as conclusões

II. VISÃO GERAL DO PROCEDIMENTO DE HANDOVER

O procedimento de *handover* se inicia com o EM monitorando diversas métricas da gNodeB servidora e suas vizinhas, como a potência do sinal de referência recebido (*Reference Signal Received Power* – RSRP). Para evitar que os relatórios causem excesso de sinalização ao serem enviados periodicamente na rede, são adotados mecanismos predefinidos para o envio dos relatórios, chamados de eventos. O evento A3 é utilizado para indicar para a gNodeB que a qualidade do sinal de pelo menos uma das gNodeBs vizinhas é maior do que a gNodeB servidora.

Para prevenir que *handovers* desnecessários ocorram motivados por flutuações rápidas no nível de sinal, um parâmetro de margem de histerese é adotado. Esse parâmetro exige que a diferença na qualidade do sinal entre a gNodeB servidora e a gNodeB vizinha exceda um certo limiar, reduzindo a ocorrência de *handover* ping-pong (fenômeno em que o EM alterna rapidamente entre duas células vizinhas devido às pequenas oscilações do sinal). Além da histerese, outra forma de evitar variações temporárias do sinal é utilizando um temporizador, chamado de *Time-To-Trigger* (TTT). O TTT é o período de tempo em que as condições do sinal que aciona um evento de *handover* devem ser mantidas antes que o *handover* seja iniciado [4].

Após receber a solicitação de *handover*, a gNodeB vizinha escolhida aloca os recursos necessários para receber a conexão. Com os recursos alocados, a gNodeB servidora enviará um comando de *handover* para que o EM desligue a sua conexão e sincronize com a gNodeB vizinha.

Ajustando os parâmetros da margem de histerese e do TTT para dificultar a ocorrência do evento A3, os *handovers* passam a ocorrer tardiamente, podendo gerar desconexão. Caso haja facilitação, os *handovers* podem ser desnecessários e causar *handover* ping-pong. As configurações inadequadas, principalmente em redes complexas como as redes 5G, podem resultar em interrupções de conexão, quedas de chamadas, perdas de pacotes de dados e uma degradação geral da qualidade do serviço, que são consideradas falhas de *handover* [5].

Para agregar informações que auxiliem a rede na execução do *handover*, o uso do *Timing Advance* (TA) foi investigado. O TA é um comando enviado pela gNodeB para ajustar o tempo da transmissão de sinais do EM. O intuito é compensar as variações de distância entre o EM e gNodeB, adiantando a transmissão. As informações de TA são enviadas primeiramente no procedimento inicial de acesso aleatório, em que o EM está no modo ocioso, mas também são mantidas durante a conexão para manter a sincronização do *uplink*.

A especificação do *timing advance* é mais avançada nas redes 5G, permitindo ajustes mais precisos do TA [6]. Enquanto nas redes 4G a precisão era de 78,12 metros, nas redes 5G essa precisão foi aprimorada para 4,88 metros.

III. ALGORITMO PROPOSTO

Um novo algoritmo de *handover* na rede 5G é proposto para o caso do evento A3, adicionando o *timing advance* como uma das métricas a serem avaliadas pela gNodeB servidora. Portanto, além do nível de sinal da gNodeB servidora e de suas vizinhas, a indicação da distância entre o EM e a gNodeB é levada em consideração. O TA é conhecido pela rede, já que é a gNodeB que informa o o TA para cada EM. São incluídos, como requisitos para *handover* no evento A3, mais duas condições relacionadas ao TA:

- Valor mínimo de TA (TA_{MIN}) – Para cada gNodeB servidora, um valor mínimo designado de TA será necessário para que o *handover* seja validado. O *handover* prosseguirá apenas se o EM tiver um valor de TA superior ao TA_{MIN} . Dessa forma, a gNodeB estabelece um perímetro mínimo que o *handover* possa ocorrer.
- Valor máximo de TA (TA_{MAX}) – Quando o EM apresentar um valor de TA maior que o TA_{MAX} , o *handover* será executado mesmo que o RSRP da gNodeB vizinha não seja maior que o da gNodeB servidora. Contudo, há a exigência de um RSRP mínimo na gNodeB vizinha para garantir a continuidade no serviço.

A utilização do TA_{MIN} e TA_{MAX} permite um controle da execução do *handover* baseado na distância entre EM e a gNodeB. A Figura 1 ilustra a atuação do algoritmo de *handover* com TA dependendo da região que o EM se encontra.

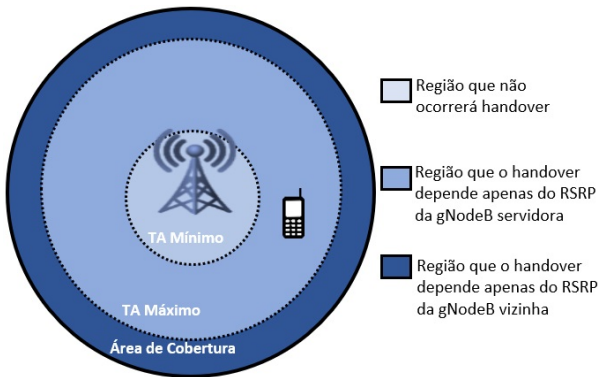


Figura 1. Algoritmo proposto baseado em TA.

IV. RESULTADOS

Para analisar o impacto do uso de TA no *handover*, foram realizadas simulações em cenários que os EMs possuem diferentes perfis de mobilidade. O simulador foi desenvolvido em C++ e implementa uma rede com 3 gNodeBs Macro e 8 *Small Cells*, com as características dos elementos (EM, antenas, potência) configurados de acordo com [7]. Foram adotados os melhores parâmetros de margem de histerese, TTT, TA_{MIN}

e TA_{MAX} que minimizassem as falhas de *handover* e os eventos de *handover* ping-pong.

A Figura 2 apresenta o percentual de falhas de *handover* e de *handovers* ping-pong encontrados utilizando o algoritmo tradicional e o algoritmo proposto em cenários com perfis de mobilidade baixo (3 km/h), médio (60 km/h) e alto (120 km/h).

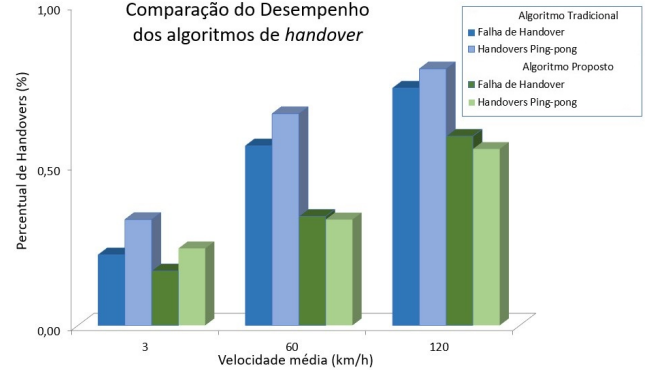


Figura 2. Comparativo do desempenho do algoritmo proposto.

O uso do *timing advance* na execução de *handovers* diminuiu o percentual de falhas de *handover* e a taxa de *handover* ping-pong em 27,4% e 36,17%, respectivamente. O resultado demonstra que o algoritmo proposto reduziu os eventos de *handover* analisados em todos os cenários investigados.

V. CONCLUSÃO

O artigo apresentou um algoritmo de *handover* que utiliza o *timing advance*, parâmetro já conhecido pela rede. Foi investigada sua utilização em cenários com EMs com diferentes velocidades médias e os resultados obtidos indicam que o uso do TA diminuiu as falhas de *handover* e os *handovers* ping-pong. Combinado com outros parâmetros de rede, como a potência do sinal recebido, o TA forneceu uma visão mais abrangente das condições geográficas do usuário, permitindo uma decisão de *handover* mais precisa.

REFERÊNCIAS

- [1] L. A. N. Oliveira, M. S. Alencar, and W. T. A. Lopes, "Evolução da Arquitetura de Redes Móveis Rumo ao 5G," *Revista de Tecnologia da Informação e Comunicação*, vol. 8, no. 2, pp. 43–50, 2018.
- [2] K. Vasudeva, M. Simsek, D. Lopez-Perez, and I. Guvenc, "Analysis of Handover Failures in Heterogeneous Networks With Fading," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6060–6074, 2017.
- [3] M. Yoon, J. Park, T. Park, J. Seo, J.-K. Yun, and K. Cho, "An Adaptive Handover Scheme to Support UE with Various Movement Speeds in 5G Network," in *2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pp. 1–4, 2022.
- [4] Lovekesh, M. Yadav, and D. Nandal, "Investigating Impact of 5G Mobility Management Over Handover," in *2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech)*, pp. 559–565, 2023.
- [5] W. K. Saad, I. Shaye, B. J. Hamza, H. Mohamad, Y. I. Daradkeh, and W. A. Jabbar, "Handover Parameters Optimisation Techniques in 5G Networks," *Sensors*, vol. 21, no. 15, 2021.
- [6] 3GPP, "Technical Specification Group Radio Access Network; NR; Physical Channels and Modulation," Technical Specification (TS) 38.211, 3rd Generation Partnership Project (3GPP), March 2024. Version 18.2.0.
- [7] J. Sultan, M. S. Mohsen, N. S. G. Al-Thobhani, and W. A. Jabbar, "Performance of Hard Handover in 5G Heterogeneous Networks," in *2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, pp. 1–7, 2021.

Planejamento de Redes 5G Small Cells Usando Dados Reais Considerando Cobertura e Capacidade: Abordagem com Clusterização e Meta-Heurística

Andrey Manhyck C. B., Flávio Henry Ferreira, Evelin Cardoso, Carlos Renato Francês, Fabrício Barros e Jasmine P. L. Araújo

Abstract—Este artigo expõe a teoria e uma simulação comparativa acerca da otimização do posicionamento de small cells usando dados reais por meio do uso de algoritmos híbridos de modo a maximizar a cobertura e a capacidade de maneira balanceada por cluster e minimizar a quantidade de clusters utilizados, por fim comparando os resultados entre priorizar somente cobertura ou balancear cobertura e capacidade.

Index Terms—Clusterização, computação bioinspirada, otimização, balanceamento, cobertura, capacidade

I. INTRODUÇÃO

O 5G e futuras redes são idealizados para provisionar casos de uso definidos pelo 3GPP, como comunicação ultra confiável de baixa latência (URLLC), banda larga móvel aprimorada (eMBB) e comunicações massivas do tipo de máquina (mMTC). Esses casos de uso devem sustentar os rígidos requisitos necessários por aplicativos como realidade virtual, veículos para infraestrutura (V2X) e comunicações de missão crítica. Bem como, as infraestruturas das redes para suporte aos serviços de 5G estão sendo delineadas, e este delineamento de redes assistido por dados reais torna-se imprescindível. Comumente, agrupar usuários pela cobertura e identificar regiões de tráfego mais alto através de clusterização permite determinar o número adequado de clusters e fornecer menor custo por bit, otimizando custos de capital e operacionais (CAPEX e OPEX) para as operadoras de telecomunicações[1].

As técnicas de machine learning (ML) e big data para redes da próxima geração (NGNs) têm sido consideradas tecnologias chave para a implementação dessas redes[2]. Desse modo, espera-se que sejam sistemas altamente complexos devido à heterogeneidade de dispositivos, redes, serviços e requisitos de aplicativos. As técnicas de ML têm realizações versáteis na análise de big data, tomada de decisão baseada em dados, correção na estimação de parâmetros e problemas de otimização multi-objetivo. Comumente, elas têm sido aplicadas a cenários e técnicas NGNs, incluindo múltiplas entradas e múltiplas saídas massivas (mMIMO), máquina para máquina (M2M), redes ad-hoc veiculares (VANETs) e Internet das Coisas (IoT). Neste estudo, o objetivo é usar os dados reais coletados de usuários LTE a partir do OpenCellid para o planejamento, avaliação de cobertura [3] e capacidade [13] para redes 5G, utilizando a

faixa de frequência de 3,5 GHz reservada para o 5G no Brasil, por meio do uso de algoritmos de otimização híbrida em conjunto com o cálculo de perda do modelo de propagação escolhido, em uma área de bairros de Belém, Pará, Brasil [9].

II. MODELO DE PERDA DE CAMINHO ECC-33

O modelo de perda escolhido é adaptado ao ambiente urbano e ambientes florestais da cidade de Belém proposto por [3] e descrito nas equações (1), (2), (3), (4) e (5):

$$PL_{ECC} = A_{fs} + A_{bm} - GF_t - GF_r \quad (1)$$

$$A_{fs} = 92,4 + 20 \log(d) + 20 \log(fc) \quad (2)$$

$$A_{bm} = 20,41 + 9,83 \log(d) + 7,89 \log(fc) + 9,56(\log(fc))^2 \quad (3)$$

$$GF_t = \log\left(\frac{h_{bs}}{200}\right)(13,958 + 5,8 \log(d))^2 \quad (4)$$

$$GF_r = 0,759(h_{ue}) - 1,862 \quad (5)$$

III. OTIMIZAÇÃO HÍBRIDA

Para o entendimento do processo da otimização híbrida se faz necessário o entendimento das duas técnicas utilizadas e sua integração, segue uma descrição teórica das técnicas empregadas:

A. Técnica de clusterização Mean Shift

O mean shift representa um procedimento geral de localização/agrupamento de modo não paramétrico, em outras palavras, está técnica busca encontrar grupos de pontos de dados semelhantes ou agrupados num conjunto de dados, além disso o algoritmo localiza os centros desses grupos e atribui os pontos de dados aos clusters correspondentes [5].

B. Algoritmo Bioinspirado Polinização de Flores (FPA)

No FPA cada solução candidata para o problema de otimização é considerada como uma "flor", e a qualidade dessa solução é determinada pela sua adaptação ou valor de aptidão. Os "agentes polinizadores" são responsáveis por explorar e modificar as soluções candidatas, transferindo informações entre elas para melhorar a qualidade das soluções ao longo do tempo [6].

C. Processo de Híbridização

O processo inicia pela extração da localização (Lat, Long) de usuários gerados aleatoriamente, de acordo com os dados obtidos do OpenCell ID. Em seguida, essas localizações são implementadas no Mean Shift de modo a encontrar os centros das regiões com mais usuários atendidos por um único cluster com um sinal por usuário de -90 dBm ou maior. O FPA, por sua vez, busca encontrar a melhor solução com base nos coeficientes da função objetiva dentro de um espaço de soluções possíveis, segue os coeficientes (Saídas): Cobertura (Cob), Potência de transmissão (Tx), Número de Clusters ($N^{\circ}_{clusters}$), informação de estado do canal (CSI) e Carga da célula (CL). Segue abaixo a função objetiva:

$$Z = (obj1)*Cob + (obj2) \cdot Tx + (obj3) \cdot N^{\circ}_{clusters} + (obj4) CSI + (obj5)*CL, \quad (6)$$

sendo os coeficientes CSI e CL responsáveis pelo balanceamento da cobertura, descritas em [7].

IV. SIMULAÇÃO E RESULTADOS

Para a simulação com foco em cobertura os pesos CSI e CL são zerados, com base neste detalhe utilizou-se as seguintes entradas: largura de banda Mean Shift (0.001 – 0.2); potência de transmissão (30 – 40). Na simulação da cobertura utilizou-se os seguintes pesos: Cob = 60%, Tx = 10%, $N^{\circ}_{clusters}$ = 30% e 0% para os pesos restantes. Por fim, obteve-se os seguintes Inputs ótimos: (0.0027, 39.8221); 68 Clusters e um erro associado de 3,5% para 1000 iterações. Segue o mapa de cobertura na imagem a seguir:

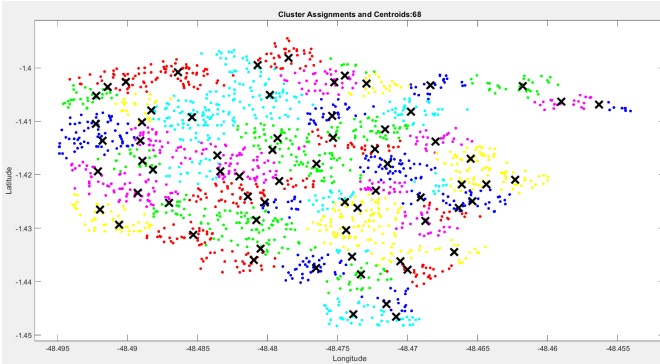


Fig. 1. Disposição de clusters com foco em cobertura

Agora, para a simulação da rede balanceando cobertura e capacidade utilizou-se os seguintes Inputs: largura de banda Mean Shift (0.001 – 0.2); potência de transmissão (30 – 40); Limite de usuários por célula (n° de usuários / Cap. máx de usuários); Carga da célula (0.01 – 1) e os seguintes pesos: cob = 30%, Tx = 10%, $N^{\circ}_{clusters}$ = 30%, CSI = 15% e CL = 15%, obteve-se os seguintes Inputs ótimos: (0.0026, 38.5885, 30, 0.3057); 70 Clusters e um erro associado de 3% para 1000 iterações. Segue o mapa de cobertura na imagem a seguir:

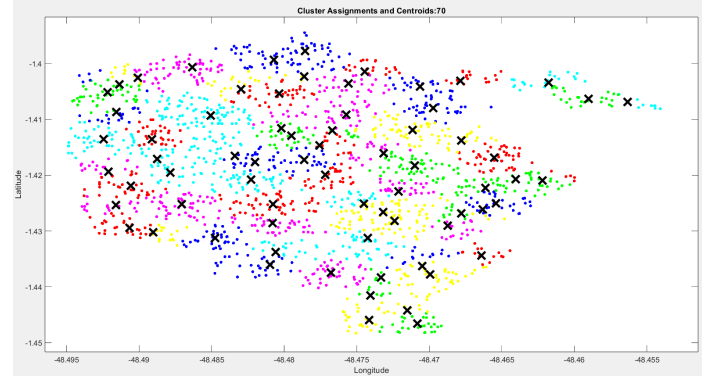


Fig. 2. Disposição de clusters balanceando Cap e Cob

Uma breve análise comparativa sobre a otimização do posicionamento de small cells balanceando cobertura e capacidade foi realizada neste artigo. Este estudo é uma melhoria do artigo [3], implementando a otimização da capacidade do artigo [7], buscando assim os melhores coeficientes da equação (6) de modo a balancear a capacidade e cobertura.

V. CONCLUSÃO

O foco na capacidade se demonstra essencial no planejamento das redes, pois por meio dessa estratégia é possível atender o maior número de usuários dentro de uma região de modo a racionalizar a potência transmitida (eficiência energética) e melhorar a qualidade da experiência (QoE) do usuário tendo em vista a previsão da quantidade de usuários dada por esta solução meta-heurística, assim permitindo que cada cluster trabalhe sob demanda, fornecendo a energia e a banda necessária para atender os usuários.

REFERENCES

- [1] S. Wang and C. Ran, "Rethinking cellular network planning and optimization," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 118125, Apr. 2016.
- [2] H.B. Yilmaz, C. -B.Chae, Y. Deng, T. O'Shea, L. Dai, N. Lee, and J. Hoydis, "Special issue on advances and applications of artificial intelligence and machine learning for wireless communications," *J. Commun. Netw.*, vol.22, no. 3, pp. 173176, Jun. 2020.
- [3] Ferreira FH, Barros FJB, Neto MCda, Cardoso E, Francês CRL, Araújo J. 2023. Hybrid computational and real data-based positioning of small cells in 5G networks. *PeerJ Comput. Sci.* 9:e1412 DOI 10.7717/peerj-cs.1412
- [4] Carvalho AA, Batalha I, Neto M, Castro B, Barros F, Araújo J, Cavalcante G. 2021. Adjusting large-scale propagation models for the amazon region using bioinspired algorithms at 1.8 and 2.6 GHz frequencies. *Journal of Microwaves, Optoelectronics and Electromagnetic Applications* 20(3):445–463 DOI 10.1590/2179-10742021v20i31099.
- [5] Derpanis, K. G. (2005). Mean shift clustering. *Lecture Notes*, 32(1), 4.
- [6] YANG, Xin-She. *Nature-Inspired Optimization Algorithms*. Elsevier, 2014. - 11.2.2 Flower Pollination Algorithm, P.
- [7] KHAN, M. Umar et al. Service-Based Network Dimensioning for 5G Networks Assisted by Real. *Parameters*, v. 1, n. B2, p. B3, 2021.

Algoritmo Recursivo para Cálculo de Medidas Cicloestacionárias

Pedro Thiago V. de Souza
Universidade Federal Rural do Semi-Árido
Centro Multidisciplinar de Pau dos Ferros
Pau dos Ferros, RN, 59900-970
Email: pedro.souza@ufersa.edu.br

Vinicius Samuel V. de Souza
Universidade Federal Rural do Semi-Árido
Centro Multidisciplinar de Pau dos Ferros
Pau dos Ferros, RN, 59900-970
Email: viniussamuel@ufersa.edu.br

Resumo—Neste trabalho propomos um novo algoritmo para o cálculo das medidas cicloestacionárias de forma recursiva. Para isso, utilizamos o algoritmo de Goertzel para o cálculo das transformadas discretas de Fourier envolvidas. O desempenho do algoritmo foi verificado em um cenário de sensoriamento espectral e os resultados encontrados demonstram a eficiência da técnica proposta em cenários com baixa relação sinal-ruído (*Signal-to-Noise Ratio* - SNR).

Index Terms—Sensoriamento espectral, cicloestacionariedade, rádio cognitivo, transformada discreta de Fourier.

I. INTRODUÇÃO

O sensoriamento espectral é uma das principais tarefas de um sistema de rádio cognitivo [1]. Um dos métodos mais difundidos para o sensoriamento espectral é o sensoriamento por análise de cicloestacionariedades [2]. Esse método é bastante robusto à ambientes com baixa SNR [3], todavia, possui alta complexidade computacional [4].

Nesse trabalho propomos um método para o cálculo das métricas cicloestacionárias de forma recursiva, não havendo necessidade do armazenamento do sinal de entrada. O algoritmo proposto é baseado nas propriedades de simetria da principal métrica de análise cicloestacionária - a função densidade de correlação espectral (*Spectral Correlation Density* - SCD) - e utiliza o algoritmo de Goertzel para o cálculo recursivo das transformadas discretas de Fourier (*Discrete Fourier Transform* - DFT).

II. ALGORITMO PROPOSTO

A SCD é definida como [2]:

$$S^\alpha(\omega) = \left\langle X\left(\tau, \omega + \frac{\alpha}{2}\right) X^*\left(\tau, \omega - \frac{\alpha}{2}\right) \right\rangle, \quad (1)$$

sendo $\langle \cdot \rangle$ o operador de média temporal e $X(t, \omega)$ a transformada de Fourier de curto termo:

$$X(\tau, \omega) = \mathcal{F}\{x(t)g(t - \tau)\} \quad (2)$$

sendo $g(t)$ uma função de janela e $\mathcal{F}\{\cdot\}$ o operador que denota uma transformada de Fourier em tempo contínuo.

Admitindo-se $x(t)$ e $g(t)$ sinais reais, e aplicando-se as propriedades de simetria da transformada de Fourier, verifica-se que:

$$X\left(\tau, \omega - \frac{\alpha}{2}\right) = X^*\left(\tau, -\omega + \frac{\alpha}{2}\right). \quad (3)$$

logo:

$$S^\alpha(\omega) = \left\langle X\left(\tau, \omega + \frac{\alpha}{2}\right) X\left(\tau, -\omega + \frac{\alpha}{2}\right) \right\rangle. \quad (4)$$

No sensoriamento espectral, as assinaturas cicloestacionárias são baseadas no módulo da SCD [6]. Utilizando as propriedades de simetria da transformada de Fourier, verifica-se que:

$$|S^\alpha(\omega)| = \left\langle \left| X\left(\tau, \omega + \frac{\alpha}{2}\right) \right|^2 \right\rangle \quad (5)$$

Pode-se obter uma estimativa da métrica apresentada na Eq. (5) amostrando-se o sinal de entrada $x(t)$, obtendo-se o sinal $x[n]$ e após isso, dividir o sinal $x[n]$ em L blocos de N amostras. Com isso, uma estimativa do módulo da SCD pode ser obtida como:

$$|\widehat{S^\alpha[k]}| = \frac{1}{L} \sum_{r=0}^{L-1} \left| X_r\left[k + \frac{\alpha}{2}\right] \right|^2, \quad r = 0, 1, \dots, L-1 \quad (6)$$

com:

$$X_r[k] = \text{DFT}\{x[n + rN]g[n]\} \quad (7)$$

sendo $\text{DFT}\{\cdot\}$ a DFT de N amostras e $g[n]$ uma função janela. Aplicando-se as propriedades da DFT, nota-se que:

$$X_r\left[k + \frac{\alpha}{2}\right] = \text{DFT}\{x[n + rN]g[n]e^{j\frac{\pi\alpha n}{N}}\}, \quad (8)$$

e utilizando-se a Eq. (8) na Eq. (6), verifica-se que:

$$|\widehat{S^\alpha[k]}| = \frac{1}{L} \sum_{r=0}^{L-1} \left| \text{DFT}\{x[n + rN]g[n]e^{j\frac{\pi\alpha n}{N}}\} \right|^2. \quad (9)$$

O termo $|\text{DFT}\{x[n + rN]g[n]e^{j\frac{\pi\alpha n}{N}}\}|^2$ pode ser calculado de forma recursiva utilizando-se o algoritmo de Goertzel [7] através do conjunto das Eq. (10) e (11):

$$v_r^k[n] = x[n + rN]g[n]e^{j\frac{\pi\alpha n}{N}} + 2 \cos\left(\frac{2\pi k}{N}\right) v_r^k[n-1] - v_r^k[n-2], \quad (10)$$

$$V_r[k] = (v_r^k[N])^2 - 2 \cos\left(\frac{2\pi k}{N}\right) v_r^k[N]v_r^k[N-1] + (v_r^k[N-1])^2, \quad (11)$$

e por fim, a estimativa do módulo da SCD é calculada como:

$$|\widehat{S^\alpha[k]}| = \frac{1}{L} \sum_{r=0}^{L-1} V_r[k]. \quad (12)$$

O sensoriamento por análise de cicloestacionariedades baseia-se no princípio que o ruído branco Gaussiano não possui componentes espectrais para $\alpha \neq 0$ [2]. Além disso, para sinais modulados, as componentes espectrais aparecem em posições bem definidas do ciclo-espectro. Desta forma, a métrica de sensoriamento aqui proposta é definida por:

$$T = |\widehat{S^\alpha[k]}| \quad (13)$$

sendo α e k valores escolhidos de acordo com o conjunto de modulações no qual deseja-se detectar. O teste de hipótese decidirá pela presença do usuário primeiro se $T > T_0$, sendo T_0 um limiar determinado pela probabilidade de falso alarme.

III. SIMULAÇÕES COMPUTACIONAIS

O limiar de detecção para foi obtido através de uma curva que relaciona o limiar de detecção e a probabilidade de falso alarme. Essa curva é apresentada na Fig. 1. Cada ponto da curva foi obtido através da média de 500 simulações, em que aplicou-se no algoritmo proposto um sinal de ruído Gaussiano de média zero e variância unitária. De forma a atender o padrão IEEE 802.22, que regulamenta uma probabilidade de falso alarme de 10% [8], obteve-se o limiar de $T_0 = 0,4639$.

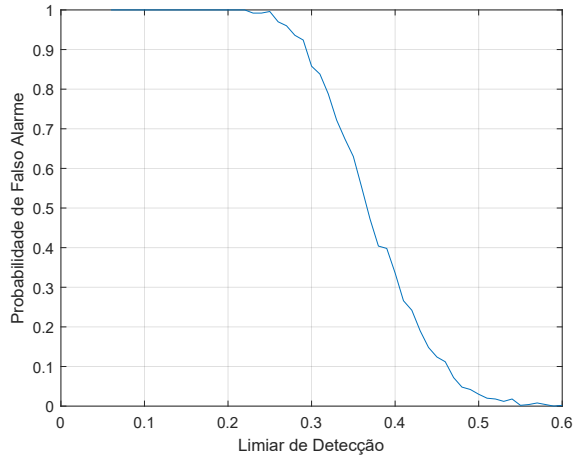


Figura 1. Limiar de detecção em função da probabilidade de falso alarme.

De forma a verificar o desempenho do algoritmo proposto, obteve-se a curva de probabilidade de detecção em função da SNR do canal. O sinal a ser sensoriado é proveniente de uma constelação BPSK com os parâmetros descritos na Tabela I. Para esse tipo de modulação, há uma raia espectral localizada em $\omega = 0$ e $\alpha = 2f_c/f_s$, de tal forma que os parâmetros k e α foram escolhidos de forma a refletirem esses valores. Os restantes dos parâmetros do algoritmo proposto estão apresentados na Tabela I.

A Fig. 2 apresenta a curva de taxa de detecção do sistema de sensoriamento proposto. Cada ponto da curva foi obtido como uma média das decisões tomadas em 500 testes. Analisando a Fig. 2, observa-se que o desempenho da arquitetura é degradado com a diminuição da SNR, possuindo taxas de detecção superiores à 90% em SNR acima de -11dB, demonstrando a eficiência do algoritmo mesmo em cenários com baixa SNR.

Tabela I
PARÂMETROS DE SIMULAÇÃO.

Característica	Parametrização
Frequência da Portadora	$f_c = 1024\text{Hz}$
Frequência de amostragem	$f_s = 4096\text{Hz}$
Sobreamostragem	16 amostras/símbolo
Fator de roll-off	0,5
Quantidade de Blocos (L)	64
Tamanho de Bloco (N)	256
α	$2f_c/f_s$
k	0
Janela $w[n]$	Hann

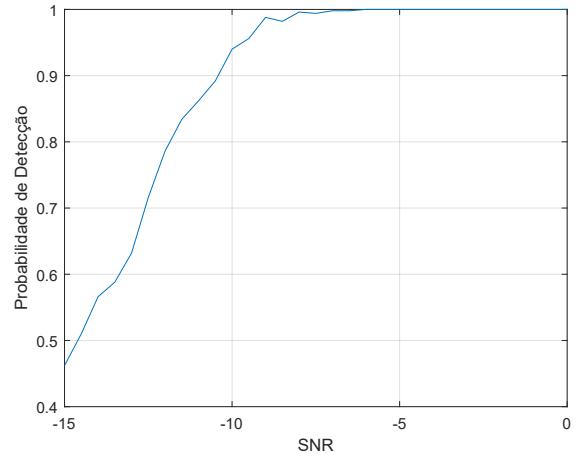


Figura 2. Taxa de detecção do usuário primário em função da SNR do canal.

IV. CONCLUSÕES

Analisando-se os resultados obtidos, pode-se constatar que o sistema proposto possui bons resultados de taxa de detecção mesmo em baixas SNR. Como trabalhos futuros pretende-se implementar o sistema proposto em bancada experimental e comparar-ló com outros algoritmos de sensoriamento.

REFERÊNCIAS

- [1] J. I. Mitola, Cognitive Radio. An integrated agent architecture for Software Defined Radio, Springer, 2002
- [2] W. Gardner, Cyclostationarity in communications and signal processing, STATISTICAL SIGNAL PROCESSING INC: YOUNTVILLE CA, 1994.
- [3] E.L. Costa - Detection and Identification of Cyclostationary Signals - Naval Postgraduate School, 1996.
- [4] A. Lima, Arquiteturas Eficientes para Sensoriamento Espectral e Classificação Automática de Modulação Usando Características Cicloestacionárias., Universidade Federal do Rio Grande do Norte – UFRN, Natal, RN, 2014.
- [5] P. Souza, A. Fontes, V. Souza, L. Silveira - Classificação de Modulação utilizando Cicloestacionariedades e Árvores de Decisão. In: X Conferência Nacional em Comunicações, Redes e Segurança da Informação, 2020, Natal. Anais da X Conferência Nacional em Comunicações, Redes e Segurança da Informação - Encom 2020, 2020. p. 159-160.
- [6] A. Farias, C.E.N. Melo e R.V. Dias - Uma Abordagem a Métodos de Sensoriamento do Espectro Aplicados a Rádio Cognitivo - UNB, Brasília, 2010.
- [7] A. Oppenheim, R. Schaffer - Digital Signal Processing. Prentice Hall, 1975.
- [8] P.S. Coutinho - Detecção de Energia para Rádios Cognitivos usando GNU Radio e USRP2 - Universidade Federal do Rio de Janeiro, 2011.

Análise de Perda de Percurso para Sistemas Multiportadoras na Faixa de 700 MHz

Grazielle de Lima, Ulysses R. C. Victor, Marcelo L. Nunes, Alysson L. Reis e Alvaro A. M. de Medeiros

Resumo—Visando uma melhor ocupação do espectro de 700 MHz, antes utilizado pela TV analógica, o Brasil destinou o uso desta faixa a implantação de sistemas celulares multiportadoras. Neste sentido, este artigo tem como objetivo avaliar o desempenho de modelos empíricos de predição de cobertura através da comparação com valores obtidos através de campanha de medição de uma operadora celular na região da Lagoa da Pampulha em Belo Horizonte. É observado que uma correção de *offset* para os modelos avaliados que apresenta significativas melhorias na adequação aos valores medidos.

Palavras-Chave—Sistemas Celulares Multiportadora, Cobertura, Perda de Percurso, Modelos Empíricos.

I. INTRODUÇÃO

Os sistemas celulares estão em aperfeiçoamento constante no que se refere à capacidade de trafegar maiores volumes de dados com maior confiabilidade [1]. Diversos fatores referentes à propagação dos sinais de comunicação tornam a evolução dos sistemas celulares desafiadora, tais como a influência de transições de diferentes tipos de ambientes, variações das alturas dos terreno e dos edifícios, tipo de vegetação, e presença de áreas cobertas de água [2]. Os modelos de propagação para a perda de percurso são importantes para o planejamento de sistemas de telecomunicações por possibilitarem uma predição da potência de sinal recebido em função da distância. Tais modelos têm historicamente evoluído através de adequações dos modelos tradicionais para as diferentes regiões analisadas.

Este artigo objetiva a caracterização de canais sem fio para sistemas celulares multiportadoras na faixa de 700 MHz (entre 698 MHz e 806 MHz). Com a migração para a TV digital, essa faixa foi distribuída entre as operadoras de telefonia móvel para implementação da tecnologia *Long Term Evolution* (LTE) e posteriormente da tecnologia *New Radio* (NR). Uma das principais vantagens do LTE na faixa de 700 MHz é a melhoria da cobertura devido à redução da frequência da portadora.

Para isso, o estudo baseia-se em medições de potência de recepção a partir de sinais transmitidos por uma estação base (*Base Station* - BS) de uma operadora de telefonia móvel na região próxima à Pampulha em Belo Horizonte. Essa região apresenta um ambiente urbano composto de diferentes fatores de propagação, tais como grandes variações de relevo, diferentes tipos de edificações (prédios, casas, supermercados, shopping centers), áreas de espelho d'água (Lagoa da Pampulha e Lagoa do Nado), áreas abertas com edificações esparsas (UFMG, Estádio Mineirão, Aeroporto da Pampulha), e áreas de vegetação (Jardim Zoológico, Jardim Botânico, Parque Ecológico).

Grazielle de Lima, Ulysses R. C. Vitor e Alvaro A. M. de Medeiros são do Programa de Pós Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora, Juiz de Fora-MG, Brasil, e-mail: {grazielle.lima, ulysses.vitor, alvaro}@engenharia.ufjf.br; Marcelo L. Nunes e Alysson L. Reis são da Agência Nacional de Telecomunicações, Belo Horizonte-MG, Brasil, email: {mlnunes, alysson}@anatel.gov.br

II. CAMPANHA DE MEDIÇÃO

Sistemas multiportadora como LTE e NR utilizam a multiplexação por divisão de frequência ortogonal (*Orthogonal Frequency Division Multiplexing* - OFDM). Tal técnica converte um fluxo de dados de alta taxa em vários fluxos de baixa taxa que são transmitidos por N canais paralelos de banda estreita, chamados de subportadoras, que podem ser mais facilmente equalizados [1].

Os principais parâmetros de monitoramento da camada física de tais sistemas baseiam-se na medição de sinais arbitrários em subportadoras específicas (sinais pilotos). A Potência Recebida do Sinal de Referência (*Reference Signal Receive Power* - RSRP) é aquele utilizado neste estudo para o cálculo da perda de percurso. O parâmetro RSRP é a potência média recebida dos sinais piloto em subportadoras espalhadas por toda a largura de banda e é considerado banda estreita. O cálculo do RSRP é baseado na energia recebida dentro da parte útil do símbolo OFDM, sendo no mínimo -20 dB de relação sinal-interferência-mais-ruído (*Signal to Interference plus Noise* - SINR) para o sinal de sincronismo secundário (*S-Synch*) necessário para detecção.

A campanha coletou dados do enlace direto (*downlink*) de 6 BSs com 3 setores cada, sendo realizada pela equipe de campo da Anatel da Gerência Regional em Belo Horizonte nos dias 14, 18 e 19 de dezembro de 2018 através de um analisador de sistemas sem fio com a função de varredura para incursões de *drive test*. Utilizou-se o equipamento TSMW da Rhode & Schwarz. O equipamento opera na faixa de 30 MHz a 6 GHz, e possui GPS integrado e receptores de RF independentes com largura de banda de 20 MHz, possibilitando-o medir simultaneamente até 12 portadoras com uma taxa de até 100 amostras por segundo com sensibilidade de até -123 dBm [3].

As medições foram feitas a uma distância máxima de 900 m da BS, a uma velocidade média de 50 km/h. Para essa análise utilizou-se as medidas do dia 18 coletada na estação cujos setores são assinalados pelos *Physical Layer Cell Identifiers* (PCIs) 156, 157 e 158, localizada no bairro São Tomaz. Na Figura 1 é possível observar uma região composta de diferentes elementos que afetam a propagação, tais como casas, prédios, galpões, vegetação densa e aeroporto.

Para todas medições foram considerados a frequência de operação, o diagrama de radiação da antena da BS, a altura da antena transmissora dentro dos limites suportados pelos modelos de propagação, as perdas nos cabos e conectores, e os ganhos do amplificador de potência da BS. A BS selecionada apresenta as seguintes configurações: *tilt* mecânico de 10,4°, ganho máximo de 15,8 dBi, potência máxima de 80 W (49 dBm). Todos estes dados possibilitaram uma estimativa mais precisa da perda de percurso conforme o procedimento descrito em [4] e [5].

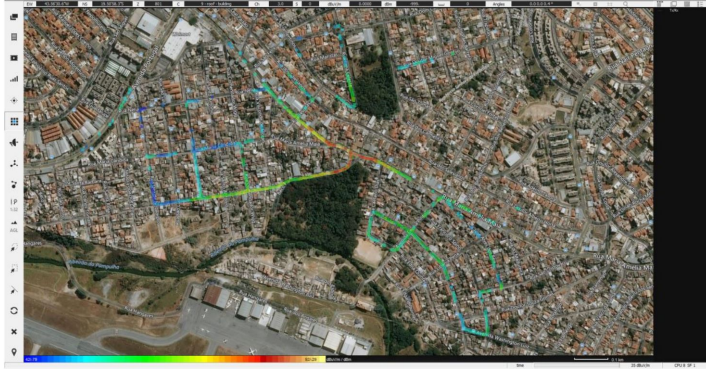


Fig. 1. Potência de recepção para BS no bairro São Tomaz.

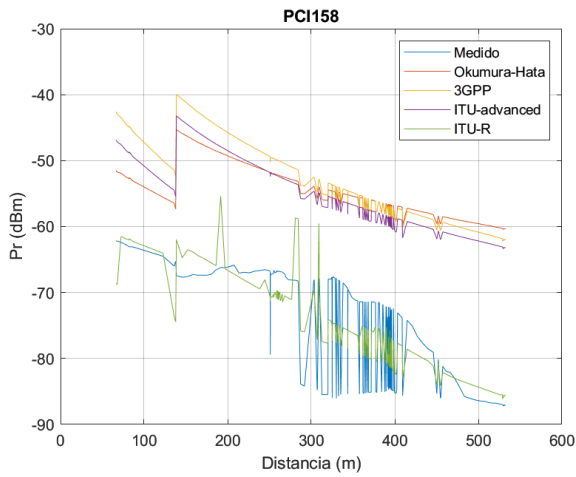


Fig. 2. Valores de potência de recepção medidos e preditos pelos modelos avaliados em função da distância para a PCI 158.

III. ANÁLISE E RESULTADOS

Após processamento dos dados coletados e cálculo da perda de percurso em função da distância da BS, os dados foram comparados com o tradicional modelo Okumura-Hata, além dos modelos de perda de percurso definidas para as especificações 3GPP e ITU-Advanced. As expressões para as perdas de percursos destes três modelos podem ser encontradas nos Apêndices do Capítulo 7 de [2].

Ademais, as medidas também foram comparadas com o modelo definido na recomendação ITU-R 1812-6 [6]. Esse método de predição de cobertura leva em consideração os seguintes elementos do modelo: as perdas em visada direta, as perdas por difração (abrangendo terrenos lisos e irregulares), perdas por dispersão troposférica, perdas por propagação anômala (reflexão e refração atmosféricas), perdas por variabilidade de localização, variação de ganho pela altura da camada morfológica e as perdas por penetração em edifícios.

Ao comparar as perdas de percurso medida e preditas, considera-se a variação de intensidade de sinal em função da distância, levando em conta o diagrama de radiação da antena da BS para cada posição da estação móvel. A Figura 2 mostra os valores de potência de recepção medidos e preditos pelos modelos avaliados em função da distância para a PCI 158. Observa-se uma significativa diferença de *offset* entre os modelos Okumura-Hata, 3GPP e ITU-Advanced em comparação com os valores medidos. A curva do modelo ITU-R 1812-6 teve um melhor ajuste com a curva de dados obtidos. O mesmo aconteceu para os outros setores da mesma BS.

A avaliação de desempenho dos modelos fez uso dos pa-

TABELA I

ANÁLISE DOS RESULTADOS DA ESTAÇÃO SÃO TOMAZ

PCI156	RMSE	Distância Média	Correlação
Okumura-Hata	4,636	-4,6573e-15	0,8584
3GPP	4,8013	-7,404e-15	0,9009
ITU-Advanced	4,6372	-1,3614e-14	0,8923
ITU-R 1812-6	4,2378	2,3725	0,9227
PCI157	RMSE	Distância Média	Correlação
Okumura-Hata	1,9672	6,1321e-15	0,9159
3GPP	3,6273	-8,7601e-16	0,9109
ITU-Advanced	3,1221	3,8934e-15	0,9122
ITU-R 1812-6	4,7036	-2,1098	0,8245
PCI158	RMSE	Distância Média	Correlação
Okumura-Hata	5,5658	2,7961e-14	0,8108
3GPP	4,7384	-2,1612e-14	0,8163
ITU-Advanced	4,966	-1,6119e-15	0,8186
ITU-R 1812-6	4,2092	-3,1038	0,8522

râmetros raiz do erro médio quadrático (*root mean squared error* - RMSE), coeficiente de correlação entre as curvas, bem como a distância média entre as curvas medida e predita. Para esta avaliação, buscou-se retirar o efeito da diferença de *offset* entre as curvas preditas e medidas através da subtração de um valor constante (nível DC) de cada curva. Tal valor é aquele que minimiza o RMSE entre a curva predita e a medida.

A Tabela III apresenta os valores de RMSE, distância média e correlação entre as curvas preditas pelos modelos e as curvas medidas de potência de recepção a partir da potência transmitida pelos 3 setores da BS avaliada. É possível observar que o modelo ITU-R 1812-6 se adequa melhor à curva medida. Nota-se também que as curvas dos modelos restantes apresentam melhor concordância com os dados medidos após a correção de *offset*.

IV. CONCLUSÕES

A caracterização da perda de percurso tem evoluído conjuntamente com os padrões de sistemas celulares, cuja gerações mais recentes se baseiam em comunicações multiportadoras. A partir de sinais de controle destes sistemas é possível medir o nível de potência recebida em função da distância e comparar com modelos de predição, buscando aquele que melhor ajusta ao ambiente investigado.

Nesse artigo, comparou-se diversos modelos de perda de percurso em uma região urbana composta de diferentes ambientes de propagação. Por considerar mais efeitos de propagação, tais como difração em edifícios, o modelo ITU-R 1812-6 apresentou melhor adequação às curvas obtidas pela campanha de medição. Contudo, modelos de mais fácil implementação computacional, tais como 3GPP e ITU-Advanced apresentam desempenho satisfatório após ajuste de *offset*.

REFERÊNCIAS

- [1] M. Sauter, *From GSM to LTE-Advanced Pro and 5G An Introduction to Mobile Networks and Mobile Broadband*. Wiley, 2021.
- [2] A. F. Molisch, *Wireless Communications: From Fundamentals to Beyond 5G*, 3rd ed. IEEE Press, 2022.
- [3] *RS TSMW and RS TSMF Universal Radio Network Analyzers - User Manual*, Rohde & Schwarz, 2017.
- [4] M. L. Nunes, "Predição de cobertura radioelétrica em ambiente urbano complexo utilizando base de dados descritiva de alta resolução," Master's thesis, Universidade Federal de Minas Gerais, 2017.
- [5] A. L. Reis, C. G. d. Rego, D. C. T. Lopez, and M. L. Nunes, "A comparative study for radioelectric coverage models applied to mobile communication systems performance analysis," in *2018 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, 2018, pp. 746–749.
- [6] ITU-R, *Recommendation 1812-6: A path-specific propagation prediction method for point-to-area terrestrial services in the frequency range 30 MHz to 6 000 MHz*, 2023.

Análise em Larga Escala da Perda de Propagação em 3,5 GHz em Ambientes 3D de Aeroporto

Mozart L. M. Junior

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
juniorlamaquias09@gmail.com

Alex S. Macedo

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
alexsanchess@gmail.com

Fabício J. B. Barros

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
fbarros@ufpa.br

Jasmine P. L. Araújo

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
jasmine@ufpa.br

Resumo—O uso de micro-ondas para comunicação tem se tornado foco de um grande número de pesquisas devido às suas características, que incluem altas taxas de transmissão e a interação das ondas eletromagnéticas com o meio. Ferramentas computacionais avançadas têm sido fundamentais para o planejamento de projetos de redes móveis. Neste contexto, este trabalho apresenta a criação de um ambiente de aeroporto 3D utilizando o software *Wireless Insite* (WI), com o objetivo de avaliar e validar sinais de transmissão reais na frequência de 3,5 GHz. A análise dos parâmetros de perda de propagação revela como aspectos do ambiente simulado influenciam a intensidade do sinal. Os resultados indicam uma diferença de RMSE entre os valores medidos e simulados de 5,09 e 4,31 para as rotas 1 e 2, respectivamente. Esses achados validam o cenário 3D e demonstram a eficácia dos modelos de propagação aplicados.

Index Terms—Micro-ondas, Perda de Propagação, Wireless Insite, Modelagem 3D, Larga escala.

I. INTRODUÇÃO

A alta demanda por comunicação confiável e de alto desempenho para ambientes internos impulsiona pesquisas sobre frequências 5G, divididas em 500 MHz-6 GHz e acima de 6 GHz (ondas milimétricas, de 10 cm a 1 mm) [1]. Experimentos atuais utilizam modelos de perda de propagação como *Close-In Free Space* (CI) e *Floating-Intercept* (FI) para analisar canais e perdas de sinal [2]. Ambientes *indoor* são estudados devido aos efeitos ambientais na intensidade do sinal [2]–[4].

Os trabalhos apresentados em [2], [3] utilizam o software WI para modelar ambientes indoor. No primeiro, foi modelado um prédio de dois andares e avaliada a perda de propagação, resultando em um RMSE (*Root Mean Squared Error*) de 4,1 no térreo e 8,7 no primeiro andar entre os valores medidos e simulados. Já no segundo, foi aplicado o traçado de raios para avaliar o perfil de potência e retardo, além de técnicas de *machine learning* para encontrar posições dentro do ambiente. Esses resultados validam a metodologia utilizada e sugerem a viabilidade de explorar outros parâmetros em pesquisas futuras.

O objetivo deste trabalho é criar um modelo 3D do saguão de desembarque do Aeroporto de Val-de-Cans, em Belém-Pará, para validar os resultados obtidos em campanhas de medição realizadas no local, conforme descrito em [4]. A ferramenta permite simular as características físicas do ambiente para realizar cálculos de propagação de ondas eletromagnéticas. Dessa forma, busca-se validar o modelo e utilizá-lo para avaliar os modelos de propagação presentes na literatura, possibilitando entender o desempenho da frequência de 3,5 GHz em ambientes internos.

II. ANÁLISE DE LARGA ESCALA

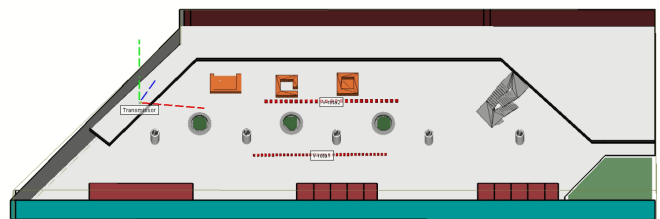


Figura 1. Ambiente de simulação

A análise do canal em larga escala envolve a compreensão dos fatores que influenciam no valor perda de propagação do ambiente. A Fig. 1 ilustra o ambiente desenvolvido para as simulações.

O ambiente é composto por materiais como ferro (cinza), madeira (laranja), cimento (cinza claro), tijolo (vermelho) e plantas (verde). Cada um desses materiais tem suas próprias características físicas como permissividade e condutividade, que podem ser alteradas de acordo com informações existentes na literatura.

Para comparar o desempenho da simulação com os valores reais (medidos) é utilizado o RMSE [2]. Para a rota 1, o RMSE

é 5,09, e para a rota 2, é 4,31. A Fig. 2 e Fig. 3 mostram os valores medidos em vermelho e os simulados em azul para as rotas 1 e 2, respectivamente.

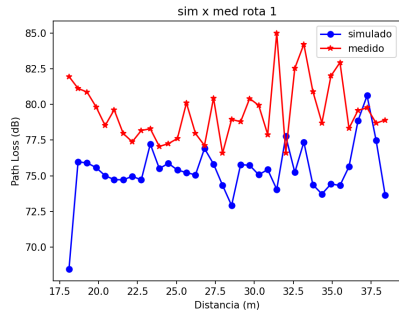


Figura 2. Rota 1.

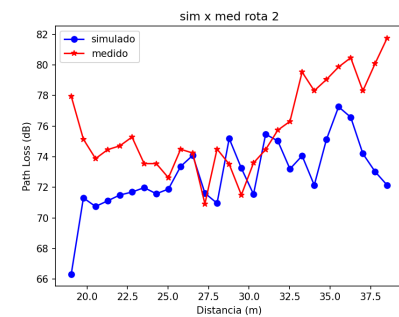


Figura 3. Rota 2.

As rotas foram analisadas para ambos os modelos, CI e FI, para encontrar os valores de PLE (*Path Loss Expoent*) para CI, α e β para o FI, necessários para medir a performance da regressão linear com os dados simulados através do RMSE. A Fig. 4 e Fig. 5 mostram as curvas dos modelos.

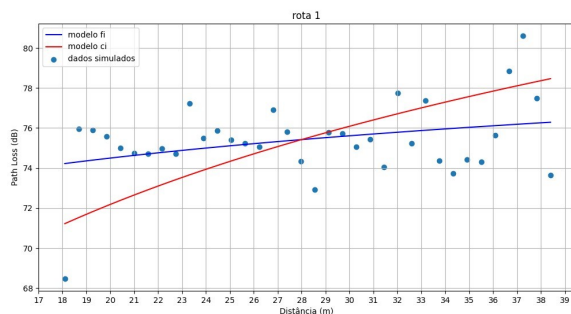


Figura 4. Modelos CI e FI para a Rota 1.

O modelo CI em vermelho tem um PLE de 2,21 e RMSE de 2,34. O modelo FI em azul apresenta α de 66,27, β de 0,632 e RMSE de 1,78. A perda se mantém entre 73 e 78 dB, devido a composição do ambiente próximo.

Na rota 2, o PLE do modelo CI foi 2,03 e o RMSE igual à 1,61. No modelo FI, o valor de α foi 48,89, β igual à 1,65

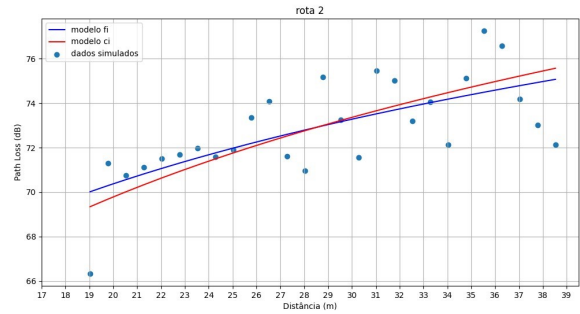


Figura 5. Modelos CI e FI para a Rota 2.

e o RMSE de 1,57. Há maior variação na perda a partir de 25 metros do transmissor devido à interação dos raios com estruturas de madeira, bancos de cimento com plantas, escada e segundo piso do hall de desembarque.

III. CONCLUSÕES

As simulações realizadas no saguão do Aeroporto Val de Cans utilizando o *software* WI revelaram uma diferença de 5,09 e 4,31 entre os dados medidos e os simulados nas rotas 1 e 2, respectivamente. Esses resultados confirmam que a simulação é condizente com o cenário real, validando a eficácia dos modelos de propagação utilizados. Além disso, os resultados obtidos abrem caminho para análises posteriores de outras métricas de comunicação, permitindo uma compreensão mais aprofundada da propagação do sinal em ambientes internos e possibilitando a otimização do desempenho da rede. Pretende-se ampliar o modelo 3D para posterior análise e avaliação dos parâmetros de pequena escala através da aplicação do modelo de traçado de raios, assim ampliando a robustez do modelo 3D e análises mais profundas em relação a projetos de rádio propagação.

REFERÊNCIAS

- [1] MacCartney, George R., et al. "Path loss models for 5G millimeter wave propagation channels in urban microcells." 2013 IEEE global communications conference (GLOBECOM). IEEE, 2013.
- [2] A. B. S. Correa, L. M. Ribeiro, F. H. C. da S. Ferreira, C. M. M. Cardoso, F. F. Ribeiro, J. P. L. Araújo, M. C. de A. Neto, F. J. B. Barros. "Medidas e Modelo de Traçado de Raios para Ambiente Indoor em 3.5GHz". ENCOM 2023. Disponível em: <https://www.iecom.org.br/encom2023/autores.html#anais>.
- [3] Clark, R. T., Engel, A. W., Shenk, T. E., Sheen, D. M. (2023, January). Indoor Wireless Localization of Uncooperative Sources Using a Ray Tracing Model. In 2023 IEEE Topical Conference on Wireless Sensors and Sensor Networks (pp. 68-70). IEEE.
- [4] Macedo, A. S., de Araujo Costa, T., de Matos, E. M. C., ERAS, L. E. C., de Araujo, J. P. L., Gonzalez, P. V. G. C. V., Barros, F. J. B. (2023). Channel analysis for 3.5 GHz frequency in airport. Journal of Communication and Information Systems, 38(1), 115-120.
- [5] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor Office Wideband Millimeter Wave Propagation Measurements and Channel Models at 28 and 73 GHz for Ultra Dense 5G Wireless Networks," IEEE Access, vol. 3, pp. 2388–2424, 2015. [Online]. Available: <https://doi.org/10.1109/access.2015.2486778>.

Estimativa de Parâmetros de Pequena Escala com Traçado de Raios em Ambiente 3D em 3,5 GHz

Mozart L. M. Junior

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
juniorlamaquias09@gmail.com

Fabício J. B. Barros

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
fbarros@ufpa.br

Alex S. Macedo

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
alexsanchess@gmail.com

Jasmine P. L. Araújo

Laboratório de Computação e Telecomunicações (LCT)
Universidade Federal do Pará (UFPA)
Belém, Brasil
jasmine@ufpa.br

Abstract—A faixa de 3,5 GHz é amplamente utilizada para a implementação do 5G, sendo considerada intermediária por proporcionar boa cobertura em áreas urbanas, suburbanas e internas. Este artigo propõe a estimativa de parâmetros de pequena escala utilizando traçado de raios em um ambiente indoor 3D de aeroporto a 3,5 GHz, com base em dados de campanhas de medição. As simulações mostraram conformidade com as medições locais, incluindo *Power Delay Profile* (PDP), *delay spread* e *mean delay* no ponto de recepção analisado. Com os resultados obtidos, foi possível validar o cenário 3D construído via *software Wireless Insite* (WI) para aplicações em análises de sistemas de comunicações móveis.

Index Terms—Micro-Ondas, Wireless Insite, Traçado de Raios, modelagem 3D, Pequena Escala.

I. INTRODUÇÃO

Ambientes *indoor* são foco de estudos de propagação de sinais de alta e baixa frequência por conta dos efeitos do ambiente na intensidade do sinal. A referência [1] utiliza uma banda de 290 GHz - 400 GHz e retorna à capacidade do canal de acordo com a condição de visada direta ou não e o tipo de obstrução utilizada no experimento. No estudo proposto por [2], analisando a Banda D (110 GHz - 170 GHz) para ambientes fechados, conseguindo apresentar resultados de perda de penetração média para cada tipo de material encontrado no ambiente (gesso, concreto e vidro) do experimento dado a sua espessura.

Em baixas frequências, a exemplo de [3], o WI foi utilizado para a obtenção de perfil de potência e retardo na faixa de 10 GHz. Já em [4], o WI foi utilizado para reproduzir um ambiente indoor de dois andares. A partir das simulações, foi possível reproduzir o comportamento do sinal propagado no ambiente na faixa de 3,5 GHz, com uma raiz do erro quadratico médio REQM de 4,1 dB no térreo e 8,7 dB no primeiro andar. Esses resultados indicam que o modelo foi eficaz em replicar o comportamento do sinal, validando a metodologia empregada.

Este trabalho propõe uma avaliação de pequena escala utilizando um modelo de traçado de raios. Para isso, foi criado um cenário 3D do saguão de desembarque do Aeroporto de Val-de-Cans, em Belém-Pará, na plataforma do WI, com o objetivo de validar os resultados medidos em campanhas realizadas no local. Pretende-se aplicar parâmetros de transmissão e recepção utilizando dados reais coletados localmente, na faixa de frequência de 3,5 GHz, que é utilizada no 5G no Brasil [5]. Dessa forma, é possível verificar parâmetros de pequena escala, como perfil de potência e retardo.

II. AMBIENTE IMPORTADO NO SOFTWARE

A construção do saguão de desembarque seguiu a planta baixa adquirida em parceria com a direção do aeroporto. O saguão, com uma área de 1480 m², inclui quiosques, escadas, bancos, estruturas metálicas e etc. A Figura 1 ilustra a localização das rotas e do transmissor. Foram criadas duas rotas de 20 metros para a medição, com pontos espaçados em 0,58 metro e 0,76 metro para as rotas 1 e 2, respectivamente.



Fig. 1. Ambiente real [5].

O , é uma ferramenta de alto desempenho e fidelidade, desenvolvida pela Remcom, no qual pode-se construir cenários, elementos irradiantes e alterar as propriedades físicas de materiais de forma que os resultados da simulação sejam os mais verossímeis o possível de acordo com o grau de compatibilidade do ambiente importado com a realidade [4].

A Tabela 1 indica as características elétricas e as cores dos materiais utilizados.

TABLE I
IDENTIFICAÇÃO DOS MATERIAIS.

	Tijolo	Vidro	Madeira	Concreto	Ferro
<i>cor</i>	vermelho	azul	laranja	cinza	branco
ϵ	4.4	4.0	2.0	5.24	3.0

As estruturas metálicas localizadas no meio do cenário, por padrão do *software* são definidas como PEC (*Perfect Electrical Conductor*), ou condutores metálicos perfeitos, são elementos que refletem totalmente a onda incidente, mas tais elementos são apenas teóricos, por isso foi definido valores de permissividade relativa para os mesmos [4].

A construção do ambiente 3D levou em consideração apenas as estruturas listadas na planta baixa adquirida, incluindo bancos, balcões de atendimento e áreas de espera designadas. Foi criada uma onda senoidal em 3,5 GHz com largura de banda de 60 MHz, utilizando antenas transmissora (TX) e receptora (RX). Para a antena TX, foi utilizada uma do tipo *Horn* (Corneta), com ganho de 16 dBi, enquanto para a RX foram usadas antenas omnidirecionais com ganho de 5 dBi, ambas com polarização vertical [5]. O transmissor possui uma potência de entrada de 51 dBm e está a uma altura de 7,9 metros. Os receptores estão a uma altura de 1,78 metros e se movem a uma velocidade de 0,58 m/s.

O modelo de análise utilizado foi o *FULL 3D* com traçado de raios *Shooting and Bouncing rays* (SBR), com 6 reflexões e 3 transmissões, esta técnica permite que várias simulações sejam realizadas sem um custo computacional muito grande. A configuração do computador no qual o programa foi executado é: CPU(s) Intel(R) Xeon(R) Silver 4210 CPU @ 2.20 GHz com 64 GB de memória RAM.

III. ANÁLISE DE PEQUENA ESCALA

A análise dos raios de comunicação envolve a compreensão da propagação de ondas eletromagnéticas, conforme descrito pela ótica geométrica. Para modelagem de canal em pequena escala, utilizamos o *Power Delay Profile* (PDP) para descrever o funcionamento do canal em determinado local, determinando a potência instantânea recebida de cada componente multipercorso [6].

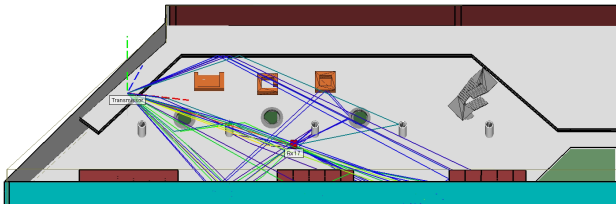


Fig. 2. Distribuição de raios no RX 17.

A Figura 2 ilustra os raios chegando ao ponto 17 da rota 1, com intensidades variando conforme a cor, sendo amarelo e verde para valores maiores e azul para menores.

A Figura 3 mostra o PDP do ponto de RX 17. O primeiro raio, com menor atraso, mantém visada direta entre TX e RX, com PDP de -24 dBm e *delay spread* de 1,98 nanosegundos (ns). O segundo raio reflete no vidro da entrada do aeroporto. O terceiro raio, com menor intensidade, sofre transmissão em uma estrutura e duas reflexões. O quarto raio, apesar de maior atraso devido à distância, experimenta menor desvanecimento, com reflexões no vidro e no piso. No geral, o ponto 17 apresenta um *mean delay* de 6,38 ns. Os resultados foram considerados condizentes com os dados medidos apresentados em [5].

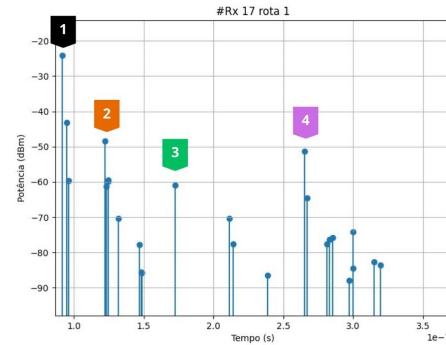


Fig. 3. PDP no Rx 17.

IV. CONCLUSÕES

Este estudo propôs uma metodologia para analisar e estimar parâmetros de pequena escala utilizando traçado de raios. O principal desafio foi criar um ambiente 3D altamente realista, ajustando os parâmetros dos materiais e de propagação para identificar os fatores que mais influenciam o comportamento do sinal. As simulações realizadas permitiram reproduzir o comportamento do sinal através do traçado de raios. Os resultados indicam que o modelo conseguiu validar a metodologia empregada, abrindo caminho para a investigação de outros parâmetros em pesquisas futuras.

REFERENCES

- [1] Khalid, N., Akan, O. B. (2016, May). Wideband THz communication channel measurements for 5G indoor wireless networks. In 2016 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.
- [2] Xing, Y., Rappaport, T. S. (2018, December). Propagation measurement system and approach at 140 GHz-moving to 6G and above 100 GHz. In 2018 IEEE global communications Conference (GLOBECOM) (pp. 1-6). IEEE.
- [3] Barbosa, B. S. S.; et.al. Metodologia para Obtenção de Perfil de Potência e Retardo na Faixa de 10 GHz. In: VIII Conferência Nacional em Comunicações, Redes e Segurança da Informação, 2018, Salvador, BA. Anais da VIII Conferência Nacional em Comunicações, Redes e Segurança da Informação, Encom 2018, 2018.
- [4] A. B. S. Correa, et.al. "Medidas e Modelo de Traçado de Raios para Ambiente Indoor em 3.5GHz". ENCOM 2023. Disponível em: <https://www.icom.org.br/encom2023/autores.html#anais>.
- [5] Macedo, A. S., et. al. Channel analysis for 3.5 GHz frequency in airport. Journal of Communication and Information Systems, 38(1), 115-120.
- [6] M. E. C. Rodrigues, "Técnicas de Tracado de Raios em Três Dimensões para Cálculo de Campos em Ambientes Interiores e Exteriores," CETUC, PUC-Rio, RJ, 2000.

Importância da Normalização na Geração de Amostras de Desvanecimento η - μ

1st Samuel Rodrigues de Miranda
Centro de Energias Alternativas e Renováveis (CEAR)
Universidade Federal da Paraíba (UFPB)
João Pessoa, Brasil
samuel.miranda@estudante.cear.ufpb.br

2nd Gustavo Fernandes Souza
Centro de Energias Alternativas e Renováveis (CEAR)
Universidade Federal da Paraíba (UFPB)
João Pessoa, Brasil
gustavof.souza@estudante.cear.ufpb.br

3rd Waslon Terllizzie Araújo Lopes
Centro de Energias Alternativas e Renováveis (CEAR)
Universidade Federal da Paraíba (UFPB)
João Pessoa, Brasil
waslon@cear.ufpb.br

4th Hugerles Sales Silva
Departamento de Engenharia Elétrica
Universidade de Brasília (UnB)
Distrito Federal, Brasília
hugerles.silva@unb.br

Resumo—Em comunicações móveis, o fenômeno de desvanecimento aparece quando o sinal transmitido chega ao receptor por meio de múltiplos percursos. A distribuição generalizada η - μ , proposta por Yacoub em 2007, modela importantes cenários de transmissão e por meio da escolha adequada dos parâmetros η e μ pode representar casos particulares, como o desvanecimento Rayleigh. Neste artigo é discutida a importância da normalização da potência das amostras de desvanecimento para a determinação correta da taxa de erro de bit em canais com desvanecimento.

Palavras-chave—Desvanecimento η - μ , fator normalizante, comunicação 5G.

I. INTRODUÇÃO

O desvanecimento do sinal é um fenômeno de suma importância nas comunicações sem fio, afetando de forma direta a confiabilidade e a qualidade do canal e dos dados nele transmitidos. Diversos fatores influenciam na ocorrência desse fenômeno, como obstáculos físicos, reflexões, refrações e interferências, causando variações na potência do sinal recebido [1].

Dessa maneira, entender e modelar o desvanecimento é essencial para desenvolver técnicas que mitiguem seus efeitos e permitam uma comunicação robusta garantindo o pleno aproveitamento da tecnologia de comunicação móvel de quinta geração (5G) e além (B5G).

O desvanecimento pode ser modelado por distribuições simples, como a Rayleigh e a Nakagami, ou por distribuições generalistas como a η - μ , proposta por Yacoub em 2007 [1], trata-se de um caso particular da distribuição generalizada α - η - κ - μ [2]. Com esta distribuição particular, é possível modelar a variação do desvanecimento em pequena escala de um sinal transmitido sem linha de visada (NLOS – *Non Line-of-Sight*) entre o transmissor e o receptor. Nesse cenário, a propagação se dá por meio de múltiplos percursos e haverá uma disparidade entre as componentes em fase e em quadratura do sinal transmitido, tendo como resultado valores distintos para a variância dessas partes.

Neste contexto, este artigo discute a importância da normalização da potência das amostras de desvanecimento η - μ para a determinação correta da taxa de erro de bit em canais com desvanecimento.

II. MODELO DE DESVANECIMENTO η - μ

A envoltória do sinal recebido através de um canal com desvanecimento η - μ é dada por [3]

$$R^2 = \sum_{i=1}^{\mu} (F_i^2 + G_i^2), \quad (1)$$

sendo F_i e G_i amostras gaussianas independentes, $E[F_i] = E[G_i] = 0$ e $E[F_i^2] = \sigma_F^2$, $E[G_i^2] = \sigma_G^2$ e $\hat{r} = \sqrt{E[R^2]}$.

Devido à ambivalência do parâmetro η , essa distribuição de desvanecimento possui dois formatos, nomeados de Formato I e II. Este trabalho considerou o Formato I em que o parâmetro η é definido como a razão entre as componentes em fase e quadratura, componentes estas, que são independentes e possuem potências distintas. O parâmetro μ trata-se da quantidade de clusters derivados dos múltiplos percursos. Embora a envoltória da distribuição seja advinda de um somatório, o parâmetro μ pode assumir qualquer valor real.

Para a geração dos coeficientes, é necessário encontrar as variâncias das duas distribuições gaussianas que compõem a envoltória do sinal. Sendo assim, é possível encontrar um desses parâmetros utilizando a relação de σ_G^2 dada por [4]

$$\sigma_G^2 = \frac{\hat{r}^2}{2\mu(1 + \eta)}. \quad (2)$$

Em seguida, o valor de σ_F^2 pode ser encontrado por meio da própria definição do parâmetro η dada por

$$\eta = \frac{\sigma_F^2}{\sigma_G^2}. \quad (3)$$

III. NORMALIZAÇÃO DA ENVOLTÓRIA DO DESVANECIMENTO

A distribuição de desvanecimento η - μ , quando gerada, apresenta um valor médio quadrático não normalizado, o que não é desejado em simulações no qual a potência média do sinal transmitido não deve ser alterada.

Portanto, é necessário que as N amostras x_i de R sejam divididas por um fator normalizante a fim de tornar o somatório dos coeficientes da envoltória x_i ao quadrado dividido pela quantidade de coeficientes igual a 1. Assim, para encontrar o fator de normalização, pode-se utilizar a seguinte expressão

$$\sum_{i=1}^N \left(\frac{x_i^2}{N} \right) = P. \quad (4)$$

O resultado do somatório é o fator normalizante, representado por P , utilizado na normalização da envoltória, dividindo cada amostra, x_i , de R por $\sqrt{P}$, originando a nova envoltória R normalizada.

IV. MODELO DO SISTEMA

No modelo do sistema de comunicações adotado neste trabalho, o sinal recebido, em um intervalo de símbolo, é dado por

$$r = \alpha \cdot s + n, \quad (5)$$

em que r é o sinal recebido, α é o coeficiente de desvanecimento, s é o sinal transmitido e n é o ruído aditivo gaussiano branco.

Considerando a modulação BPSK, o sinal transmitido s pode assumir os valores ± 1 na sua representação passabaixa [5].

V. RESULTADOS

Foram implementadas rotinas em C++ para geração de amostras do desvanecimento η - μ usando as expressões da Seção II. No primeiro conjunto de simulações, foram geradas 10^6 amostras da distribuição η - μ e, após, foi feita a normalização das amostras. As Figuras a seguir apresentam os histogramas com e sem normalização, considerando $\eta = 1$ e $\mu = 1$. Pode-se observar, então, que a potência (valor médio quadrático) das amostras estão distribuídas em torno de 1 na Figura 2.

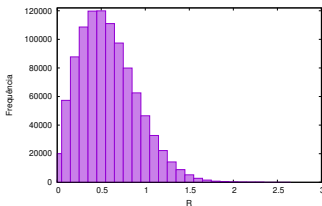


Figura 1: Sem normalização.

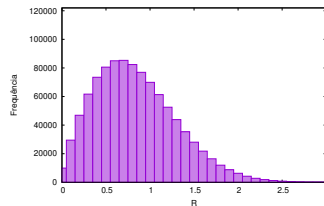


Figura 2: Normalizada

O segundo conjunto de simulações consistiu na transmissão de sinais BPSK por um canal sujeito ao desvanecimento η - μ e ruído aditivo. A Figura 3 apresenta a taxa de erro de bit em termos da relação sinal-ruído por bit (E_b/N_0).

Considerando $\eta=\mu=1$, pode-se observar que há uma diferença de aproximadamente 3 dB entre as curvas normalizada e não-normalizada, conforme esperado, pois $\sigma_F^2 = \sigma_G^2 = 1$.

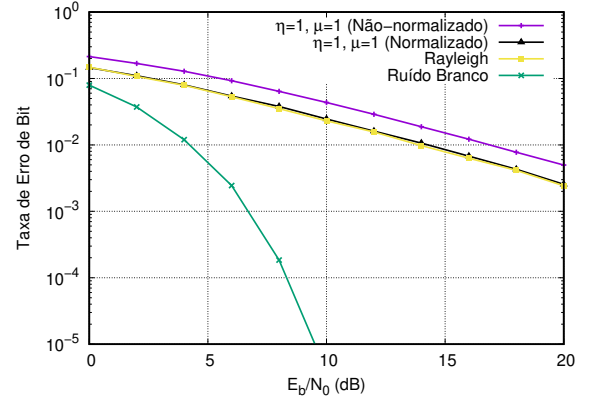


Figura 3: Taxa de erro de bit em função da relação sinal-ruído.

Para fins de comparação, também estão traçadas na Figura 3 as curvas relativas a taxa de erro de bit considerando o canal com desvanecimento Rayleigh e para o canal contaminado apenas com o ruído Gaussiano. É importante ressaltar que a equivalência entre o desempenho do sistema para o desvanecimento Rayleigh e para o desvanecimento η - μ , pois, de fato, o desvanecimento Rayleigh é um caso particular do desvanecimento η - μ , quando $\eta = \mu = 1$ [6].

VI. CONCLUSÃO

Este trabalho abordou o problema da geração de amostras de desvanecimento η - μ com potência normalizada, isto é, $E[R^2] = 1$. Por meio de simulações, foi mostrado que a taxa de erro de bit de sistemas de comunicação depende fortemente desta normalização. Em particular, um erro de 3 dB foi verificado entre o desempenho entre o sistema normalizado e o não normalizado. Adicionalmente, os resultados obtidos com $\eta = \mu = 1$ coincidiram com o desvanecimento clássico do tipo Rayleigh, conforme esperado.

REFERÊNCIAS

- [1] M. D. Yacoub, "The κ - μ Distribution and the η - μ Distribution". *IEEE Antennas and Propagation Magazine*, vol. 49, no. 1, pp. 68–81, February 2007.
- [2] A. A. dos Anjos, T. R. R. Marins, R. A. A. de Souza and M. D. Yacoub, "Higher Order Statistics for the α - η - κ - μ Fading Model". *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 6, pp. 3002–3016, March 2018.
- [3] F. C. P. Barros, F. C. M. Oliveira, M. L. M. G. Alcoforado and W. T. A. Lopes, "Performance of Polar Codes over Generalized Fading Channels". *Journal of Communication and Information Systems*, vol. 37, no. 1, pp. 1–9, February 2022.
- [4] F. Yang, H. Mao, X. Zeng, X. Jian, S. Fu, B. Wu, X. Tan and J. Zhou, "Nondata-Aided Error Vector Magnitude Analysis of η - μ Fading Channels in Device-to-Device Communications". *IEEE Access*, vol. 7, pp. 82101–82109, 2019.
- [5] B. P. Lathi and Z. Ding, *Modern Digital and Analog Communication Systems*. Oxford University Press, New York, fourth edition, 2010.
- [6] T. R. R. Marins, "Estatísticas de Ordem Superior para o Modelo de Envoltória α - η - κ - μ ". Tese de doutorado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, SP, 2020.

Machine Learning-Based Methods for CQI Prediction in Mobility 5G Scenarios

Debora S. Barros, Janderson C. Meireles, Bruno S. da Silva, Daniel da S. Souza, Hugo O. da Cruz and José M. Filho

Abstract—In 5G systems, the user equipment (UE) reports channel quality indicator (CQI) parameters to the access network. This indicator, which relates to the status of the propagation channel, is used as an input parameter to the packet scheduler when resource blocks need to be allocated. This paper proposes an approach to predict the CQI by obtaining internal measurement data from the 5G UE modem in mobility scenarios and applying multivariate time series forecasting techniques to predict the next samples for the CQI, thereby optimizing and accelerating the allocation of resources. To predict the CQI, four models were used, the first based on Long Short-Term Memory (LSTM), a deep learning approach of RNN (Recurrent Neural Network), and the other using a classic machine learning algorithms: Adaptive Boosting (AdaBoost), Random Forest (RF) and Support Vector Machine (SVM). The LSTM technique proved to be the best choice, achieving the lowest RMSE (0.87) compared to the others.

Keywords—5G, CQI, Prediction, LSTM, Adaboost, Random Forest, SVM.

I. INTRODUCTION

The 5G New Radio mobile communication system provides support for various scenarios, from ultra-high data transfer rates to the possibility of ultra-low latency communication. During downlink transmission, the 5G base station (gNB) needs to allocate available resources to different users based on the channel condition and select the appropriate modulation and coding scheme (MCS) for each UE connected to the network. The UE reports channel quality indicators (CQIs) to gNB. The CQI index is an input parameter to the packet scheduler that ranges from 0 to 15, where higher is better. The reported CQI values provide the requirements for resource allocation decisions. Users with good channel quality would achieve better communication performance at high data rates [1].

Assuming that the CQI is transmitted correctly and the channel does not change abruptly, the selected MCS must ensure robust and high-quality communication and keep the block error rate (BLER) below 10%, a quality standard used for many applications. However, a correctly received CQI does not predict a sudden change in channel condition, which may make this CQI redundant and cause the network to incorrectly allocate resources, directly affecting the spectral efficiency and increasing BLER.

SIDIA Institute of Science and Technology, Manaus-AM, e-mail: {debora.barros-e, janderson.correa, bruno.satiro, daniel.ssilva, hugo.cruz}@sidia.com and José M. Filho, Universidade Tecnológica Federal do Paraná, {jmiracy@alunos.utfpr.edu.br}. This work was presented as part of the result of Project AMAN, executed by Sidia Institute of Science and Technology in partnership with Samsung Eletrônica da Amazônia LTDA, according to Informatics Law n.8387/91 and Art.39 of Decree 10.521/2020.

Various problems with delayed CQI and possible solutions are described in the literature [2], [3]. The aim of this work is to use machine learning (ML) techniques to predict CQI index in a mobility scenario based on the parameters provided over the time series by the UE modem. Our approach differs from others dealing with the same problem because we propose to predict future values of the CQI index with certain approximations that reflect the state of the channel directly sensed by the UE modem.

In addition to this introduction, the paper is organized as follows: Section II presents the data and ML techniques used. The prediction results and comparison between the ML techniques are presented in Section III. Finally, Section IV contains the conclusion.

II. METHODOLOGY

For the development of the analysis, four ML techniques - AdaBoost, SVM, RF and LSTM - were used to predict CQI index for 100 future steps, taking as input metrics reference signals: Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Received Signal Strength Indicator (RSSI), Signal Interference Noise Ratio (SINR) and a metric related to MCS network resources allocation, extracted from the on-site UE log. The signal dataset was obtained from field tests carried out in some cities in Latin America, primarily in Brazil. Field tests are carried out in which the UE modem runs through various scenarios. During the field test, 5G signals are processed in the UE modem and metrics are stored. This process has been systematized for several smartphones, 3200 samples were collected, which corresponds to a network time of 128000 milliseconds and can be used for prediction, 80% being used for training and 20% testing.

The time series forecasts through making predictions of 100 future values using 50 samples past observations. By implementing the sliding window approach, time series data can be transformed into a supervised learning problem by generating overlapping windows of a predefined size. This prediction experiment was carried out thirty times for each technique in order to obtain a statistically reliable result. Furthermore, four machine learning techniques were selected to carry out this experiment.

The AdaBoost enhances weak learners by combining them into a strong predictor. This model iteratively trains weak regressors, adjusting weights based on prediction errors to focus on harder-to-predict segments [4]; The SVM finds the optimal hyperplane that best separates these vectors based on their future values. The kernel trick handles non-linear

relationships by transforming data into higher-dimensional space; The Random Forest technique constructs multiple trees using random feature subsets to ensure diversity. Final predictions are averaged across all trees, reducing overfitting and improving generalization [5]; The LSTM is a type of recurrent neural network, which captures long-term dependencies in sequential data. LSTM networks predict future values by retaining information over longer sequences, making them powerful for complex time series with seasonality and trends [6]. In order to find the best configurations of the hyperparameters, several configurations were tested for the different techniques.

III. RESULTS AND DISCUSSION

To evaluate the results obtained (measured and predicted), the Root Mean Squared Error (RMSE) metric was defined, which is the square root of the mean squared difference between the actual value of the data set and the predicted values of the model. Standard deviation is an assessment precision that shows how much the values deviate from the mean of the data set, and the mean absolute error (MAE), which represents the difference in the absolute mean between the actual values of the data set and the values predicted by the model.

Considering the lowest RMSE (y-axis) found between the measured and predicted data for the 30 tested seeds (x-axis), Fig. 1 compares the four implemented techniques. It can be seen that the LSTM approach outperformed the others and produced the best result in Seed 6, with the lowest RMSE value of 0.87 and the following hyperparameters: 0.00437 learning rate, 0.1 dropout, activation function “tanh”, 2 layers (256, 128 neurons) and optimizer “adam”.

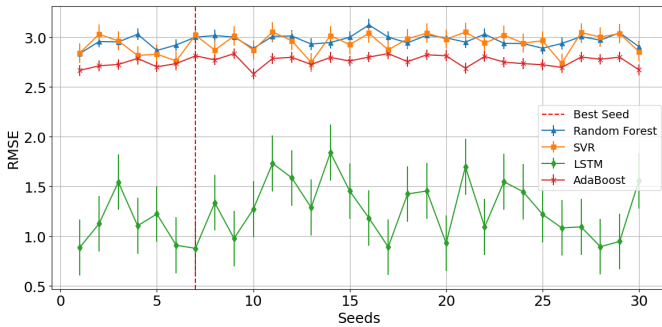


Fig. 1. Comparison of ML techniques by seed.

The average values obtained from the metrics evaluated by the 30 seeds for the ML techniques are listed in Table I.

TABLE I
AVERAGE EVALUATION OF THE 30 SEEDS TO ESTIMATE CQI

Technique	RMSE	STD Error	MAE
Random Forest	2.96	2.96	2.21
SVR	2.94	2.93	2.23
LSTM	1.25	0.75	1.248
AdaBoost	2.75	2.75	2.08

The prediction result using the best trained model (LSTM) is presented in Fig. 2, showing that it is possible to predict 100 future samples with an RMSE of 0.87 using only 50 historical samples, which is considered good in a mobility scenario, improving the network’s ability to predict upcoming events.

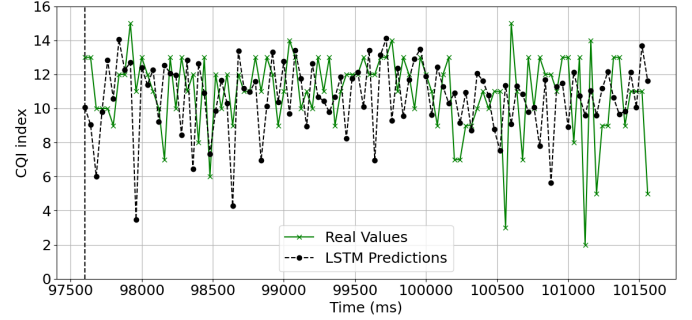


Fig. 2. Comparison of the Real and Predicted values of the best model.

The LSTM network excels in time series forecasting due to its ability to handle long-term dependencies. This is achieved through its input, forget and output gates that regulate the flow of information. Unlike algorithms like Random Forest, AdaBoost, and SVM, that treat each data point independently and cannot capture temporal sequences, LSTM processes sequential data, dynamically maintaining and adjusting memory. This allows LSTM to learn complex temporal patterns and make more accurate predictions by effectively using information over time.

IV. CONCLUSIONS

In this paper, a machine learning approach was proposed to predict CQI in 5G networks. CQI prediction is a very helpful function, especially in mobility scenarios where the propagation channel may suddenly change. Using variables extracted from the UE modem, the proposed method uses 50 samples from the past to 100 prediction samples in the future. The network can have 4.000 ms future information, allowing for better resource allocation and more predictability in decision making.

REFERENCES

- [1] M. Fuentes, J. L. Carcel, C. Dietrich, L. Yu, E. Garro, V. Pauli, F. I. Lazarakis, O. Grøndalen, Bulakci, J. Yu, W. Mohr, and D. Gomez-Barquero, “5g new radio evaluation against imt-2020 key performance indicators,” *IEEE Access*, vol. 8, pp. 110 880–110 896, 2020.
- [2] A. Balieiro, K. Dias, and P. Guarda, “Addressing the CQI feedback delay in 5G/6G networks via machine learning and evolutionary computing,” *Intell. and Converged Netw.*, vol. 3, no. 3, pp. 271–281, Sep. 2022.
- [3] P. S. Pati, S. S. Sahoo, D. Krishnaswamy, and R. Datta, “A novel machine learning approach for link adaptation in 5g wireless networks,” in *2020 2nd PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS)*, 2020, pp. 1–2.
- [4] R. E. Schapire, “Explaining AdaBoost,” in *Empirical Inference*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–52.
- [5] K. Faceli, A. C. Lorena, J. Gama, T. A. d. Almeida, and A. C. P. L. F. d. Carvalho, *Inteligência artificial: uma abordagem de aprendizado de máquina*. LTC, 2021.
- [6] R. C. Staudemeyer and E. R. Morris, “Understanding lstm—a tutorial into long short-term memory recurrent neural networks,” *arXiv preprint arXiv:1909.09586*, 2019.

Proposta de Melhoria no Processo de Testes de Personalização dos Serviços de IMS em Dispositivos Móveis

Gaspar H. A. Mota*, Flávia C. M. de Oliveira* e André J. de França*

Resumo—Neste trabalho apresenta-se a análise de um novo procedimento de testes no tocante à personalização dos serviços do IP Multimedia Subsystem (IMS) em dispositivos móveis. Com a crescente demanda por novos dispositivos e tecnologias, a disponibilidade de um laboratório especializado, torna-se um gargalo no processo de entrega do produto. A partir da automação de uma parte do escopo de testes, foi possível reduzir o tempo de execução das tarefas destinadas ao time SEL e, por consequência, foi possível obter um aumento na disponibilidade do laboratório. No contexto brasileiro, em um cenário de alta demanda, o uso desse novo procedimento pode proporcionar uma redução 24,5% no tempo de entrega de uma requisição de teste.

Index Terms—Teste de Software, Automação de Testes, Melhoria de Processo, Dispositivos Móveis.

I. INTRODUÇÃO

O IP Multimedia Subsystem (IMS) proporciona às operadoras a estrutura necessária para a implementar diversos serviços de comunicações, como: multimídia, dados, voz e vídeo. Sua concepção foi direcionada às preferências dos usuários e já prevê o acesso à múltiplas redes móveis, bem como o uso concomitante de vários destes serviços [1]. Com o advento da quinta geração de redes móveis (5G), que visa garantir a conectividade com alta confiabilidade para um grande número de dispositivos, novos serviços estão sendo adicionados, como por exemplo Voz sobre New Radio (VoNR) e Vídeo sobre New Radio (ViNR) [2]–[4]. Os dispositivos móveis mais recentes já suportam a rede 5G e, além disso, já acessam os seus serviços de IMS. No entanto, o desenvolvimento do *software* neles embarcados requer uma etapa de testes, que visa garantir a qualidade do produto [3].

Estas novas tecnologias necessitam de laboratórios especializados para testes. Esses, por sua vez, devem permitir simulações de redes, interação com dispositivos móveis e garantir um ambiente perfeitamente controlado, livre de interferências externas [4]. Além disso, faz-se necessária uma equipe com capacidade de realizar testes e garantir que as funcionalidades do *software* e todos os requisitos dos clientes (operadoras) estejam corretamente aplicados [5]. Esse é o papel do time conhecido como *Software Engineering Lab* (SEL) que realiza um conjunto de testes nos dispositivos do fabricante a cada lançamento de um novo modelo ou a cada nova atualização do sistema operacional (SO) Android¹ [6].

O presente trabalho contempla uma automação que foca no acesso e na extração de requisitos e funcionalidades de IMS.

*Sidia Instituto de Ciência e Tecnologia. Manaus, Amazonas, Brasil. E-mails: gaspar.mota@sidia.com, flavia.oliveira@sidia.com e andre.franca@sidia.com

¹<https://www.android.com/>

A proposta é avaliar o uso dessa nova abordagem no escopo de testes das operadoras brasileiras e, dessa forma, evidenciar a viabilidade de seu uso em um cenário de alta demanda, visando a diminuição do tempo de entrega das requisições de testes e o aumento da disponibilidade do laboratório.

Na próxima Seção, tem-se um resumo das atividades de testes desempenhadas pelo time SEL no tocante aos serviços de IMS.

II. TESTES DE SOFTWARE NOS SERVIÇOS DE IMS

As demandas atribuídas ao SEL são bastante elevadas devido à quantidade de novos dispositivos lançados pelo cliente e, também, devido às atualizações de SO. Como o tempo de execução dos testes varia de acordo com a quantidade de serviços que os dispositivos suportam e que as operadoras oferecem, o número de requisições de testes torna-se proporcional e, consequentemente, eleva-se o tempo de execução dentro do laboratório. O escopo de teste é dividido em duas partes: a funcionalidade dos serviços e a validação da personalização. A Tabela I mostra o tempo médio em minutos de execução, segundo os serviços e a personalização das operadoras no cenário do Brasil. Os valores são referidos ao tempo gasto no processo padrão de teste executado pelo time do SEL, em média são gastos, 50min, 40min e 20min respectivos às operadoras A, B e C.

Diante deste cenário, faz-se necessária a busca de alternativas que automatizem o processo de testes e que atuem na validação de requisitos, bem como dos serviços. Nesse trabalho são apresentados a criação e o desenvolvimento de uma automação que agiliza e melhora o processo de testes.

Tabela I
TEMPO MÉDIO DE EXECUÇÃO DOS TESTES DE IMS - ESCOPO BRASIL

Operadora	Tempo [min]	
	Serviço	Personalização
A	30	20
B	30	10
C	15	5

III. TÉCNICA PROPOSTA

A proposta deste trabalho é a utilização de uma automação para extrair informações diretamente do *software* embarcado no dispositivo. Estas informações estão espalhadas dentro do código-fonte e, por meio de sua varredura, deseja-se validar os requisitos das operadoras, que anteriormente eram observados

no momento das simulações em laboratório. Assim, espera-se que o tempo de execução dos testes seja reduzido, uma vez que o uso do laboratório ficaria restrito aos testes de serviços.

A automação consiste em um programa desenvolvido utilizando Python², Java³ e a ferramenta de programação *Android Debug Bridge* (ADB)⁴. Seu funcionamento foi restringido à parte da personalização em que, por meio da varredura do código, procura por marcadores específicos de cada funcionalidade a ser testada. Ao final da execução, é gerado um relatório com as validações efetuadas e os seus respectivos resultados, se os casos de testes passaram ou falharam. Assim, antes do colaborador dirigir-se ao laboratório, deve executar o programa em cada dispositivo e obter o resultado da validação.

Na Seção seguinte são apresentados os resultados que foram obtidos ao ser aplicado o novo processo de teste. Também são realizados cálculos do impacto do uso desta nova abordagem em um quadro de alta demanda.

IV. RESULTADOS

Com o uso da automação foi constatado que o tempo de execução da varredura, validação dos testes e geração do relatório é de, aproximadamente, 1 minuto. Isso faz com que o tempo médio gasto na execução dos testes diminua substancialmente. Observa-se na Tabela II que, em média, são gastos 36min, 31min e 16min respectivos às operadoras A, B e C para um único projeto. Uma observação a ser feita é que o tempo na execução dos serviços se manteve nesse novo processo, dado que eles permanecerão sendo executados no laboratório.

Tabela II
TEMPO MÉDIO DE EXECUÇÃO DOS TESTES DE IMS - NOVO PROCESSO

Operadora	Tempo [min]		Total [min]
	Serviço	Personalização	
A	30	6	36
B	30	1	31
C	15	1	16

Num cenário de alta demanda de testes, se compararmos os dois processos, o impacto dessa melhoria torna-se ainda mais evidenciado. Por exemplo, a perspectiva é que na próxima atualização do Android a gama de projetos que será atendida deve ultrapassar a marca de 30 projetos. Desta forma, como exposto na Tabela III, o time SEL poderá atender aos projetos brasileiros com um total de 41,5h, que representa uma redução de, aproximadamente, 13,5h em comparação ao processo anterior.

A maior redução ficará a cargo da operadora A, que possui mais serviços. Para essa operadora, espera-se entregar todos os 30 projetos com 7h de antecedência, ou seja, o tempo de entrega será reduzido em 28%, quando comparado com o processo anterior. As demais operadoras, em termos percentuais, será possível obter 22,5% e 20%, para as operadoras B e C, respectivamente.

²<https://www.python.org/>

³<https://www.java.com/>

⁴<https://developer.android.com>

Tabela III
COMPARAÇÃO DO TEMPO MÉDIO DE EXECUÇÃO DOS TESTES DE IMS

Operadora	n° de Projetos	Tempo [h]	
		Processo Anterior	Processo Novo
A	30	25	18
B	30	20	15,5
C	30	10	8
TOTAL		55	41,5

V. CONCLUSÃO

Este trabalho apresentou uma proposta de melhoria de processo de testes de serviços de IMS e nas personalizações aplicadas aos dispositivos móveis. Foi visto que, no período de alta demanda, o time SEL sofre com a baixa disponibilidade do laboratório especializado e que, no âmbito das operadoras do Brasil, utilizando-se esta nova abordagem o time pode reduzir o tempo de execução em até 24,5% do tempo previsto para os testes, o que implica diretamente no aumento da disponibilidade do uso do laboratório. Para trabalhos futuros, pretende-se realizar coletas de dados obtidos na prática (após a alta demanda), englobando o escopo da América Latina.

VI. AGRADECIMENTOS

Este artigo é o resultado do projeto de PD&I (AMAN), realizado pela Sidia Instituto de Ciência e Tecnologia em parceria com a Samsung Eletrônica da Amazônia Ltda., usando recursos da Lei Federal nº 8.387/1991, estando sua divulgação e publicidade em conformidade com o previsto no artigo 39.º do Decreto nº 10.521/2020

REFERÊNCIAS

- [1] E. Bertin, I. B. Yahia, and N. Crespi, "Modeling ims services," *Journal of Mobile Multimedia*, pp. 150–167, 2007.
- [2] J. J. A. Arnez and C. B. B. De Souza, "Analysis of inter-mean-opinion-score (imos) and qos metrics for 5g voice over lte (volte) and voice over new radio (vonr) services over ip-multimedia-subsystem (ims) in 5g mobile networks," in *2023 6th International Conference on Advanced Communication Technologies and Networking (CommNet)*. IEEE, 2023, pp. 1–7.
- [3] M. F. De Aguiar, J. K. S. Carvalho, V. D. A. Rodrigues, A. F. Chabi, J. V. D. S. Campos, L. R. Lopes, and J. O. De Sousa, "Performance analysis of 5g voice solutions via eps fallback and vonr," in *2023 IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE, 2023, pp. 1–6.
- [4] S. Pereira Pedrozo, V. Silva Aquino, V. Reboucas Amorim, R. Queiroz Peixe, and P. Cesar Fonseca, "Improving testing process to maximize validation efficiency on advanced telecommunications laboratory: An experience report," in *Proceedings of the 5th International Conference on Management Engineering, Software Engineering and Service Sciences*, 2021, pp. 8–16.
- [5] A. Brígido, C. Souza, F. Moura, M. Reis, A. Neto, and B. Bonifácio, "An industrial case study on applying software testing automated in global software development environment," in *23rd International Conference on Global Software Engineering and Technology on January, 28-29, 2021 at New York, USA (2021)*, 2021, pp. 1456–1459.
- [6] A. J. D. Franca, G. H. A. Mota, K. I. Sahdo, L. Tiago, F. Oliveira, and L. Chaves, "Linkdoc: An automated process in the delivery of documentation in a global software development environment," in *Proceedings of the 2023 5th World Symposium on Software Engineering*, 2023, pp. 9–16.

Análise experimental de conexões ópticas com adaptadores do tipo SC

Rafael Rodrigues de Santana
Laboratório de Telecomunicações
Universidade Federal de Itajubá (UNIFEI)
Itajubá, Brasil
d2021021230@unifei.edu.br

Ramon Maia Borges
Instituto de Engenharia de Sistemas e Tecnologias da
Informação
Universidade Federal de Itajubá (UNIFEI)
Itajubá, Brasil
ramonmb@unifei.edu.br

Resumo—Este trabalho apresenta uma análise experimental de conexões ópticas realizadas com uso de adaptadores do tipo *subscriber* (SC). Tais adaptadores podem ser encontrados comercialmente com as especificações adicionais: monomodo ou multimodo; polimento reto ou angulado. Investiga-se o impacto imposto por estas diferentes versões de adaptadores SC na atenuação de uma dada conexão óptica, estabelecida de forma exemplificativa entre dois cordões de fibra monomodo terminados em conectores do tipo SC. Os resultados obtidos em laboratório indicam que adaptadores SC com especificações distintas não causam alteração significativa na atenuação da conexão óptica, que permaneceu em $\approx 0,2$ dB para conectores de mesmo polimento.

Palavras-chave—adaptadores ópticos, comunicações ópticas, medidas de potência, atenuação

I. INTRODUÇÃO

A infraestrutura de fibra óptica cresceu significativamente nos últimos anos e segue com perspectiva de expansão, impulsionada principalmente pela demanda por banda larga fixa e vazão de dados aos sistemas celulares de quarta/quinta geração (4G/5G). Segundo a Agência Nacional de Telecomunicações (Anatel), a conectividade por fibra óptica alcançou mais de 70% dos municípios brasileiros em 2023, sendo direcionados esforços para elevar este percentual a 100% até 2027 [1]. Em paralelo, os sistemas de comunicações ópticas seguem em constante evolução, o que proporciona a ampliação e diversificação progressiva dos segmentos de aplicação. Nota-se, no cenário atual, soluções ópticas nas áreas de telecomunicações, sensoriamento, engenharia biomédica, indústria 4.0, aviação, dentre outras. No âmbito da pesquisa científica, inclusive, são numerosos os experimentos envolvendo óptica para o desenvolvimento de dispositivos optoeletrônicos e soluções fotônicas inovadoras [2-4].

Seja na implementação de redes ópticas ou em montagens experimentais, são vários os elementos envolvidos e faz-se necessário conectá-los rumo à aplicação desejada. Em geral, os dispositivos ópticos são empregados com suas terminações conectorizadas com conectores ópticos, que possuem especificações de tipo e polimento da face seguindo o formato “tipo/polimento”. Como exemplo, os conectores SC/PC (*subscriber connector/physical contact*) e SC/APC (*subscriber connector/angled physical contact*) apresentam forma quadrada e face com diâmetro de 2,5 mm, referente ao tipo SC. São, porém, diferenciados pelo polimento/corte reto (PC) ou angulado (APC) na extremidade cilíndrica. Existem ainda vários outros tipos de conectores com características físicas diferentes, como o FC (*Ferrule connector*) e o LC (*Lucent connector*), seguidos pela especificação de polimento.

Cabe mencionar o uso comum de cores para auxiliar na identificação, sendo o azul, o verde e o bege atribuídos aos conectores de polimento reto, de polimento angulado, e compatíveis com fibra multimodo (MM), respectivamente [5].

As conexões ópticas entre os dispositivos conectorizados demandam o uso de adaptadores ópticos, definidos como elementos passivos aptos a acoplar conectores similares ou de tipo diferente (caso da conexão híbrida) garantindo alinhamento. Tais adaptadores recebem classificações como SC, FC, LC, SC/FC e SC/LC dentre outras, referente ao tipo de conector suportado. São também comercializados com especificações adicionais quanto ao polimento e tipo de fibra. A Fig. 1 mostra adaptadores SC com as diferentes especificações comerciais: SC/PC, SC/APC e SC/MM. Nota-se a distinção por cores conforme padrão atribuído aos conectores, apesar da igualdade nas dimensões. É importante destacar que uma conexão óptica (conjunto adaptador + conectores) é considerada adequada ao resultar em atenuação típica inferior a 0,5 dB por conexão, devendo ser mantida a característica de polimento dos conectores na mesma [6-8].

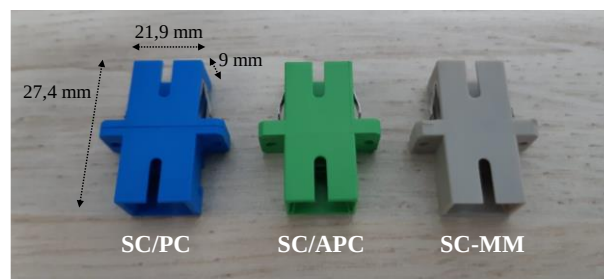


Fig. 1. Adaptadores do tipo SC com diferentes especificações.

Este trabalho reporta uma análise experimental de conexões ópticas baseadas em adaptadores SC, amplamente difundidos no mercado, considerando as diferentes versões expostas na Fig. 1. O objetivo é investigar a atenuação em uma mesma conexão óptica ao empregar cada um destes adaptadores, ou seja, analisar se as especificações adicionais do adaptador SC (polimento e tipo de fibra) afetam a perda de potência óptica na conexão. De acordo com o conhecimento dos autores, trata-se de um estudo/reporte inédito na literatura que contribui para a realização de conexões adequadas em redes e ensaios ópticos. O artigo é estruturado em três seções. A Seção II descreve os detalhes da montagem experimental em ambiente de laboratório e expõe os resultados obtidos, enquanto a Seção III apresenta as conclusões relevantes e as sugestões para trabalhos futuros.

II. MONTAGEM E RESULTADOS EXPERIMENTAIS

As Figs. 2(a) e 2(b) apresentam o diagrama em blocos e uma fotografia da montagem experimental, respectivamente.

Este trabalho foi realizado com o apoio do CNPQ, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

O laser ML925B45F controlado pelos dispositivos controladores de corrente (LDC205C) e de temperatura (TED200C e TCLDM9) emite luz contínua no comprimento de onda (λ) igual a 1550 nm. Com o auxílio de um microposicionador, a luz no espaço livre é confinada em um cordão óptico monomodo (SM) com 1 m de comprimento (cordão 1, azul) e conectores FC/PC e SC/PC nas extremidades. Tal cordão é acoplado no adaptador SC em teste, que faz a conexão com outro cordão óptico SM de 1 m (cordão 2, amarelo) terminado em conectores SC/PC. O sinal óptico chega então ao medidor de potência óptica (conjunto S122C + PM400), onde é realizada a leitura da potência recebida. Fez-se necessário o trecho de óptica no espaço livre (≈ 1 cm) pelo fato de o laser disponível para uso não ser conectorizado.

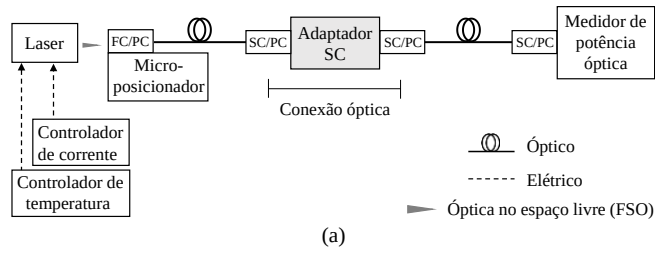
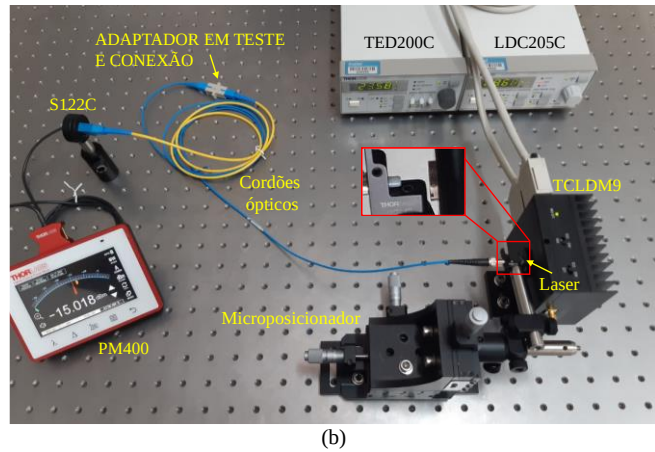


Fig. 2. Montagem experimental: (a) diagrama em blocos; (b) fotografia.

A metodologia de análise consiste em fixar as características da montagem descrita e alternar o adaptador óptico em teste na conexão segundo as versões SC/PC, SC/APC e SC-MM, medindo para cada caso a potência óptica recebida. É também estabelecida uma potência óptica de referência na recepção (-14,8 dBm) em um esquema *back-to-back* (B2B), onde o primeiro cordão óptico chega diretamente ao S122C, ou seja, cenário em que a conexão óptica envolvendo o adaptador em teste é eliminada. A comparação entre a potência recebida no cenário B2B e a potência recebida na montagem com adaptador SC permite estimar a atenuação imposta pela conexão óptica alvo do estudo, em função do adaptador SC utilizado. A perda de potência óptica no cordão 2 pode ser desprezada, já que fibras SM padrão apresentam atenuação típica de 0,2 dB/km em 1550 nm [2].

A Fig. 3(a) apresenta os resultados de potência óptica medida no PM400 ao longo do tempo em função do adaptador SC em uso, bem como a curva de potência medida no cenário B2B. A pequena oscilação nos valores de potência é oriunda de instabilidade do laser. Nota-se que a troca do adaptador não altera de forma significativa a potência medida (diferença $\leq 0,1$ dB) e que a atenuação na conexão permanece

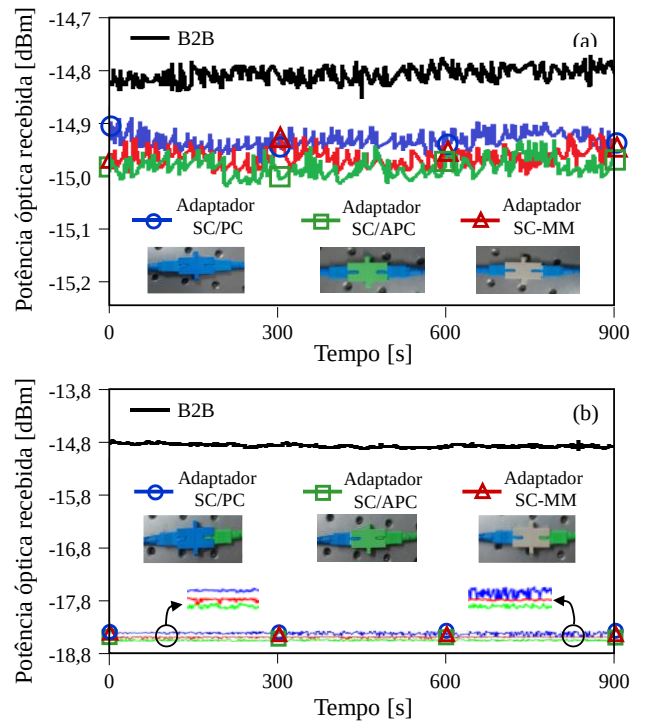


Fig. 3. Potência óptica no tempo em função do adaptador SC em uso: (a) conexão com mesmo polimento; (b) conexão com polimentos distintos.

abaixo de 0,25 dB, valor condizente com o esperado para uma conexão adequada (conectores de mesmo polimento na face). Em caso de conectores com polimentos distintos, observa-se (Fig. 3(b)) que a potência óptica é reduzida em mais de 3dB em relação ao cenário B2B. A troca dos adaptadores neste caso foi também irrelevante ao valor da atenuação na conexão.

III. CONCLUSÕES

O trabalho relatou uma análise experimental de conexões ópticas com adaptadores SC. Os resultados demonstraram que o uso de tais adaptadores com diferentes especificações de polimento (PC; APC) e tipo de fibra (SM; MM) não impacta significativamente a perda de potência em uma conexão óptica (diferença $\leq 0,1$ dB) entre conectores de polimentos semelhantes ou distintos. O comportamento exposto flexibiliza a escolha do adaptador óptico para uso na implementação de redes ópticas e em montagens experimentais. Trabalhos futuros em curso referem-se à análise similar com adaptadores de outros tipos e à proposta de adaptador óptico para acoplar conectores de polimento distintos sem prejuízo relevante na atenuação da conexão.

REFERÊNCIAS

- [1] Anatel, "Relatório anual de gestão 2023," Apr. 2024. [Online].
- [2] P. J. Winzer, D. T. Neilson, A. R. Chraplyvy, "Fiber-optic transmission and networking: the previous 20 and the next 20 years," *Opt. Express*, vol. 26, no.18, pp. 24190-24239, Sep. 2018.
- [3] H. R. D. Filgueiras et al., "Wireless and optical convergent access technologies toward 6G," *IEEE Access*, vol. 11, Jan. 2023.
- [4] E. S. Lima, R. M. Borges, N. Andriolli, E. Conforti, G. Contestabile, A. Cerqueira, "Integrated optical frequency comb for 5G NR Xhuals," *Scientific Reports*, vol. 12, 16421, Sep. 2022.
- [5] Furukawa Electric, "Performance de conectores ópticos," Especificação técnica ET03770 v6, 2023. [Online].
- [6] Fibracem, "Adaptador óptico," Ficha técnica R00. [Online].
- [7] Furukawa Electric, "Kit de adaptadores ópticos," Especificação técnica ET02221 v7, 2022. [Online].
- [8] Intelbrás, "Adaptador óptico," Ficha técnica - XFA1 e XFA2. [Online].

Automatização e controle de microposicionador para alinhamento em sistemas FSO

Elaine Aparecida Inácio
Laboratório de Telecomunicações
Universidade Federal de Itajubá
Itajubá, Brasil
elaineinacio@unifei.edu.br

José Alberto Ferreira Filho
Instituto de Engenharia de Sistemas e
Tecnologias da Informação
Universidade Federal de Itajubá
Itajubá, Brasil
jose.alb@unifei.edu.br

Ramon Maia Borges
Instituto de Engenharia de Sistemas e
Tecnologias da Informação
Universidade Federal de Itajubá
Itajubá, Brasil
ramonmb@unifei.edu.br

Resumo—Este artigo propõe e apresenta a validação experimental de uma solução automatizada para alinhamento de sistemas de óptica no espaço livre (FSO). Mais especificamente, trata-se da motorização e controle sem fio, via Bluetooth, de microposicionador óptico rumo à execução ágil e eficaz do alinhamento de feixes de luz não-visíveis. Resultados práticos obtidos em ambiente de laboratório demonstram o controle à distância do microposicionador com precisão de 10 μm , bem como o impacto positivo da solução proposta no acoplamento da luz em fibra óptica monomodo através de FSO.

Palavras-chave—alinhamento, FSO, microposicionador, Bluetooth

I. INTRODUÇÃO

Sistemas de óptica no espaço livre (FSO - *Free Space Optics*), apontados como uma das soluções técnicas promissoras para as redes móveis de quinta e sexta geração (5G e 6G), permitem a transmissão de informações em feixes de luz não confinados em meio físico. Viabilizam, por exemplo, o enlace entre lasers e fotodetectores sem uso de infraestrutura de cabo óptico, assim como o acoplamento da luz oriunda de *chips* ópticos, tipicamente não conectorizados, em segmentos de fibra óptica. Os cenários de aplicação da tecnologia FSO incluem: enlaces em áreas com obstáculos à instalação de fibra e/ou de difícil acesso; enlaces em áreas urbanas, especialmente entre edifícios ou torres elevadas; enlaces híbridos, onde faz-se a combinação com rádio sobre fibra (RoF); comunicação entre *chips* ópticos; comunicação entre *chip* óptico e fibra óptica; dentre outros [1-3].

Entre os desafios da tecnologia FSO, pode-se destacar a influência de condições climáticas e a dificuldade de alinhamento dos dispositivos ópticos nas extremidades do sistema de comunicação [4]. O processo de alinhamento, alvo do presente estudo, conta frequentemente com microposicionadores ópticos. Trata-se de estruturas mecânicas com plataforma móvel de ajuste micrométrico nos eixos “XYZ”, sobre a qual posiciona-se o dispositivo ou fibra óptica para alinhamento [5]. O ajuste da posição da plataforma, ou seja, o ajuste do microposicionador, é realizado até que se alcance a condição de alinhamento onde o acoplamento preciso da luz é estabelecido. Os microposicionadores são, assim, importantes nas fases de instalação, ativação e manutenção dos sistemas FSO.

Existem diversas opções de microposicionadores, variando em tamanho, capacidade de carga, precisão e método de controle. Este último inclui: uso de ajuste manual através de bastões rotacionáveis nos três eixos [5]; uso de materiais piezoelétricos que se deformam quando uma tensão elétrica é aplicada, permitindo movimentos precisos da plataforma em resposta a pequenos sinais elétricos [4], uso de campos

eletromagnéticos gerando forças que movem componentes mecânicos de forma precisa [6]; uso de hidráulica e pneumática para movimentação da plataforma através da aplicação de pressão e fluidos [7]; dentre outros métodos. Existem ainda no mercado microposicionadores motorizados, tipicamente controlados por cabos conectados ao controlador via Ethernet, USB ou RS-232 [5].

Este trabalho apresenta uma solução para a automatização e controle sem fio de microposicionador óptico destinado a aplicações de FSO. O objetivo é criar alternativa eficiente e de baixo custo para ajustar microposicionadores de forma remota, sem a necessidade de cabos. A proposta inova ao contar com a tecnologia Bluetooth, além de microcontrolador e motor de passo. A prova de conceito acontece com a automatização e controle sem fio de um microposicionador comercial, originalmente de controle manual.

II. MICROPOSICIONADOR AUTOMATIZADO E RESULTADOS EXPERIMENTAIS

O projeto de automatização e controle do microposicionador envolve o uso de um motor de passo NEMA 17, conectado a um *driver* L298N e controlado por uma ESP32 via Bluetooth, como ilustra a Fig. 1. O eixo giratório do motor é acoplado fisicamente ao bastão rotacionável do microposicionador através de um arranjo com correia dentada e polias. O esquema permite comandos remotos para girar o motor em ambas as direções e limita os movimentos ao comprimento do eixo do microposicionador, para que se obtenha precisão. O motor de passo NEMA 17, com 200 passos por rotação (1.8° por passo), foi selecionado devido às especificações do microposicionador alvo da automatização. Tal microposicionador possui uma faixa de medição de 0 a 25 mm, com uma resolução de 0.01 mm por divisão, e é capaz de realizar 50 voltas completas no eixo giratório. Para alcançar um deslocamento linear de 0.01 mm, o motor precisa rotacionar aproximadamente 7.19°, o que equivale a cerca de 4 passos. Uma vez estabelecida a automatização e controle, o microposicionador é utilizado para alinhamento em FSO.

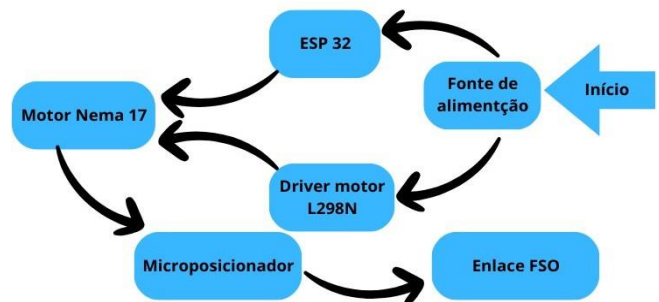


Fig. 1. Diagrama em blocos da proposta.

A solução alvo de estudo foi testada no acoplamento de luz não-visível entre um laser não-conectorizado e um cordão óptico, com um dos seus conectores fixado na plataforma móvel do microposicionador automatizado e controlado via Bluetooth. A outra extremidade do cordão óptico foi conectada a um medidor de potência óptica (conjunto S122C + PM400) para o monitoramento da potência recebida. O comprimento de onda (λ) da luz no ensaio foi de 1550 nm. A Fig. 2 apresenta uma fotografia da montagem experimental, enquanto a Tab. 1 lista os dispositivos utilizados.

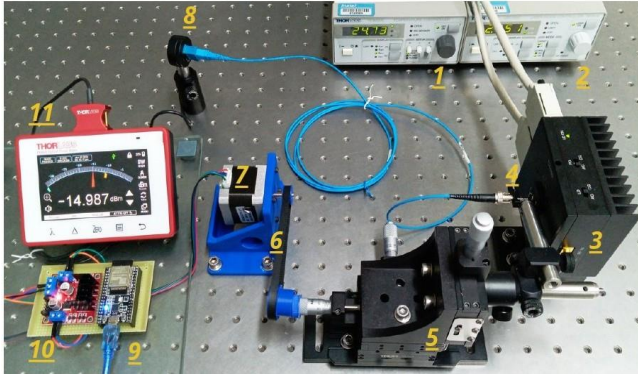


Fig. 2: Fotografia da montagem experimental.

Tabela 1: Dispositivos utilizados

Nº	Equipamento	Especificação
1	TED200C	Controlador de Temperatura
2	LDC205C	Controlador de Corrente
3	TCLDM9	Estrutura para o Laser
4	Laser	FSO de 1 cm
5	Microposicionador	Microposicionador manual
6	Correia	Correia dentada
7	Motor de Passo	Motor de passo NEMA 17
8	S122C	Sensor de Potência Óptica
9	Microcontrolador	Microcontrolador ESP32
10	Driver	Driver L298N
11	PM400	Medidor de Potência Óptica

A Fig. 3 demonstra os comandos enviados via Bluetooth ao ESP32 para mover o eixo e, consequentemente, a plataforma do microposicionador. Esses comandos permitem a rotação em ambas as direções, ajustando a quantidade de passos desejados dentro dos limites pré-estabelecidos do eixo do microposicionador. Cabe mencionar que os modelos de microposicionadores estudados não oferecem suporte para conexão via Bluetooth, uma tecnologia reconhecida por sua universalidade. A conexão via Bluetooth permite comunicação eficaz em distâncias de até aproximadamente 50 metros em linha reta, como comprovado por testes positivos de controle a distâncias desta ordem de grandeza. Integrar essa conectividade aos dispositivos não apenas simplifica a operação, mas permite o controle e gerenciamento remoto.

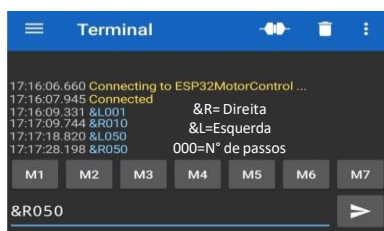


Fig. 3: Exibição dos comandos via Bluetooth.

Os testes realizados mostraram que a inclusão do motor de passo não apenas contribuiu para o estabelecimento do trecho em FSO na montagem experimental, mas também melhorou a precisão do alinhamento no eixo X do microposicionador, quando comparado ao alinhamento manual. É importante mencionar que, nesta prova de conceito inicial, os testes foram conduzidos apenas para um eixo do microposicionador, com os demais eixos sendo ajustados manualmente. A Fig. 4 reporta uma análise da potência óptica medida no PM400 em função do tempo, considerando o ajuste manual e automatizado do microposicionador. Observa-se que os ajustes manuais iniciais causam uma variação abrupta na potência óptica acoplada no cordão óptico, até que se chegue ao valor mais alto de potência medido ($\approx -12,6$ dBm). Na sequência, o ajuste automatizado com diferentes alternativas de passos permite alcançar o mesmo valor de potência, inclusive viabilizando o ajuste fino preciso, estável e à distância no alinhamento. Os trechos de decaimento sutil da potência óptica durante o controle automatizando indicam os momentos em que foram dados passos maiores propositalmente a fim de demonstrar o impacto no controle do alinhamento.

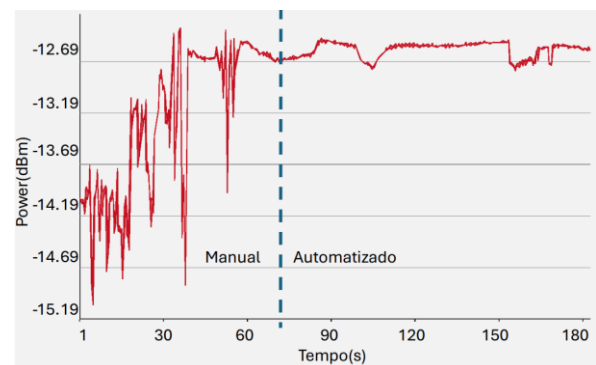


Fig. 4: Potência óptica acoplada no cordão óptico ao longo do tempo em função do controle manual e automatizado do microposicionador.

III. CONCLUSÕES

O trabalho apresentou uma solução para automatização e controle via Bluetooth de um microposicionador óptico mecânico. Os resultados da prova de conceito inicial e experimental em cenário FSO validaram o potencial do controle sem fio à distâncias de até 50m, com precisão de 10 μ m. Trabalhos futuros visam miniaturizar o setup proposto.

IV. REFERENCES

- [1] H. R. D. Filgueiras et al., "Wireless and optical convergent access technologies toward 6G," *IEEE Access*, vol. 11, Jan. 2023.
- [2] H. Kaushal, G. Kaddoum, "Optical communication in space: challenges and mitigation techniques," *IEEE Commun. Surveys & Tutorials*, vol. 19, 2017.
- [3] E. Erdogan, I. Altunbas, G. Karabulut Kurt, H. Yanikomeroglu, "The Secrecy Comparison of RF and FSO Eavesdropping Attacks in Mixed RF-FSO Relay Networks," *IEEE Photonics J.*, vol. 14, no. 1, 2022.
- [4] H. J. Shieh, P. K. Huang, "Adaptive Tracking Control of a Piezoelectric Micropositioner," In: *IEEE International Conference on Industrial Electronics and Applications*, Singapore, 2006.
- [5] Thorlabs, "Plataformas de flexura e posicionadores de 3 eixos," Plataformas multi-eixos. [Online].
- [6] M.-Y. Chen, H.-H. Huang, S.-K. Hung, L.-C. Fu, "Design and Implementation of a New 3-DOF Electromagnetic Micropositioner Utilizing Flexure Mechanism," In: *American Control Conference*, Minneapolis, USA, 2006.
- [7] D.A. Horsley, A. Singh, A.P. Pisano, R. Horowitz, "Angular micropositioner for disk drives," in *IEEE Int. Workshop on Micro Electro Mechanical Systems*, 1997.

Controle Fuzzy para Otimização da Pressão em Sistemas de Distribuição de Água

Taciano Douglas P. Prudêncio

Departamento de engenharia elétrica

Universidade Federal da Paraíba

João Pessoa, Brasil

taciano.prudencio@estudante.cear.ufpb.br

José Hélio B. da Silva

Departamento de engenharia elétrica

Universidade Federal da Paraíba

João Pessoa, Brasil

jose.dasilva@cear.ufpb.br

Juan Moises M. Villanueva

Departamento de engenharia elétrica

Universidade Federal da Paraíba

João Pessoa, Brasil

jmauricio@cear.ufpb.br

Resumo—Este trabalho apresenta o desenvolvimento de um controlador de pressão fundamentado na lógica Fuzzy para sistemas altamente dinâmicos e não lineares de abastecimento de água. O objetivo é controlar a pressão através da variação da frequência de operação do conjunto motor-bomba resultando na diminuição de perdas hídricas e aumentando eficiência energética.

Palavras-chave—lógica fuzzy, controle de pressão, abastecimento de água, eficiência energética.

I. INTRODUÇÃO

De acordo com levantamento divulgado pelo Instituto Trata Brasil, com ano base em 2022, mais de 37% da água captada pelos reservatórios brasileiros é desperdiçada ao longo de todo o processo até chegar ao consumidor. A eficiência na rede de distribuição está associada as fugas de água, deteriorização da rede, consumo energético do conjunto motor-bomba e ao controle da pressão e vazão [1]. Podendo ser esses dois últimos parâmetros eficientemente controlados a partir do controle da velocidade do conjunto motor bomba.

Além disso, 90% do consumo de energia elétrica em sistemas de bombeamento de água é resultante da operação do conjunto motor bomba [2]. Paralelo ao desafio do consumo energético, há as perdas por fugas de água que existem por consequência de pressões elevadas nas tubulações [3]. Portanto, é necessário investimento das concessionárias na melhoria da eficiência energética e no controle de pressão por meio de técnicas de controle inteligente que possam adaptar-se a cenários dinâmicos e não lineares.

Na literatura, alguns métodos para controle da pressão estão se destacando, como: aplicação da lógica difusa (*Fuzzy*) [4], controle adaptativo [5] e controle Proporcional Integrativo e Derivativo (PID) [6]. Nesse contexto, o objetivo deste trabalho é criar um controlador de pressão fundamentado na lógica Fuzzy para um sistema de abastecimento de água composto por um conjunto motor bomba.

II. REVISÃO BIBLIOGRÁFICA

O estudo realizado em [4] utilizou um modelo baseado em dados históricos e um controlador fuzzy para atuar sobre os inversores de frequência. Resultados mostraram melhoria na estabilidade da pressão e redução no consumo de energia,

mas houve instabilidades devido a falhas na comunicação, indicando a necessidade de otimização do sistema.

Enquanto o trabalho apresentado em [6] descreve a aplicação de um algoritmo de controle PID, ajustando os parâmetros do controlador para um sistema de abastecimento simulado. A aplicação resultou em uma redução significativa no consumo energético, contudo, a não linearidade do sistema e possíveis perturbações podem torná-la ineficaz.

Considerando o exposto, foi possível identificar tanto as oportunidades quanto os desafios no desenvolvimento de sistemas de controle em redes de distribuição de água, visto que envolve a aplicação de técnicas tradicionais de controle, como o PID, e abordagens mais avançadas, como a lógica Fuzzy.

III. PROJETO DO CONTROLADOR

Este estudo foi conduzido em bancada experimental representada na Figura 1, localizada no Laboratório de Eficiência Energética e Hidráulica em Saneamento (LENHS). A estrutura da bancada inclui um conjunto motor-bomba (CMB), transdutores de pressão, eletroválvulas, válvulas mecânicas, transdutores de vazão e a tubulação.

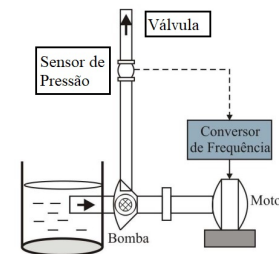


Fig. 1. Modelo esquemático do sistema estudado.

O conversor de frequência é um equipamento eletrônico empregado para modular a velocidade de operação do motor elétrico. O intervalo de atuação do inversor deste estudo é de 3 Hz a 60 Hz. O sensor de pressão mede a variação de pressão ou altura manométrica da tubulação. A eletroválvula é utilizada neste estudo para emular o comportamento dinâmico da pressão em tubulações reais, onde a demanda por água varia ao longo do dia.

Um controlador Fuzzy típico pode ser projetado para simular a capacidade humana de raciocinar, ou seja, o processo pelo

qual as pessoas inferem conclusões com base nas informações que já possuem [7]. Os três principais passos para criação de controladores Fuzzy são a fuzzificação, a inferência e a defuzzificação.

Fuzzificação é o passo no qual valores de entrada, como temperatura ou velocidade, são convertidos em graus de pertinência nos conjuntos Fuzzy predefinidos. Oferecendo uma representação mais flexível e adaptável do mundo real. A inferência é o método onde opera com regras linguísticas do tipo "se *antecedente* então *consequente*", onde as variáveis de entrada são combinadas para produzir um conjunto de saídas Fuzzy.

Na etapa de defuzzificação, os resultados Fuzzy gerados pela inferência são transformados em valores de saída compreensíveis para o atuador, em consonância com as observações em [8].

O controlador proposto é baseado em duas entradas: o erro e a derivada do erro (Derro). O erro, neste contexto, refere-se à discrepância entre o valor desejado (ou referência) e o valor atual obtido pelo sistema. Por outro lado, a Derro descreve a taxa de mudança do erro ao longo do tempo, indicando se o sistema está se aproximando ou se afastando da referência e com qual velocidade isso ocorre.

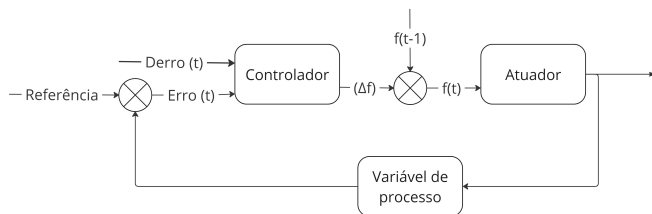


Fig. 2. Diagrama de blocos do controlador Fuzzy.

A saída do controlador é o incremento na frequência (Δf), que representa o ajuste necessário para corrigir o sistema em direção à referência desejada. Assim, o controlador utiliza informações sobre a magnitude e a direção do erro, bem como sua variação ao longo do tempo, para determinar o ajuste adequado na frequência. Uma representação esquemática desse controlador pode ser observada na Figura 2.

IV. RESULTADOS

O impacto do controlador Fuzzy foi analisado efetuando manobras hidráulicas para variar o ângulo de fechamento da válvula (CV-1) ao longo do experimento para representar diferentes padrões de consumo. Com essas variações, foi possível avaliar como o controlador influencia os custos energéticos, comparando a energia consumida com e sem a aplicação do controle Fuzzy. Ou seja, o custo energético foi medido para o sistema em malha aberta e malha fechada. Os resultados estão dispostos nas Tabelas I e II

A análise desses dados revela que o sistema em malha fechada obteve uma redução significativa no consumo de energia: 14,20% para CV-1 em 30°, 17,07% para 50° e 15,28%

Tabela I
CONSUMO DO SISTEMA EM MALHA ABERTA.

Tempo (min)	CV-1 (°)	Corrente (A)	Potência (W)	Energia (kWh)
10	0	3,65	1652	0,2754
10	30	3,45	1562	0,2603
10	50	3,28	1485	0,2475
10	70	3,14	1421	0,2369

Tabela II
CONSUMO DO SISTEMA EM MALHA FECHADA.

Tempo (min)	CV-1 (°)	Corrente (A)	Potência (W)	Energia (kWh)
10	0	3,65	1652	0,2754
10	30	2,96	1340	0,2233
10	50	2,72	1231	0,2052
10	70	2,66	1204	0,2007

para 70°, em comparação com o motor operando a 45 Hz. Em média, isso representa uma diminuição de 15,51%.

V. CONCLUSÃO

Neste artigo foi apresentado o desenvolvimento de um controle de pressão baseado em lógica Fuzzy para uma bancada que simula um sistema de abastecimento de água. A lógica de controle utilizou a aquisição dos dados de pressão para inferir modificações atuando diretamente na frequência de operação do conjunto motor bomba.

Apesar dos sinais de pressão e de frequência não terem uma relação linear entre si, o sistema elaborado possibilitou o controle inteligente, visto que na análise de eficiência energética houve uma redução média de 15,51% em relação ao funcionamento em malha aberta, representando uma solução para melhorar o consumo energético no sistema de abastecimento de água e mitigar os efeitos negativos da variação de pressão.

REFERÊNCIAS

- [1] D. B. Cavalcanti, J. J. P. Moura, J. M. M. Villanueva, e H. P. Gomes. Modelagem de um Sistema de Bombeamento para o Controle de Vazão por Conversor de Frequência Baseado na Teoria de Identificação de Sistemas. In *Anais do Congresso Brasileiro de Automática (CBA 2018)*, pp. —, 2018. DOI: 10.20906/cps/cba2018-0313.
- [2] S. Bezerra *et al.* Energy savings in pumping systems: Application of a fuzzy system. *Ciência & Engenharia*, v. 24, n. 1, pp. 71–78, 2015.
- [3] R. A. Sobrinho e P. C. Borja. Management of water and energy losses in embasa water supply systems: A study of factors involved in rms. *Engenharia Sanitária e Ambiental, SciELO Brasil*, v. 21, pp. 783–795, 2016.
- [4] J. V. S. de Araújo *et al.* Fuzzy control of pressure in a water supply network based on neural network system modeling and IoT measurements. *Sensors*, v. 22, n. 23, 2022. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/22/23/9130>.
- [5] T. K. S. Flores *et al.* Adaptive pressure control system based on the maximum correntropy criterion. *Sensors, MDPI*, v. 21, n. 15, p. 5156, 2021.
- [6] A. C. L. Monteiro *et al.* Aplicação de um controlador PID para redução do consumo de energia elétrica em sistemas de abastecimento de água. 2016.
- [7] M. G. Simões e I. S. Shaw. Controle e modelagem fuzzy. [S.l.]: Editora Blucher, 2007.
- [8] M. P. Schoen. Introduction to Intelligent Systems, Control, and Machine Learning using MATLAB. [S.l.]: Cambridge University Press, 2023.

Communication in Robotic Systems with ROS2 and Arduino using AADL

Igor Gabriel dos Santos

DIATINF

IFRN

Email: igor1208gabriel@gmail.com

Ivanilson França Vieira Junior

DIATINF

IFRN

Email: ivanilson.junior@ifrn.edu.br

Jorgiano Marcio Bruno Vidal

DIATINF

IFRN

Email: jorgiano.vidal@ifrn.edu.br

Abstract—This study uses ROS2 and Arduino, employing a low-power wireless module to model wireless communication in robotic systems. The research examines the interaction between a desktop computer and an Arduino microcontroller. Using the Architecture Analysis and Design Language (AADL), we model the system architecture and analyze its communication performance. This paper establishes a foundation for future advancements in integrating microcontrollers with advanced robotic frameworks.

Index Terms—ROS2, Arduino, Modeling, AADL, Robotic Systems, communication

I. INTRODUCTION

Effective communication is the backbone for the functioning of embedded systems, particularly in robotic applications where reliable and real-time data exchange is crucial. Communication in robotic systems ensures the seamless interaction between various components, such as sensors, actuators and processing units. In modern robotics, the integration of diverse hardware and software components requires robust communication protocols to maintain synchronization and coordination among different system parts.

This study presents an interaction approach for wireless communication between a desktop computer, responsible for running the Robot Operating System (ROS2) [1] framework, and an Arduino microcontroller equipped with an HC-06 Bluetooth module. Despite the Arduino's limited compatibility with ROS2, this study demonstrates an approach to bridging this gap, thereby enabling seamless integration for microcontrollers within advanced robotic frameworks.

A key challenge addressed in this study is the direct communication between ROS2 and an Arduino microcontroller without intermediary microcomputers like Raspberry Pi, which may not be feasible due to power constraints. This research fills a significant gap by demonstrating a method to achieve this communication via Bluetooth, which has not been extensively covered in existing literature.

Architecture Analysis and Design Language (AADL) [2] is used to meticulously model the system architecture, profiling and deploying hardware and software components and analyzing communication performance and reliability. This dual focus on modeling and practical implementation provides a comprehensive framework for future enhancements in robotic systems integration.

ROS2 capabilities are essential for developing complex robotic systems, offering standardized interfaces for message

passing through topics and a suite of development, debugging, monitoring and configuration tools. This allows deploying services onto different platforms and reduce time-to-market for applications [3]. Despite the simplicity of the robotic application presented, using ROS2 infrastructure ensures scalability and ease of understanding, facilitating for more sophisticated implementations.

The system infrastructure includes an Arduino board, Bluetooth module and a desktop with ROS2. This configuration can be adapted with minimal modifications, showcasing the flexibility and applicability of the proposed approach.

This paper is structured as follows: Section II details the robotic system infrastructure, including hardware and software. Section III delves into the modeling process with AADL. Section IV discusses the results and future work, summarizing key findings and their implications. All source code is available on <https://github.com/Igor1208gabriel/Arduino-ROS-Bluetooth.git>

By addressing these aspects, this study contributes to the foundational understanding of integrating microcontrollers with ROS2, providing a blueprint for future advancements in robotic communication systems

II. ROBOTIC SYSTEM INFRASTRUCTURE

To demonstrate this interaction approach, a very simple system is used: An Arduino Mega board with an HC-06 Bluetooth module connected to a desktop computer using Ubuntu 22.04 distribution with ROS2 Iron. The Arduino collects data from an ultrasonic sensor and transmits the readings via Bluetooth to the desktop computer, which processes the data using ROS2 nodes.

The hardware components consists of: An Arduino Mega board, an HC-06 Bluetooth module, an HC-SR04 ultrasonic sensor and the desktop computer

The software components consists of: An Arduino sketch, A ROS2 BTSender node and a ROS2 BTRceiver node.

III. MODELING AND AADL

AADL is utilized to model the system architecture, providing a detailed view of the whole system[2]. That includes hardware (Arduino, Bluetooth module) and software (ROS2 nodes) components and a view of how the software is deployed onto the hardware (Communication paths). By addressing

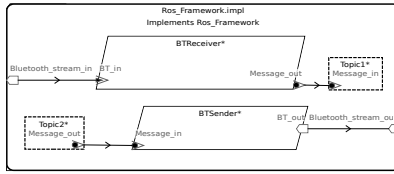


Fig. 1. AADL software model

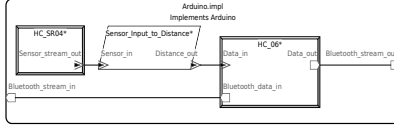


Fig. 2. AADL hardware model

these aspects, this study contributes to the foundational understanding of integrating microcontrollers with ROS2, providing a blueprint for future advancements in robotic communication systems. Secondly, another point of using AADL is to verify and analyze the system's architecture, ensuring it meets the desired performance criteria.

OSATE2 [4], the tool used to write AADL models allows for writing models for the three aforementioned views and for checking the models rightness and performance [5]. The primary goal of performance analysis is to identify potential bottlenecks and areas for optimization. By simulating different scenarios, it is possible to evaluate the system's communication latency, reliability, and overall performance [6].

Although this study does not optimally predict and analyze communication latency and performance, AADL models can incorporate fault injection and error propagation analysis to evaluate the system's reliability and fault tolerance, allowing for more robust and scalable systems. Additionally, modeling robotic systems also allow for optimization opportunities and validating the systems.

To effectively model the system architecture, AADL is employed, which offers a comprehensive view of both software and hardware components, and their interactions. The model visually represents the interaction and communication flow within the system, highlighting the integration of various components.

Figure 1 shows the software components: BTReceiver, receives the stream of information given by the HC-06 module and publishes it to Topic1; BTSender, receives information from Topic2' and sends it to the Arduino Mega board via the HC-06 module; Sensor_input_to_distance (Arduino sketch) reads data from the HC_SR04 sensor, calculates the distance and sends it to the desktop via the HC-06 module. Additionally, it handles incoming data from the desktop. Figure 2 shows the hardware components: an Arduino Mega board, a desktop computer, an HC-06 Bluetooth module and an HC-SR04 ultrasonic sensor which collect distance measurements to be processed by the Arduino

Interaction flow: The HC-SR04 sensor collects distance measurements and sends the raw sensor data to the Arduino Mega. The Arduino Mega processes this data using the

Sensor_Input_to_Distance sketch and transmits the distance information to the desktop via the HC-06 Bluetooth module. On the desktop, the BTReceiver node receives the Bluetooth data and publishes it to Topic1 in the ROS2 framework. BTSender subscribes to Topic2 and sends commands back to the Arduino via the HC-06 Bluetooth module, enabling bidirectional communication.

IV. RESULTS AND FUTURE WORK

The limited compatibility of Arduino with ROS2 presents a significant challenge for embedded and robotic systems developers. ROS2, being a powerful middleware for robotic systems, requires computational resources which Arduino lacks. This limitation hinders the direct deployment of ROS2 nodes onto the microcontrollers, complicating the integration process.

To address this issue, the study leverages Bluetooth to bridge the gap between Arduino and a more capable ROS2 environment running on a desktop computer. This approach allows the Arduino to collect sensor data and execute actuator commands while offloading the computationally intensive ROS2 processing to the desktop.

By demonstrating a practical solution to this compatibility issue, this study provides a valuable framework for other developers, facilitating more seamless integration of Arduino microcontrollers into advanced robotic systems.

This study demonstrates a novel approach to integrating Arduino microcontrollers with ROS2 via Bluetooth, using AADL for system modeling. Although detailed performance analysis was not conducted, the successful implementation demonstrates the potential for this method in robotic systems, addressing a significant gap in existing literature. Future work will focus on comprehensive performance testing and model analysis to further validate and optimize system models. Additionally, efforts will be directed towards automatic code generation from AADL models to streamline the development process.

REFERENCES

- [1] O. Robotics. Ros, powering the world's robots. [Online]. Available: <https://www.ros.org/>
- [2] P. H. Feiler, B. A. Lewis, and S. Vestal, "The SAE Architecture Analysis & Design Language (AADL) a standard for engineering performance critical systems," in *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*, 2006, pp. 1206–1211.
- [3] E. Senn, L. W. J. Bourdon, and D. Blouin, "Multi-paradigm modeling for early analysis of ros-based robotic applications using a library of aadl models," in *Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings*, ser. MODELS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 677–683.
- [4] Welcome to osate. [Online]. Available: <https://osate.org/>
- [5] E. SENN and L. BOURDON, "Introducing CPU load Analysis from AADL Models for ROS applications : a use case," in *2021 Forum on specification & Design Languages (FDL)*, 2021, pp. 01–08.
- [6] E. Senn, "ROS communications profiling for bus load analysis from AADL," in *11th European Congress on Embedded Real Time Systems 2022*, ser. ERTS 2022 proceedings. Toulouse, France: 3AF SEE, Jun. 2022. [Online]. Available: <https://hal.science/hal-03818603>

Algoritmo de Filtragem de Sinais de Medição de Sensores em Cenários com Ruído Impulsivo

Pedro Thiago V. de Souza
Universidade Federal Rural do Semi-Árido
Centro Multidisciplinar de Pau dos Ferros
Pau dos Ferros, RN, 59900-970
Email: pedro.souza@ufersa.edu.br

Vinicius Samuel V. de Souza
Universidade Federal Rural do Semi-Árido
Centro Multidisciplinar de Pau dos Ferros
Pau dos Ferros, RN, 59900-970
Email: viniussamuel@ufersa.edu.br

Resumo—Neste trabalho propomos um novo algoritmo para a filtragem de sinais obtidos através da medição de sensores em cenários em que o ruído impulsivo está presente. O algoritmo é baseado na média móvel de Kaufman (KAMA - *Kaufman Adaptive Moving Average*). A partir dos resultados obtidos, constatou-se que o algoritmo é robusto à *outliers*, principalmente quando comparado ao média móvel.

Index Terms—Ruído impulsivo, filtro adaptativo, instrumentação, controle, α -estável

I. INTRODUÇÃO

Em sistemas de controle e automação, os sinais obtidos por sensores são degradados devido ao ruído. De forma geral, o ruído nesses sistemas é modelado como Gaussiano [1]. Todavia, em alguns casos, há existência de *outliers*, os quais não podem ser modelados por uma distribuição Gaussiana [2]. Nesses cenários, é mais prudente modelar o ruído através de um modelo impulsivo [3].

Embora a modelagem de ruído por um modelo impulsivo seja bastante aceita pela comunidade científica, boa parte dos algoritmos de processamento de sinal foram propostos para cenários em que o ruído é Gaussiano [2]. Um dos algoritmos mais utilizados para minimizar o efeito do ruído em medições de sensores é a média móvel [4]. Todavia, esse algoritmo é projetado assumindo que o ruído é Gaussiano, e o seu desempenho em cenários impulsivos é sub-ótimo [1].

Nesse trabalho, propõem-se um novo método para filtragem de sinais obtidos através de sensores em que o ruído impulsivo está presente. O algoritmo consiste em um filtro de primeira ordem adaptativo, em que o filtro torna-se mais seletivo quando ocorre um *outlier* no sinal de entrada.

II. SISTEMA PROPOSTO

O sistema proposto corresponde à um filtro em tempo discreto passa-baixas de primeira ordem, causal, com entrada $x[n]$ e saída $y[n]$, descrito pela seguinte equação de diferenças:

$$y[n] - ay[n-1] = (1-a)x[n] \quad (1)$$

em que $a \in \mathbb{R}$ e $|a| < 1$. O parâmetro a define a frequência de corte de 3dB desse filtro, podendo ser calculado como:

$$a = 2 - \cos \omega_c - \sqrt{\cos^2 \omega_c - 4 \cos \omega_c + 3} \quad (2)$$

no qual ω_c é a frequência de corte em 3dB do filtro.

O sistema aqui proposto consiste em ajustar, de forma dinâmica, o parâmetro a . A regra de ajuste é escolhida tal que quando o sinal de entrada apresentar uma grande variação de amplitude (ocasionada por um *outlier*, por exemplo), o filtro torne-se mais seletivo, eliminando-se essa perturbação.

A regra de ajuste adotada é baseada na KAMA. A KAMA é bastante utilizada no mercado financeiro, em que ocorrem grandes variações dos valores das ações [5]. No sistema proposto, inicialmente calcula-se a volatilidade da entrada ($v[n]$) e a mudança total ($c[n]$), definidos como:

$$v[n] = |x[n] - x[n-1]|, \quad (3)$$

$$c[n] = |x[n] - x[n-N+1]|, \quad (4)$$

após isso, o índice de eficiência ($e_r[n]$) é calculado como:

$$e_r[n] = \frac{c[n]}{\sum_{k=n-N+1}^n v[k]}. \quad (5)$$

Nas Equações (3)-(5), N indica a quantidade de amostras de entrada utilizadas para o cálculo adaptativo dos parâmetros do filtro. Por fim, o parâmetro a é calculado como:

$$a = e_r[n](a_{max} - a_{min}) + a_{min} \quad (6)$$

sendo a_{max} e a_{min} os valores máximo e mínimo de a , podendo ser determinados através da frequência de corte mínima e máxima do filtro, respectivamente, e calculadas utilizando-se a Equação (2).

III. SIMULAÇÃO COMPUTACIONAL

As simulações computacionais foram realizadas considerando-se um sinal de entrada como uma senoide de 200 Hz, amostrada através de um conversor discreto-contínuo ideal com frequência de amostragem de 20 kHz. Para simular a perturbação, gerou-se um ruído impulsivo através da distribuição α -estável com relação de sinal-ruído geométrica (GSNR - *Geometric Signal-to-Noise*) [6] de 5dB. O parâmetro α da distribuição foi variado de $\alpha = 1,1$ até $\alpha = 2,0$, esse último correspondente à um cenário de ruído Gaussiano. O ruído com distribuição α -estável gerado através do método de Weron & Weron [7].

Utilizou-se como frequência de corte mínima e frequência de corte máxima, respectivamente, os valores de 300Hz e 600Hz. Esses valores devem ser escolhidos de forma que todo o espectro de sinal de entrada esteja na banda passante do filtro. Foram utilizadas $N = 10$ amostras de entrada para o cálculo dos parâmetros adaptativos do filtro.

A Fig. 1 apresenta (a) o sinal senoidal original, (b) o sinal senoidal contaminado por ruído, (c) o sinal após a filtragem do sistema proposto e (d) o sinal após a filtragem por média móvel com 10 amostras. Os resultados foram obtidos considerando $\alpha = 1, 5$. Obviamente, por tratar-se de resultados estocásticos, esses podem variar com o ruído de entrada, mas observa-se que o sistema proposto é mais robusto do que um média móvel, principalmente quando ocorre um *outlier* de entrada.

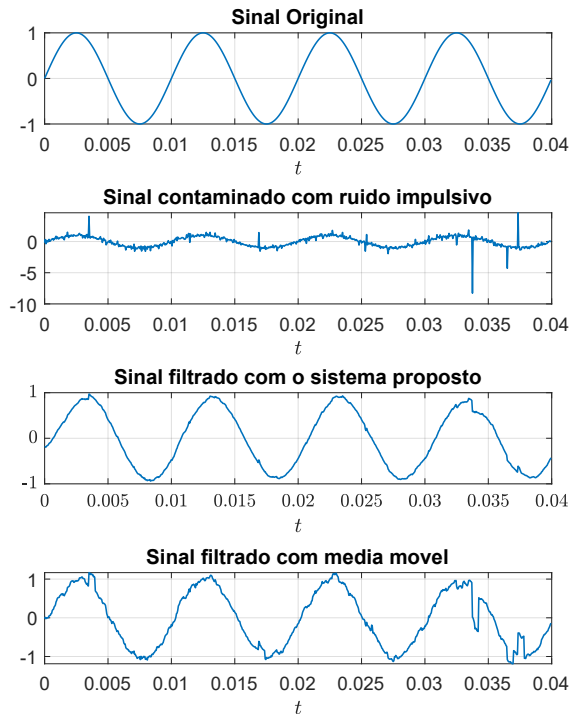


Figura 1. Comparação do sinal original, sinal contaminado por ruído impulsivo, sinal filtrado com o sistema proposto e o sinal filtrado com a média móvel considerando $\alpha = 1, 5$ e $GSNR = 5$ dB.

A Fig. 2 apresenta o coeficiente de correlação médio entre o sinal senoidal e o sinal de saída do sistema proposto e de um filtro de média móvel. Para cada valor de α , o coeficiente de correlação foi obtido através da média de 500 simulações. Nota-se que o desempenho dos algoritmos é degradado com a diminuição de α , haja vista que o canal torna-se mais impulsivo. Além disso, em cenários mais impulsivos, o desempenho do algoritmo proposto é mais eficiente do que o média móvel. Para valores de $\alpha < 1, 8$, o algoritmo proposto é mais eficiente do que o média móvel. Por sua vez, quanto mais próximo de $\alpha = 2$, mas Gaussiano será o canal e melhor será o desempenho da média móvel.

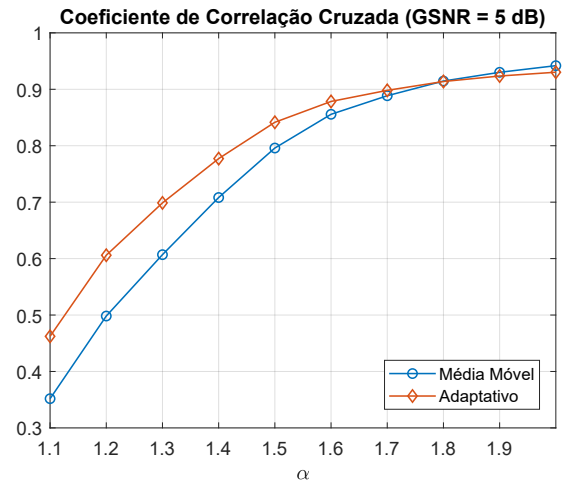


Figura 2. Coeficiente de correlação cruzada entre o sinal de saída e o sinal original considerando $GSNR = 5$ dB.

IV. CONCLUSÕES

Analisando-se os resultados obtidos, pode-se constatar que o sistema proposto é robusto à presença de *outliers*. Como trabalhos futuros pretende-se implementar o sistema proposto em bancada experimental utilizando-se FPGA (*Field Programmable Gate Array*). Vale citar que embora o cenário de análise seja para processamento de sinais de sensores, o algoritmo proposto pode ser utilizado em outros cenários, tal como em sistemas de comunicações.

REFERÊNCIAS

- [1] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, Jul. 1993, doi: 10.1109/5.231338.
- [2] P. T. V. de Souza, A. I. R. Fontes, V. S. V. De Souza and L. F. Silveira, "A Novel Signal Detector in MIMO Systems Based on Complex Correntropy," in *IEEE Access*, vol. 7, pp. 137517–137527, 2019, doi: 10.1109/ACCESS.2019.2942822.
- [3] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: new methods and results for class A and class B noise models," in *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1129–1149, May 1999, doi: 10.1109/18.761256.
- [4] Z. L. Chong, K. M. Chan, J. Wang, J. -C. Malela-Majika and S. C. Shongwe, "Overall Performance Comparison of Homogeneously Weighted Moving Average and Double Homogeneously Weighted Moving Average Schemes," *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore, Singapore, 2021, pp. 1225–1229, doi: 10.1109/IEEM50564.2021.9672787.
- [5] Craig A. Ellis and Simon A. Parbery, "Is smarter better? A comparison of adaptive, and simple moving average trading strategies," *Research in International Business and Finance*, Volume 19, Issue 3, 2005, pp. 399–411, doi: 10.1016/j.ribaf.2004.12.009.
- [6] S. Luan, J. Li, W. Chen, J. Zhang, T. Liu and B. Lv, "MMCC-Music: A Robust Direction of Arrival Estimator based On Maximum Mixture Correntropy Criterion Under Alpha-Stable Distributed Noise," *2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)*, Xi'an, China, 2022, pp. 01–06, doi: 10.1109/MLSP55214.2022.9943365.
- [7] Weron, A., Weron, R. (1995). Computer simulation of Lévy α -stable variables and processes. In: Garbaczewski, P., Wolf, M., Weron, A. (eds) *Chaos — The Interplay Between Stochastic and Deterministic Behaviour*. Lecture Notes in Physics, vol 457. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/3-540-60188-067>

Avaliação de Técnicas de Inicialização Aplicadas a Versões Modificadas do Algoritmo LBG

Verusca Severo
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife-PE, Brasil
E-mail: verusca.severo@poli.br

Jair Guimarães
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife-PE, Brasil
E-mail: jgdan@poli.br

Francisco Madeiro
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife-PE, Brasil
E-mail: madeiro@poli.br

Resumo—A quantização vetorial (QV) é comumente empregada na compressão de sinais. A escolha do dicionário inicial do algoritmo LBG (Linde-Buzo-Gray), que é amplamente utilizado para produção de dicionários, repercute no desempenho da QV. Este trabalho avalia o impacto de técnicas de inicialização no algoritmo LBG e em suas versões aceleradas, no tocante à qualidade das imagens reconstruídas. Os resultados das simulações mostram que a utilização de técnicas de inicialização permite obter resultados de PSNR (do inglês, *peak signal to noise ratio*) superiores aos obtidos com a inicialização aleatória.

Palavras-chave—Quantização Vetorial, LBG, Técnicas de inicialização.

I. INTRODUÇÃO

Dentre as técnicas usadas para diminuir o número de *bits* para representar um sinal, destaca-se a quantização vetorial (QV), que permite uma alta taxa de compressão [1]. Na QV, cada vetor do sinal, $\vec{x} \in \mathbb{R}^K$, é mapeado, de acordo com o critério de distância mínima, para um vetor pertencente a um subconjunto finito W de $\mathbb{R}^K$, denominado dicionário.

O projeto de W parte comumente de um conjunto de treino, em que se busca encontrar N vetores-código para representar os vetores de treino de modo a minimizar a distorção introduzida na representação [1]. O algoritmo Linde-Buzo-Gray (LBG) [2] é o mais utilizado para projetar W . O LBG parte de um dicionário inicial, W_0 , que é melhorado iterativamente.

A escolha de W_0 repercute na qualidade da QV. Estudos apontam que a utilização de técnicas de inicialização [3–5] podem fornecer melhorias na qualidade dos sinais reconstruídos.

Neste trabalho, é investigada a utilização de três técnicas de inicialização, no tocante à qualidade das imagens reconstruídas, do algoritmo LBG e de suas versões aceleradas [6, 7].

II. ALGORITMO LBG

Seja $Y = \{\vec{y}_m, m = 1, \dots, M\}$ um conjunto de treino e $W_0 = \{\vec{w}_n, n = 1, \dots, N\}$ o dicionário inicial, ambos de dimensão K . Seja t o número de iteração e γ um valor pré-definido. O algoritmo LBG pode ser descrito como segue:

- **Passo 1:** Aloque cada vetor do dicionário W_t na região de Voronoi S_n de acordo com

$$S_n = \{\vec{y}_m : d(\vec{y}_m, \vec{w}_n) < d(\vec{y}_m, \vec{w}_i) \forall i \neq n\}, \quad (1)$$

em que $d(\vec{y}_m, \vec{w}_n)$ é a distância Euclidiana quadrática. Calcule a distorção total de acordo com

$$D_t = \sum_{n=1}^N \sum_{\vec{y}_m \in S_n} d(\vec{y}_m, \vec{w}_n). \quad (2)$$

- **Passo 2:** Se $(D_{t-1} - D_t)/D_t \leq \gamma$, o algoritmo deve encerrar, retornando W_t como dicionário final. Caso contrário, continue.
- **Passo 3:** Calcule os novos vetores-código como a média dos vetores de cada região de Voronoi, como segue

$$w_{nj} = \frac{1}{|S_n|} \sum_{\vec{y}_m \in S_n} y_{mj}, \quad (3)$$

em que $|S_n|$ é quantidade de vetores de treino na região S_n e y_{mj} é a j -ésima componente de $\vec{y}_m$. Faça $t = t + 1$ e retorne ao **Passo 1**.

A. LBG acelerado

Em [6], Lee, Baek e Sung propuseram o uso de um fator de escala na atualização do dicionário do LBG (**Passo 3**).

Os vetores-código são calculados como segue

$$w_{nj}^{\text{nov}} = w_{nj}^{\text{atual}} + s \cdot (w_{nj}^{\text{calculado}} - w_{nj}^{\text{atual}}), \quad (4)$$

em que $w_{nj}^{\text{calculado}}$ é a j -ésima componente do vetor-código calculada pela Equação (3) e s é o fator de escala. Se $s = 1$, tem-se que a versão acelerada corresponde ao LBG. Em [6], os autores sugerem $1 < s < 2$.

B. LBG acelerado com escala variável

Em [7], Paliwal e Ramasubramanian propuseram o uso de um fator escala que varia em função de t , como segue

$$s = 1 + \frac{h}{h + t}, \quad (5)$$

em que h é uma constante positiva.

A Equação (5) satisfaz a sugestão proposta em [6] de utilizar $1 < s < 2$ visando obter uma convergência mais rápida do que o LBG padrão.

III. INICIALIZAÇÕES

Há na literatura diversas técnicas destinadas à geração de dicionários iniciais. Esta seção apresenta algumas delas.

A. KATSA

Proposta por Katsavounidis, Kuo e Zhang em [8], a técnica baseia-se na ideia de que os vetores de treino com a maior distância dentre todos os vetores têm mais chances de pertencerem a regiões de Voronoi diferentes.

A técnica inicia calculando a norma de cada vetor de treino. Seleciona-se o vetor com a maior norma como o primeiro vetor-código do dicionário. Em seguida, calcula-se a distância entre todos os vetores de treino restantes e o primeiro vetor-código para que o vetor com a maior distância seja definido como o segundo vetor-código. O processo se repete com o cálculo da distância entre os vetores de treino e os vetores-código já selecionados, até se obterem N vetores-código.

B. MEIM

Proposto por Nyeck e Tosser-Roussey em [9], a técnica MEIM (do inglês, *Maximum Entropy Initialization Method*) busca garantir que as regiões de Voronoi iniciais tenham tamanhos aproximadamente iguais.

Seja $X = \{\vec{x}_q, q = 1, \dots, N\}$ um dicionário inicial aleatório. Seja S_q a q -ésima região de Voronoi de X , com N_q vetores treino. Faça $N_q = 1 \forall q$. Se

$$N_q d(\vec{y}_m, \vec{x}_q) < N_p d(\vec{y}_m, \vec{x}_p) \forall p \neq q, \quad (6)$$

então o vetor de treino $\vec{y}_m$ é alocado em S_q e $N_q = N_q + 1$. O processo é repetido para todo o conjunto de treino Y . Em

seguida, é feito o cálculo dos centroides de acordo com a Equação (3). O dicionário final é definido pelos vetores de X que mais se assemelham com os centroides calculados.

C. DSICS

Proposta por Hu *et al.* em [10], a técnica DSICS (do inglês, *Double Sorting-based Initial Codewords Selection*) é resultante da junção de duas técnicas de inicialização [11, 12].

Sejam Y_1 e Y_2 cópias do conjunto de treino Y . Para cada vetor de Y_1 , encontre a componente com o menor valor e subtraia esse valor de cada componente em seu respectivo vetor. Em seguida, calcule a distância Euclidiana de cada vetor em Y_1 à origem e ordene Y_1 em termos dessas distâncias. Para cada vetor de Y_2 , calcule a soma de seus componentes e classifique Y_2 em termos dessas somas. Divida Y_1 e Y_2 em N subgrupos. Para cada subgrupo Y_{1i} e Y_{2i} , com $i = 1, \dots, N$, se $Y_{1i} \cap Y_{2i} \neq \emptyset$, o ponto médio de $Y_{1i} \cap Y_{2i}$ é definido como um vetor-código. Caso contrário, o i -ésimo vetor-código é definido como o ponto médio de Y_{1i} .

IV. RESULTADOS

Os resultados foram obtidos para dicionários projetados com $K = 16$ (blocos de 4×4 pixels) e $N = 32, 64, 128, 256$ e 512 , utilizando como conjunto de treino imagens de tamanho 256×256 pixels, codificadas a 8 bpp (256 níveis de cinza). Foi utilizado $\gamma = 0,001$ e $s = 1,8$.

Os dicionários iniciais foram criados a partir dos conjuntos de treino. Na inicialização “aleatória”, tem-se um conjunto de 10 dicionários inicializados aleatoriamente. Esse conjunto foi considerado na técnica MEIM (que faz uso de aleatoriedade). Foram realizadas 30 execuções dos algoritmos com as inicializações aleatória e MEIM.

A Tabela I apresenta os resultados de PSNR médio das imagens Clock e Lena reconstruídas, obtidos com o algoritmo LBG. Para ambas as imagens e todo N , a inicialização Katsa apresenta melhor desempenho em comparação com as demais técnicas. O ganho mais significativo obtido com Katsa sobre a inicialização aleatória é observado em $N = 512$ com ganho de 4,76 dB para imagem Clock e 1,72 para a imagem Lena.

TABELA I: Valores médios de PSNR (dB) para o LBG padrão.

Imagem	Método	N				
		32	64	128	256	512
Clock	LBG (aleat.)	26,25	27,17	28,23	29,45	30,74
	LBG (Meim)	26,26	27,16	28,25	29,41	30,67
	LBG (Katsa)	26,87	28,03	29,59	31,64	35,50
	LBG (Dsics)	29,98	27,16	28,03	29,35	30,91
Lena	LBG (aleat.)	26,59	27,72	28,77	29,87	31,08
	LBG (Meim)	26,57	27,69	28,73	29,78	30,95
	LBG (Katsa)	26,60	27,73	28,94	30,62	32,80
	LBG (Dsics)	26,49	27,73	28,71	29,87	31,24

Para o algoritmo LBG com escala fixa, os resultados da Tabela II indicam um comportamento similar ao obtido com o algoritmo LBG, em que para Clock e Lena, considerando todo N , a inicialização Katsa apresenta o melhor desempenho entre as técnicas de inicialização.

TABELA II: Valores médios de PSNR (dB) para o LBG com escala fixa (LBG-1).

Imagem	Método	N				
		32	64	128	256	512
Clock	LBG-1 (aleat.)	26,37	27,31	28,41	29,72	31,20
	LBG-1 (Meim)	26,40	27,35	28,42	29,69	31,11
	LBG-1 (Katsa)	26,90	28,23	29,73	31,90	35,48
	LBG-1 (Dsics)	26,21	27,33	28,30	29,63	31,38
Lena	LBG-1 (aleat.)	26,63	27,81	28,96	30,22	31,65
	LBG-1 (Meim)	26,63	27,77	28,93	30,15	31,54
	LBG-1 (Katsa)	26,59	27,87	29,16	30,77	33,04
	LBG-1 (Dsics)	26,63	27,80	29,06	30,14	31,71

A Tabela III apresenta os valores de PSNR médio para o LBG com escala variável. Para Lena, com $N \geq 64$, a inicialização Katsa apresenta os melhores resultados; para $N = 32$ a inicialização Dsics apresenta o melhor desempenho.

Para Clock, a substituição da inicialização aleatória por Katsa leva a ganhos de PSNR superiores a 0,52 dB; particularmente, para $N = 512$, tem-se um ganho de 4,18 dB.

TABELA III: Valores médios de PSNR (dB) para o LBG com escala variável (LBG-2).

Imagem	Método	N				
		32	64	128	256	512
Clock	LBG-2 (aleat.)	26,31	27,25	28,40	29,71	31,24
	LBG-2 (Meim)	26,32	27,28	28,39	29,67	31,14
	LBG-2 (Katsa)	26,83	28,05	29,54	31,91	35,42
	LBG-2 (Dsics)	26,13	27,14	28,24	29,59	31,41
Lena	LBG-2 (aleat.)	26,62	27,79	28,97	30,22	31,72
	LBG-2 (Meim)	26,62	27,76	28,88	30,13	31,57
	LBG-2 (Katsa)	26,54	27,88	29,15	30,73	33,04
	LBG-2 (Dsics)	26,63	27,76	29,00	30,18	31,75

V. CONCLUSÃO

Este trabalho apresentou a avaliação de três diferentes técnicas de inicialização do algoritmo LBG e de suas versões aceleradas (com escala fixa e escala variável).

As técnicas de inicialização avaliadas permitiram obter dicionários com qualidade superior aos obtidos com o uso de inicialização aleatória, o que levou à obtenção de imagens reconstruídas com melhor qualidade em termos de PSNR. Com a substituição da inicialização aleatória pela técnica Katsa, foram obtidos resultados com diferença de PSNR médio superior a 4 dB para a imagem Clock com $N = 512$.

AGRADECIMENTOS

Este trabalho conta com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERÊNCIAS

- [1] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. USA: Kluwer Academic Publishers, 1991.
- [2] Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design,” *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.
- [3] V. Severo, F. B. S. Ferreira, R. Spencer, A. Nascimento, and F. Madeiro, “On the initialization of swarm intelligence algorithms for vector quantization codebook design,” *Sensors*, vol. 24, no. 8, 2024.
- [4] A. Nascimento, R. Spencer, V. Severo, and F. Madeiro, “Impacto da inicialização no desempenho do algoritmo FSS-LBG,” in *XI Conferência Nacional em Comunicações, Redes e Segurança da Informação. Campina Grande, PB, Brasil: IECOM, 2021, 2021*, pp. 1–7.
- [5] R. Spencer, N. Barros, P. Amorim, V. Severo, and F. Madeiro, “Inicialização do algoritmo do vagalume aplicado ao projeto de dicionários,” in *XIV Congresso Brasileiro de Inteligência Computacional. Curitiba, PR: ABRICOM, 2019*, pp. 1–7.
- [6] D. Lee, S. Baek, and K. Sung, “Modified K -means algorithm for vector quantizer design,” *IEEE Signal Processing Letters*, vol. 4, no. 1, pp. 2–4, 1997.
- [7] K. K. Paliwal and V. Ramasubramanian, “Comments on “modified K -means algorithm for vector quantizer design,”,” *IEEE Transactions on Image Processing*, vol. 9, no. 11, pp. 1964–1967, 2000.
- [8] I. Katsavounidis, C.-C. J. Kuo, and Z. Zhang, “A new initialization technique for generalized Lloyd iteration,” *IEEE Signal Processing Letters*, vol. 1, no. 10, pp. 144–146, 1994.
- [9] A. Nyeck and A. Tosser-Roussey, “Maximum entropy initialisation technique for image coding vector quantiser design,” *Electronics Letters*, vol. 28, no. 3, pp. 273–274, 1992.
- [10] K.-C. Hu, C.-H. Chen, C.-W. Tsai, and M.-C. Chiang, “An enhanced initialization method for codebook generation,” in *2015 IEEE International Conference on Consumer Electronics-Taiwan*. IEEE, 2015, pp. 92–93.
- [11] M. Yedla, S. R. Pathakota, and T. M. Srinivasa, “Enhancing K -means clustering algorithm with improved initial center,” *International Journal of Computer Science and Information Technologies*, vol. 1, no. 2, pp. 121–125, 2010.
- [12] S. Vimala, “Techniques for generating initial codebook for vector quantization,” in *2011 3rd International Conference on Electronics Computer Technology*, vol. 4, 2011, pp. 201–208.

Ferramenta de auxílio ao estudo de transformadas DCT e DWT na compressão de imagens

Gabriel de Assunção Ferreira
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brasil
gaf@poli.br

Francisco Madeiro
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brasil
madeiro@poli.br

Verusca Severo de Lima
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brasil
verusca.severo@poli.br

Resumo—O trabalho apresenta uma ferramenta de apoio ao aprendizado da transformada discreta do cosseno (DCT, do inglês *discrete cosine transform*) e da transformada de wavelet discreta (DWT, do inglês *discrete wavelet transform*) aplicada à compressão de imagens. A ferramenta foi utilizada nas disciplinas de sinais e sistemas e processamento digital de sinais pertencentes ao curso de engenharia da UPE (Universidade de Pernambuco) e tem como objetivo auxiliar o aprendizado dos principais aspectos das transformadas de maneira prática. Por fim, o trabalho apresenta a avaliação dos alunos sobre a utilização da ferramenta em sala de aula, a qual apontou que a maioria dos alunos concordam que a ferramenta auxilia na compreensão dos tópicos apresentados em sala de aula e que trouxe a percepção do processamento abordado com clareza.

Palavras-chave—Processamento Digital de Sinais, Sinais e Sistemas, Compressão de Imagens, Ferramenta didática.

I. INTRODUÇÃO

O uso de transformadas tem importante papel em trabalhos que abordam processamento de áudio [1], imagens [2] e vídeos [3], supressão de ruído [4], telemedicina [5] e em aprendizagem de máquina [6].

O desenvolvimento de jogos para educação [7] ou ferramentas de apoio ao aprendizado [8] para assuntos já estabelecidos na literatura, com o propósito de facilitar o ensino e engajamento dos alunos das áreas de tecnologia, tem se tornado cada vez mais imprescindíveis.

Nesse contexto, o trabalho tem como propósito a elaboração e apresentação de uma ferramenta didática para a abordagem de características e familiarização com aspectos da transformada discreta do cosseno [9] e da transformada wavelet discreta [10] aplicada à compressão de imagens digitais.

II. METODOLOGIA

A ferramenta proposta foi desenvolvida com a linguagem de programação “Python” para o *backend*, com a utilização da biblioteca “cv2” para processamentos da transformada DCT e “pywt” para a transformada DWT. Para que o usuário pudesse interagir com a ferramenta, foi utilizado o *framework* “React”, permitindo que o usuário possa utilizar tanto computadores ou *notebooks* quanto o celular. O objetivo foi permitir uma maior flexibilização e facilidade na hora da escolha do equipamento a ser utilizado ao se realizar a atividade prática. A aplicação esteve hospedada nos servidores do “Render”.

O trabalho foi dividido em duas etapas: uma aula para abordagem teórica e discussão dos principais tópicos das transformadas e, posteriormente, foi proposto um guia prático para a utilização da ferramenta contendo exercícios para fixação do conteúdo e familiarização com os aspectos das transformadas. Ao fim, os alunos responderam, de forma

anônima, um questionário de avaliação sobre a ferramenta disponibilizado no “Google Forms”.

III. FERRAMENTA

A ferramenta é dividida em dois módulos, os quais apenas diferem no tipo da transformada a ser utilizada para realizar o processamento, seus principais aspectos e resultados. Ambos os módulos retornam uma imagem após o processamento escolhido e suas respectivas métricas de MSE (*Mean Squared Error*), SSIM (*Structural Similarity Index Measure*) e PSNR (*Peak Signal-to-Noise Ratio*). O processamento consiste na aplicação da transformada escolhida sobre a imagem, a condição de quais coeficientes devem ser mantidos e, por fim, a aplicação da transformada inversa.

A Fig. 1. apresenta o primeiro módulo da ferramenta e tem como objetivo a utilização da DCT. O aluno pode selecionar a imagem no banco de dados de imagens, o qual é composto por 8 imagens (*Barbara, Boat, Clock, Elaine, Goldhill, Lena, Mandril, Peppers e Tiffany*) em níveis de cinza e dispostos em 256 por 256 pixels, o tamanho do bloco (composto pelos valores de 4x4, 8x8, 16x16 etc.) a ser utilizado na transformada DCT e quantidade de coeficientes (1, 3, 5, 10, 20) a serem mantidos no processamento.

O segundo módulo é apresentado na Fig. 2. e diz respeito ao uso da transformada DWT. O usuário seleciona uma imagem, o número de níveis (1 ou 2 níveis) de decomposição desejado e quais das sub-bandas devem ser mantidas (as demais têm os coeficientes zerados). Selecionado apenas 1 nível, o usuário pode selecionar as sub-bandas de aproximação, de detalhe horizontal, vertical e diagonal tanto separadamente quanto combinadas entre si. Já em 2 níveis, o usuário pode selecionar a sub-banda de coeficientes de aproximação de nível 2, a sua combinação com os coeficientes de detalhes de nível 2 ou a combinação de todas as sub-bandas.



Fig. 1. Primeiro módulo da ferramenta: transformada DCT.

IV. RESULTADOS

Na análise dos questionários de avaliação da ferramenta, o total de 19 alunos concordaram sobre o auxílio na compreensão de tópicos apresentados na disciplina, sendo que

aproximadamente metade apenas concordou e a outra metade concordou totalmente com a declaração.

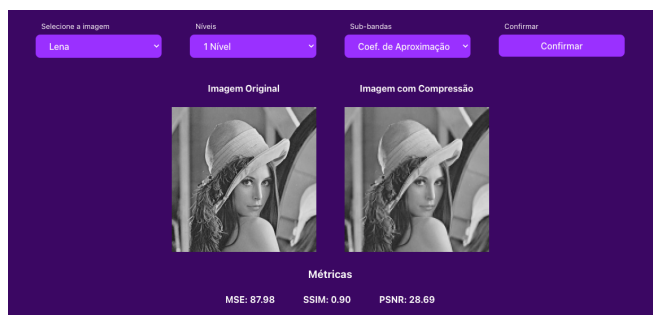


Fig. 2. Segundo módulo da ferramenta: transformada DWT.

Em relação à declaração: “As imagens resultantes da aplicação da transformada trouxeram a percepção do processamento abordado em sala de aula com clareza.”, foi possível observar que todos os alunos estiveram de acordo, conforme pode ser visto na Fig. 3.

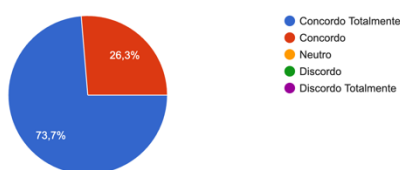


Fig. 3. Resultado do questionário sobre a declaração: As imagens resultantes da aplicação da transformada trouxeram a percepção do processamento abordado em sala de aula com clareza.

No tocante à visualização na prática do conteúdo abordado em sala de aula a partir da ferramenta, 63,2% dos alunos concordaram totalmente com a declaração, 31,6% concordaram e apenas 5,3% se mantiveram neutros.

Considerando a facilidade ao utilizar a ferramenta, enquanto apenas 10,5% dos alunos se mantiveram neutros, o restante concordou ou concordou totalmente que a ferramenta é de fácil uso, conforme ilustra a Fig. 4. Além disso, apenas 21,1% mencionaram que ocasionalmente necessitou consultar especialistas para obter maior entendimento sobre a utilização da ferramenta, enquanto o restante raramente necessitou ou não precisou.

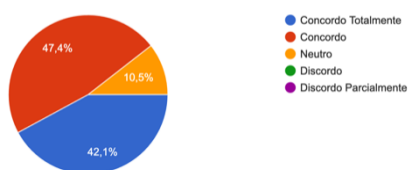


Fig. 4. Resultado do questionário sobre a declaração: Foi fácil utilizar a ferramenta.

Em relação à afirmação “A ferramenta estimula o aluno a desenvolver um Trabalho de Conclusão de Curso, um Projeto de Pesquisa ou uma dissertação futura de mestrado com temas relacionados ao experimento.”, 52,6% afirmaram que concordam totalmente com a declaração, 31,6% concordaram e apenas 15,8% dos participantes se posicionaram neutro, resultando em um índice de concordância com a declaração maior do que 80%, conforme pode ser visto na Fig. 5.

Entre algumas das sugestões dos alunos, é válido citar a recomendação de adição de balões de *tooltips* para pequenas explicações de conceitos como, por exemplo, as definições de SSIM, PSNR e MSE e a possibilidade de se colocar lado a

lado as imagens resultantes dos processamentos da DCT e DWT para fins de comparação. Essas sugestões serão levadas em consideração para melhoria da ferramenta e futuros projetos.

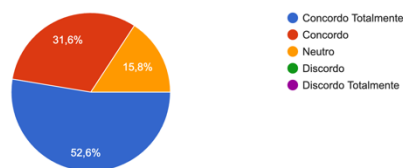


Fig. 5. Resultado do questionário sobre a declaração: A ferramenta estimula o aluno a desenvolver um Trabalho de Conclusão de Curso, um Projeto de Pesquisa ou uma dissertação futura de mestrado com temas relacionados ao experimento.

V. CONCLUSÃO

Com base nas informações obtidas no questionário de avaliação, é possível afirmar que o trabalho proposto apresenta bons resultados em termos do auxílio na compreensão de transformadas. Além disso, o trabalho tem como resultado o incentivo ao aluno desenvolver um trabalho de conclusão de curso, iniciação científica ou um possível trabalho de pós-graduação com temas que foram abordados na ferramenta.

AGRADECIMENTOS

Este trabalho conta com apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERÊNCIAS

- [1] S. K. Abu, J. R. Jorquia, J. M. Mendoza, C. T. Tolentino and C. R. Lucas, "Lossless Audio Compression using DWT, DCT and Huffman-based LZW Encoding," 2023 22nd International Symposium on Communications and Information Technologies (ISCIT), Sydney, Australia, 2023, pp. 357-362.
- [2] H. Wen, Y. Huang and Y. Lin. High-quality Color Image Compression-Encryption Using Chaos and Block Permutation. Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 8, 2023, pp. 101660.
- [3] Zhang, F. et al. "Exploring Consistent Spatio-Temporal Distortion and Stable 3-D DCT Coefficients for Robust Blind Video Watermarking". 2024 IEEE International Conference on Acoustics, Speech and Signal Processing, Seoul, Korea, Republic of, 2024, pp. 4595-4599.
- [4] V. Van Duong, T. N. Huu, J. Yim and B. Jeon, "Hybrid Light Field Image Denoising Network using 4D-DCT Separated Transform," 2023 IEEE International Conference on Visual Communications and Image Processing (VCIP), Jeju, Korea, Republic of, 2023, pp. 1-5
- [5] R. Ch. N. Vivek K, G. Srivastava and R. Gadekallu. ECDSA-based Tamper Detection in Medical Data Using a Watermarking Technique. International Journal of Cognitive Computing in Engineering, vol. 5, 2024, pp. 78-87.
- [6] J. Lee and H. Kim, "DCT-ViT: High-Frequency Pruned Vision Transformer With Discrete Cosine Transform," in IEEE Access, vol. 12, pp. 80386-80396, 2024.
- [7] T. H. Laine and R. S. N. Lindberg, "Designing Engaging Games for Education: A Systematic Literature Review on Game Motivators and Design Principles," in IEEE Transactions on Learning Technologies, vol. 13, no. 4, pp. 804-821, 1 Oct.-Dec. 2020.
- [8] N. F. Duarte Filho e R. F. L. Rodrigues, "O Uso de Ferramentas Tecnológicas e o Estudo das Bases Conceituais da Educação Profissional e Tecnológica: Intersecções Entre Teoria e Prática no Mestrado ProfEPT", RENOTE, vol. 21, nº 2, p. 13-22, dez. 2023.
- [9] N. Ahmed, T. Natarajan and K. R. Rao, "Discrete Cosine Transform," in IEEE Transactions on Computers, vol. C-23, no. 1, pp. 90-93, Jan. 1974.
- [10] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674-693, July 1989

Uma Nova Técnica de Projeto de Quantizadores Vetoriais Baseada no Algoritmo *Cuckoo Search*

Rodrigo Calazans
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife-PE, Brasil
E-mail: rshc@poli.br

Verusca Severo
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife-PE, Brasil
E-mail: verusca.severo@poli.br

Francisco Madeiro
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife-PE, Brasil
E-mail: madeiro@poli.br

Resumo—Este trabalho apresenta um novo algoritmo de projeto de dicionário para quantização vetorial que combina o algoritmo bioinspirado *Cuckoo Search* (CS) com o algoritmo LBG (Linde-Buzo-Gray). A nova técnica proposta é denominada Cuckoo-LBG e foi aplicada à compressão de imagens baseada em quantização vetorial. O Cuckoo-LBG se mostrou superior ao algoritmo LBG para todas as imagens avaliadas em termos de PSNR (*Peak Signal-to-Noise Ratio*) das imagens reconstruídas.

Palavras-chave—Quantização Vetorial, Algoritmo Bioinspirado, *Cuckoo Search*, LBG.

I. INTRODUÇÃO

A Quantização Vetorial (QV) é uma técnica de compressão de sinais [1]. Em se tratando de compressão de imagens baseada em QV, a qualidade das imagens reconstruídas está fortemente atrelada à qualidade do dicionário projetado.

Algoritmos para projeto de dicionário têm sido amplamente explorados na literatura. Dentre eles, pode ser citado o LBG (Linde-Buzo-Gray) [2], cujo desempenho depende do dicionário inicial [3].

O projeto de dicionários pode se beneficiar de técnicas de inicialização [4–9], bem como de algoritmos que fazem uso da inteligência de enxames [10, 11].

Neste trabalho, é apresentado um novo método de projeto de dicionário, denominado Cuckoo-LBG, que combina a otimização do algoritmo *Cuckoo Search* com o LBG.

II. ALGORITMO LBG

O LBG é um algoritmo iterativo em que um dicionário inicial $W = \{\vec{w}_i, i = 1, \dots, N\}$ é fornecido como entrada. Dado um conjunto de treino $X = \{\vec{x}_j, j = 1, \dots, M\}$, geralmente com $M \gg N$, e uma dimensão K dos vetores $\vec{w}_i$ e $\vec{x}_j$, a cada iteração é realizado um particionamento em regiões de *Voronoi* como segue

$$S_i = \{\vec{x} : d(\vec{x}, \vec{w}_i) < d(\vec{x}, \vec{w}_p) \forall p \neq i\}, \quad (1)$$

com

$$d(\vec{x}_m, \vec{w}_i) = \sum_{q=1}^K (x_{m,q} - w_{i,q})^2, \quad (2)$$

em que $d(\cdot, \cdot)$ é a distância Euclidiana quadrática, $x_{m,q}$ é a q -ésima componente do vetor $\vec{x}_m$ e $w_{i,q}$ é a q -ésima componente do vetor $\vec{w}_i$.

Em seguida, a distorção é calculada. O critério de parada é avaliado: quando a redução percentual da distorção é menor que um valor pré-estabelecido, o algoritmo para; caso contrário, o mesmo prossegue com o cálculo dos centroides

$$\vec{w}_i = \frac{1}{M_i} \sum_{\vec{x}_m \in S_i} \vec{x}_m, \quad (3)$$

em que M_i é a quantidade de vetores $\vec{x}_m$ alocados em S_i . O novo vetor-código da região passa a ser o centroide atual e o algoritmo se repete com um novo particionamento.

III. CUCKOO SEARCH

O *Cuckoo Search* é inspirado no hábito de reprodução dos pássaros da família *Cuculidae*, conhecidos popularmente por cucos [12]. As fêmeas cucos põem seus ovos em ninhos de outras espécies, fazendo com que as fêmeas hospedeiras choquem e alimentem o filhote cuco [13].

A equação da atualização do movimento de um cuco é [14]

$$\vec{w}_i^{t+1} = \vec{w}_i^t + \alpha \cdot s \otimes H(p_a - \epsilon) \otimes (\vec{w}_i^t - \vec{w}_j^t), \quad (4)$$

em que t representa a iteração atual, α é um fator de escala do passo, s é o tamanho do passo, $\otimes$ é uma operação de multiplicação entre os elementos em posições correspondentes, $H(\cdot)$ é a função de *Heaviside* com descontinuidade em a , p_a é a probabilidade de os ovos depositados serem descobertos, representando os piores resultados, $\epsilon \in [0, 1]$ e $\vec{w}_j$ é um vetor escolhido aleatoriamente.

IV. CUCKOO-LBG

No Cuckoo-LBG, cada dicionário é um cuco e a função objetivo corresponde ao inverso da distorção média. Os cucos iniciais podem ser gerados a partir de estratégias de inicialização ou de forma aleatória.

Inicia-se com a etapa de particionamento, em que todos os dicionários são particionados e a distorção média de cada um deles é calculada para definir o dicionário de maior *fitness*. Na sequência, os dicionários são atualizados seguindo a Equação (4). Um novo voo é realizado para uma fração p_a dos dicionários com os piores desempenhos. Neste caso, a equação de atualização, Equação (4), é alterada para

$$\vec{w}_i^{t+1} = \vec{w}_i^t + \alpha \cdot s \cdot (\vec{w}_i^t - \vec{w}_{best,i}^t), \quad (5)$$

em que $\vec{w}_{best,i}^t$ é o dicionário com o melhor desempenho.

Em seguida, o critério de parada é avaliado. Se a condição de parada não for alcançada, atualize os dicionários a partir do cálculo dos centroides realizado pelo LBG e retorne ao passo do particionamento. Se o critério de parada for atingido, o melhor cuco será o dicionário final projetado pelo algoritmo.

V. RESULTADOS

Para os resultados, foram utilizadas as imagens *Barbara*, *Clock*, *Lena* e *Peppers* no formato PGM (do inglês, *Portable Gray Map*), de tamanho 256×256 pixels, codificadas a 8 bpp (256 níveis de cinza).

Os dicionários foram projetados com $K = 16$, ou seja, blocos de 4×4 pixels, e tamanhos $N = 32, 64, 128, 256$ e 512 . Dessa forma, foram obtidas taxas R iguais a 0,3125 bpp, 0,375 bpp, 0,4375 bpp, 0,5 bpp e 0,5625 bpp, respectivamente. O limiar utilizado como critério de parada foi $\mu = 0,001$.

Foram realizadas 30 execuções e os resultados foram obtidos em termos da média aritmética ao final das execuções.

As Tabelas I e II apresentam os valores médios de PSNR (*Peak Signal-to-Noise Ratio*) das imagens reconstruídas utilizando os dicionários projetados pelas técnicas LBG e Cuckoo-LBG.

TABELA I: Valores médios de PSNR (dB) para as imagens *Barbara* e *Clock*.

N	<i>Barbara</i>		<i>Clock</i>	
	LBG	Cuckoo-LBG	LBG	Cuckoo-LBG
32	24,75	24,80	26,23	26,72
64	25,67	25,79	27,23	27,97
128	26,68	26,84	28,29	29,25
256	27,76	28,08	29,50	30,80
512	29,06	29,63	30,87	32,40

TABELA II: Valores médios de PSNR (dB) das imagens *Lena* e *Peppers*.

N	<i>Lena</i>		<i>Peppers</i>	
	LBG	Cuckoo-LBG	LBG	Cuckoo-LBG
32	26,61	26,67	26,07	26,15
64	27,74	27,88	27,36	27,45
128	28,83	29,14	28,52	28,78
256	29,90	30,52	29,65	30,27
512	31,11	32,25	30,93	32,22

Para a imagem *Barbara*, o Cuckoo-LBG foi superior em termos de PSNR médio das imagens reconstruídas quando comparado com o algoritmo LBG. Observa-se que esta superioridade aumenta com o tamanho do dicionário, N .

Para a imagem *Clock*, o Cuckoo-LBG leva aos maiores valores de PSNR. O ganho de PSNR médio para o tamanho $N = 256$ é de 1,30 dB e de 1,53 dB para $N = 512$.

Para a imagem *Lena*, para $N \geq 256$, a substituição de dicionários LBG por dicionários Cuckoo-LBG leva a ganhos de PSNR superiores a 0,60 dB. Para $N = 512$, particularmente, a substituição supracitada leva a um ganho de 1,14 dB.

Para a imagem *Peppers*, a superioridade do Cuckoo-LBG sobre o LBG, em termos de PSNR das imagens reconstruídas, é observada para todos os valores de N . A superioridade aumenta com o tamanho do dicionário.

VI. CONCLUSÃO

Neste trabalho, foi apresentada uma nova técnica para o projeto de dicionários, denominada Cuckoo-LBG. A nova técnica se mostrou superior ao algoritmo LBG para todas as imagens avaliadas em termos de PSNR das imagens reconstruídas.

Como trabalhos futuros pode-se citar a avaliação dos impactos de estratégias de inicialização de dicionário no algoritmo proposto e a avaliação do Cuckoo-LBG em outras aplicações, tais como segmentação de imagens e compressão de nuvens de pontos 3D.

AGRADECIMENTOS

Este trabalho conta com apoio financeiro da Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERÊNCIAS

- [1] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 2012, vol. 159.
- [2] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.
- [3] G. Patané and M. Russo, "The enhanced LBG algorithm," *Neural Networks*, vol. 14, no. 9, pp. 1219–1237, 2001.
- [4] K.-C. Hu, C.-H. Chen, C.-W. Tsai, and M.-C. Chiang, "An enhanced initialization method for codebook generation," in *2015 IEEE International Conference on Consumer Electronics-Taiwan*. IEEE, 2015, pp. 92–93.
- [5] A. Nyeck and A. Tosser-Roussey, "Maximum entropy initialisation technique for image coding vector quantiser design," *Electronics Letters*, vol. 28, no. 3, pp. 273–274, 1992.
- [6] I. Katsavounidis, C.-C. J. Kuo, and Z. Zhang, "A new initialization technique for generalized Lloyd iteration," *IEEE Signal Processing Letters*, vol. 1, no. 10, pp. 144–146, 1994.
- [7] X. Ma, Z. Pan, Y. Li, and J. Fang, "High-quality initial codebook design method of vector quantisation using grouping strategy," *IET Image Processing*, vol. 9, no. 11, pp. 986–992, 2015.
- [8] B. Mirzaei, H. Nezamabadi-Pour, and D. Abbasi-Moghadam, "An effective codebook initialization technique for LBG algorithm using subtractive clustering," in *2014 Iranian Conference on Intelligent Systems (ICIS)*. IEEE, 2014, pp. 1–5.
- [9] S. Chen and F. Li, "Initial codebook method of vector quantisation in Hadamard domain," *Electronics Letters*, vol. 46, no. 9, pp. 630–631, 2010.
- [10] M. Minu and R. A. Canessane, "An efficient squirrel search algorithm based vector quantization for image compression in unmanned aerial vehicles," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. IEEE, 2021, pp. 789–793.
- [11] K. Chiranjeevi and U. R. Jena, "Image compression based on vector quantization using cuckoo search optimization technique," *Ain Shams Engineering Journal*, vol. 9, no. 4, pp. 1417–1431, 2018.
- [12] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*. IEEE, 2009, pp. 210–214.
- [13] R. B. Payne and M. D. Sorensen, *The cuckoos*. Oxford University Press, 2005, vol. 15.
- [14] X.-S. Yang and S. Deb, "Cuckoo search: recent advances and applications," *Neural Computing and Applications*, vol. 24, no. 1, pp. 169–174, 2014.

Arranjo Refletor Pixelado empregando Algoritmo Genético para Aplicações em Redes NTN

Lucas de Oliveira Veiga

Laboratório de Telecomunicações (LabTel)
Universidade Federal de Itajubá (UNIFEI)

Itajubá, Brasil

lucas.oliveiraveiga@outlook.com

Danilo Henrique Spadoti

Laboratório de Telecomunicações (LabTel)
Universidade Federal de Itajubá (UNIFEI)

Itajubá, Brasil

spadoti@unifei.edu.br

Arismar Cerqueira Sodré Junior

Laboratório WOCA
Instituto Nacional de Telecomunicações (INATEL)
Santa Rita do Sapucaí, Brasil
arismar@inatel.br

Ramon Maia Borges

Laboratório de Telecomunicações (LabTel)
Universidade Federal de Itajubá (UNIFEI)
Itajubá, Brasil
ramonmb@unifei.edu.br

Resumo—Este trabalho apresenta um arranjo refletor com elementos em forma de *pixels*, projetado utilizando algoritmos genéticos para aplicações em Redes Não Terrestres (NTN). A célula refletora foi projetada em *Python*, utilizando a biblioteca PyAEDT e o software HFSS para simulações eletromagnéticas. O arranjo, composto de material RO3303 com permissividade elétrica igual a 3 e espessura de 0,51 mm, foi otimizado para minimizar o coeficiente de transmissão selecionado como função de aptidão. O algoritmo genético identificou as configurações mais eficientes, alcançando valores de S_{21} em torno de -30 dB na oitava interação após 15 horas de simulações, totalizando 256 simulações no HFSS. O método de *Floquet* foi utilizado para a análise das estruturas periódicas. O arranjo final, com dimensões de 12x12 cm, apresentou ganho entre 20,25 e 21,5 dBi na faixa de 17 a 22 GHz.

Palavras-chave—5G, algoritmo genético, arranjo refletor, NTN, *python*

I. INTRODUÇÃO

Os arranjos refletores, que combinam as vantagens das antenas parabólicas e planas, são importantes para as Redes Não Terrestres (NTN). Esses arranjos oferecem uma solução compacta e versátil para direcionamento de feixes, otimizando a transmissão e recepção de sinais. Diversos estudos investigam materiais, *design* e integração de tecnologias para maximizar o desempenho dos arranjos refletores. Entretanto, a modelagem e simulação numérica desses sistemas são tipicamente complexas e requerem grande capacidade computacional, prolongando o processo de simulação e aumentando a necessidade de *hardware* avançado [1], [2].

Avanços recentes em inteligência artificial (IA) têm sido importantes na otimização de projetos de antenas e arranjos refletores. Pois, permitem a análise simultânea de diversos parâmetros com maior agilidade, eliminando a necessidade de ajustes manuais extensivos e prevendo o desempenho das antenas sob diferentes condições. Tais fatores contribuem para a redução significativa do tempo e dos recursos necessários para o desenvolvimento. A aplicação de IA está expandindo

a capacidade de fornecer soluções de comunicação mais flexíveis e eficientes, melhorando a precisão dos modelos numéricos e maximizando o desempenho das antenas [3].

O presente trabalho propõe um arranjo refletor de antenas para aplicações em redes não terrestres, operando na faixa de frequência de 19,7 a 21,2 GHz, e reporta a metodologia e os resultados numéricos do projeto realizado com uso de algoritmo genético (GA). O objetivo é obter um arranjo refletor otimizado, que apresente características favoráveis à aplicação em NTN. Como principal contribuição, aponta-se uma solução com células refletoras em forma de *pixels*, que leva ao chamado arranjo refletor "pixelado", operando com ganho acima de 20 dBi na faixa de frequências estabelecida pelo 3GPP (*Third-Generation Partnership Project*) para NTN. Cabe enfatizar que a estrutura em forma de *pixels* traz vantagens de otimização topológica. O uso de *Python* para realizar e integrar todo o processo de desenvolvimento, desde a criação do modelo do arranjo ao uso de GA, complementa as contribuições do artigo.

II. PROJETO DA CÉLULA DO ARRANJO REFLETOR

O arranjo refletor apresentado na Fig. 1, onde destaca-se a célula refletora em forma de *pixels*, foi projetado em *Python* utilizando a biblioteca PyAEDT e o *High-Frequency Structure Simulator* (HFSS) da *Ansys* para a simulação eletromagnética de cada elemento. A célula emprega o material RO3303, com permissividade relativa (ϵ_r) igual a 3 e espessura de 0,51 mm. O desenvolvimento envolveu um algoritmo genético para explorar e otimizar diversas configurações. O coeficiente de transmissão foi selecionado como a função de aptidão, fundamental para avaliar e maximizar a eficiência da célula.

O método de *Floquet* foi empregado na simulação da célula refletora, sendo importante para a análise de estruturas periódicas. Este método permite a avaliação precisa da resposta em frequência dessas estruturas, considerando a periodicidade do campo eletromagnético em uma célula unitária

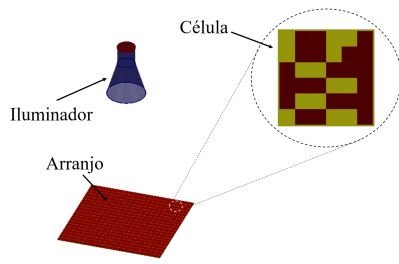


Fig. 1. Arranjo refletor.

e a interação entre os harmônicos resultantes. Com o método de *Floquet*, é possível determinar a resposta da estrutura a diferentes frequências e ângulos de incidência, assegurando uma simulação precisa do comportamento eletromagnético [4].

III. RESULTADOS NUMÉRICOS

A Fig. 2 reporta a aptidão (coeficiente de transmissão) em função das interações. O coeficiente de transmissão (S_{21}) alcançou uma média inferior a 0,001, equivalendo - 30 dB na oitava interação. Foram necessárias aproximadamente 15 horas, totalizando 256 simulações no HFSS, para que o GA conseguisse obter esse valor de aptidão pré-configurada. Apesar do tempo medido em horas, tem-se notoriamente agilidade no processo de otimização em comparação com projetos baseados em metodologias tradicionais.

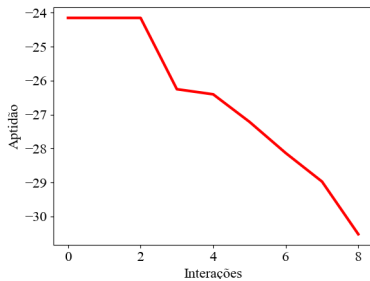


Fig. 2. Aptidão versus interações do GA.

A Fig. 3 apresenta o gráfico dos parâmetros S_{21} e S_{11} em função da frequência para a célula refletora. O parâmetro S_{11} indica o coeficiente de reflexão e permanece próximo de 0 dB ao longo de toda a faixa de frequência analisada, sugerindo alta reflexão do sinal incidente. Por outro lado, o parâmetro S_{21} apresentou um valor mínimo de -62,5 dB em 19,4 GHz. Os resultados indicam que a célula refletora foi projetada para maximizar a reflexão do sinal.

O ganho em função da frequência para o arranjo refletor com 400 células otimizadas, está apresentado na Fig. 4. Observa-se que o ganho varia entre 19,9 dBi e 21,7 dBi ao longo da faixa de frequência de 17 GHz a 22 GHz, demonstrando um desempenho consistente do arranjo refletor otimizado.

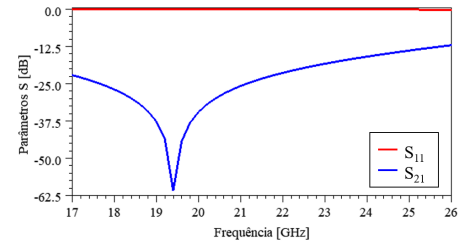


Fig. 3. Parâmetros S.

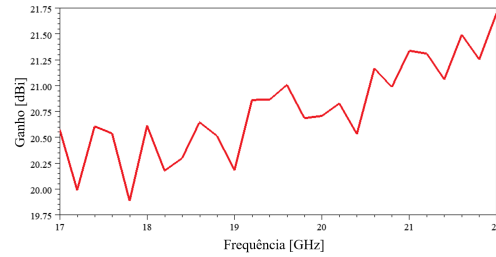


Fig. 4. Ganho do arranjo refletor de 12x12 cm.

IV. CONCLUSÕES

O estudo do arranjo refletor pixelado por meio de algoritmos genéticos mostrou-se eficaz na otimização de projetos de arranjos refletores para aplicações em redes não terrestres (NTN). A combinação das ferramentas PyAEDT e HFSS permitiu um *design* com otimização topológica, sendo o coeficiente de transmissão a função de aptidão principal. Os resultados obtidos demonstram que o arranjo final pixelado, com dimensões de 12x12 cm, apresentou um ganho consistente na faixa de 17 a 22 GHz, variando entre 20,25 e 21,5 dBi, e alcançando o valor da aptidão na oitava interação após 15 horas de simulações. Este estudo destaca a importância da integração de técnicas avançadas de otimização e ferramentas computacionais modernas para o desenvolvimento de soluções de comunicação mais eficientes e adaptativas.

V. AGRADECIMENTOS

Os autores agradecem o apoio financeiro da CAPES, CNPq-402093/2022-4, FAPEMIG e FAPESPE 2021/06569-1.

REFERENCES

- [1] Joy, J. A., Palaniswamy, S. K., Kumar, S., Kanagasabai, M., Bashyam, S. e Hussein, M. I. (2024). "Modern Reflectarray Antennas: A Review of the Design, State-of-the-Art, and Research Challenges". IEEE Access.
- [2] Theoharis, P. I., Raad, R., Tubbal, F., Khan, M. U. A. e Jamalipour, A. (2022). "Wideband reflectarrays for 5G/6G: A survey," IEEE Open Journal of Antennas and Propagation, 3, 871-901.
- [3] Prado, D. R., Lopez-Fernandez, J. A., Arrebola, M. e Goussetis, G. (2018). "Support vector regression to accelerate design and crosspolar optimization of shaped-beam reflectarray antennas for space applications," IEEE Transactions on Antennas and Propagation, 67(3), 1659-1668.
- [4] Bhattacharyya, A. K. (2006). "Phased array antennas: Floquet analysis, synthesis, BFNs and active array systems," (Vol. 179). John Wiley and Sons.

Avaliação de Conformidade de Equipamentos de Telecomunicações com Requisitos de Segurança Cibernética

Guilherme Rocha[†], Diego Pereira[†], Victor Pimentel[†], Ramon Fontes*, Augusto Neto* and Roger Immich*

*Universidade Federal do Rio Grande do Norte, Natal, RN

[†]Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, Parnamirim, RN

Abstract—O Ato 77 da Anatel define um conjunto de requisitos de segurança cibernética para dispositivos de telecomunicações. O objetivo é reduzir ou corrigir vulnerabilidades, seja através de atualizações de *software* ou *firmware*, ou por meio de recomendações de configuração. Este artigo descreve a realização de um trabalho de avaliação da conformidade de diversos produtos classificados como equipamentos terminais com conexão à Internet e equipamentos de infraestrutura de redes de telecomunicações com os requisitos estabelecidos pelo Ato 77. A pesquisa envolve uma análise dos critérios relacionados à área capacidade de mitigação de ataques. Os resultados revelam níveis variados de conformidade, destacando áreas críticas para melhorias e propondo recomendações para assegurar o pleno atendimento às exigências regulatórias. Este estudo visa contribuir para a melhoria da segurança cibernética dos equipamentos de telecomunicações no Brasil, a partir de análises de equipamentos disponíveis no mercado nacional.

Index Terms—Segurança Cibernética, Equipamentos de telecomunicações, Avaliação da conformidade, Análise de lacunas, Ato 77

I. INTRODUÇÃO

A segurança cibernética tem se tornado uma preocupação central à medida que o número de dispositivos conectados à Internet continua a crescer exponencialmente. Em resposta a essa crescente demanda por segurança, regulamentações específicas têm sido desenvolvidas por órgãos reguladores em todo o mundo para assegurar que os equipamentos de telecomunicações atendam a critérios rigorosos de proteção. No Brasil, o Ato 77 da Agência Nacional de Telecomunicações (Anatel) [1] é um exemplo significativo dessas iniciativas, estabelecendo um conjunto de requisitos de segurança cibernética para equipamentos de telecomunicações.

Publicado em janeiro de 2021, o Ato 77 foi elaborado com o objetivo de minimizar ou corrigir vulnerabilidades em equipamentos de telecomunicações por meio mecanismos automatizados para atualização segura de *software/firmware*, a implementação de métodos robustos de autenticação e criptografia, e a capacidade de mitigar ataques cibernéticos. A importância dessas medidas é reforçada pela Estratégia Nacional de Segurança Cibernética [2], que destaca a

necessidade de proteger a infraestrutura crítica do país contra vulnerabilidades e ataques.

Nesse contexto, a conformidade com o Ato 77 é fundamental para garantir que os dispositivos não apenas atendam aos padrões de segurança atuais, mas também sejam capazes de se adaptar a novas ameaças à medida que estas surgem. Ele deve ser aplicado a dispositivos presentes na Lista de Referência de Produtos para Telecomunicações¹, abrangendo tanto equipamentos terminais com conexão à Internet quanto equipamentos de infraestrutura de redes de telecomunicação. As suas diretrizes alinham-se com padrões internacionais, refletindo uma preocupação global com a integridade e a segurança das redes de telecomunicações.

Este estudo visa avaliar a conformidade de diversos produtos classificados como equipamentos terminais com conexão à Internet e de infraestrutura de redes de telecomunicações com os requisitos do Ato 77. A metodologia envolve uma análise detalhada dos dispositivos para verificar sua adesão aos critérios de segurança estabelecidos com a área de capacidade de mitigar ataques. Os resultados dessa avaliação são parte de um estudo inicial para identificar lacunas na conformidade e propor melhorias, contribuindo para o desenvolvimento da segurança cibernética na infraestrutura de telecomunicação de produtos comercializados no Brasil.

II. REVISÃO DA LITERATURA

A. Segurança Cibernética e Normas Internacionais

A segurança cibernética tem se tornado uma prioridade global, impulsionada pelo aumento significativo no número de dispositivos conectados à Internet. Diversos países têm estabelecido normas e regulamentos para garantir a proteção dos dados e a integridade das redes de telecomunicações. Nos Estados Unidos, por exemplo, o padrão *Federal Information Processing Standard* (FIPS) 140-3 define requisitos rigorosos para módulos criptográficos usados por agências governamentais e empresas que lidam com informações sensíveis. Esta norma especifica requisitos para o projeto e a implementação de algoritmos de criptografia, gestão de chaves criptográficas, mecanismos de autenticação, segurança física, dentre outras áreas.

¹Instituto Metrópole Digital, Universidade Federal do Rio Grande do Norte, Natal, RN.

²Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, Parnamirim, RN. Email: diego.pereira@ifrn.edu.br

¹<https://informacoes.anatel.gov.br/legislacao/atos-de-certificacao-de-produtos/2020/1493-ato-7280>

De forma semelhante, na União Europeia, o Regulamento Geral sobre a Proteção de Dados (*General Data Protection Regulation* - GDPR) [3] e a Diretiva de Segurança das Redes e da Informação (*NIS Directive*) são os principais marcos regulatórios que visam proteger dados pessoais e assegurar a segurança das redes de informação. A GDPR, em particular, impõe obrigações rigorosas sobre o tratamento e a proteção de dados pessoais, exigindo que organizações implementem medidas técnicas e organizacionais adequadas para garantir a segurança dos dados. A *NIS Directive*, por sua vez, foca na segurança das redes e dos sistemas de informação, exigindo que os operadores de serviços essenciais e os prestadores de serviços digitais implementem medidas de segurança robustas para proteger suas infraestruturas.

Outros países, como Japão, Singapura, Austrália e Índia também possuem regulamentações que tratam de forma específica o estabelecimento de requisitos de segurança cibernética para proteger as suas infraestruturas críticas de telecomunicações contra ataques.

B. Regulamentação de Segurança Cibernética no Brasil

No Brasil, a Anatel desempenha um papel fundamental na regulamentação e supervisão da segurança dos equipamentos de telecomunicações. O Ato 77, conforme destacado anteriormente, estabelece requisitos específicos de segurança cibernética para esses dispositivos. Este ato foi desenvolvido em consonância com um conjunto de legislações vigentes, tais como, Lei Geral de Telecomunicações [4], Regulamento para Avaliação da Conformidade e Homologação de Produtos para Telecomunicações [5], a Estratégia Nacional de Segurança Cibernética, a Lei Geral de Proteção de Dados Pessoais (LGPD) [6] e o Regulamento de Segurança Cibernética Aplicada ao Setor de Telecomunicações.

De forma geral, o Ato 77 busca minimizar ou corrigir vulnerabilidades através de atualizações seguras de *software/firmware*, fornecer recomendações detalhadas de configurações, a implementação de métodos adequados de autenticação e criptografia, e a capacidade de mitigar ataques cibernéticos. Além disso, o Ato exige que os equipamentos possuam mecanismos para gerenciamento remoto seguro e proteção de dados pessoais, alinhando-se às melhores práticas internacionais de segurança.

Enquadram-se produtos que possuem função de equipamento terminal com conexão à Internet ou de equipamento de infraestrutura de redes de telecomunicações, tais como roteadores, *gateways*, *firewalls*, modems, telefones celulares, computadores e dispositivos IoT. Devido à limitação de espaço, apenas parte dos requisitos definidos para o processo de conformidade e homologação junto à norma serão destacados no presente estudo.

O processo de certificação e homologação dos equipamentos, é realizado a partir do requerimento de homologação que contém uma declaração do requerente informando a quais requisitos listados o produto e o fornecedor atendem. A declaração deve contemplar o uso do princípio de *security by design*, proposto pelo

Open Project Application Security Project (OWASP), para desenvolvimento do produto, que consiste em aplicar as melhores práticas e padrões de segurança cibernética desde a fase de concepção e projeto dos equipamentos, visando minimizar as vulnerabilidades e maximizar a proteção dos dados e das informações [7].

O modelo *security by design* incorpora princípios como a minimização de dados, a defesa em profundidade, a segregação de funções, o princípio do menor privilégio, a criptografia por padrão, entre outros. O uso desse modelo pode contribuir para aumentar a confiança e a qualidade dos equipamentos de telecomunicações no Brasil, bem como para atender às expectativas dos usuários e às normas regulatórias.

Também é importante destacar que, a qualquer tempo, por meio do programa de Supervisão de Mercado, a Anatel pode avaliar a conformidade do produto e seu fornecedor com a declaração inserida no requerimento de homologação. Além disso, quaisquer falhas de segurança cibernética identificadas em equipamentos homologados que afetem a segurança de seus usuários, prestadoras ou das redes de telecomunicações do país podem ser objeto de avaliação pela Anatel, ainda que a característica afetada não tenha sido objeto da declaração que compõe o requerimento de homologação.

Assim, identificada uma falha ou vulnerabilidade, a Anatel notificará o responsável para saná-la dentro de um prazo adequado, considerando a severidade conforme o *Common Vulnerability Scoring System* (CVSS). Este prazo pode ser prorrogado pela Anatel. Se não corrigido no prazo estabelecido, a homologação pode ser suspensa e o produto recolhido ou substituído. A suspensão permanece até que a vulnerabilidade seja superada ou o risco mitigado. Após o prazo máximo, a homologação será cancelada caso a falha não for resolvida.

C. Requisitos de segurança cibernética do Ato 77

Os requisitos do Ato 77 são organizados em várias seções que cobrem aspectos essenciais da segurança cibernética para equipamentos de telecomunicações. Primeiramente, é avaliado se o dispositivo foi desenvolvido com base no princípio de *security by design*. Em seguida, são delineadas diretrizes para atualização segura de software e firmware (4 requisitos), gerenciamento remoto (2 requisitos), instalação e operação dos dispositivos (6 requisitos), acesso à configuração dos equipamentos (9 requisitos), serviços de comunicação de dados (4 requisitos), proteção de dados pessoais e sensíveis (4 requisitos), e capacidade de mitigar ataques (3 requisitos). Para cada requisito, deve-se apresentar justificativas específicas caso não seja atendido.

Esta pesquisa ficará restrita a seção de capacidade de mitigar ataques. Para obter conformidade com este seção, o equipamento deve ter conformidade três requisitos, são eles:

- a) Possuir mecanismo para limitação da taxa de transmissão de dados de saída (upload), além do usualmente necessário, a fim de minimizar sua utilização como vetor em ataques a outros equipamentos ou sistemas (ataque de negação de serviço).

- b) Implementar mecanismos para validação do endereço de origem dos pacotes de dados, filtrando pacotes com endereço de origem falsificados (filtro antispoofing), em especial na transmissão de dados de saída (upload).
- c) Ser projetado para mitigar os efeitos de ataques de negação de serviço em andamento, sendo resistentes a um número excessivo de tentativas de autenticação, por meio de, por exemplo: priorização de sua capacidade de processamento às sessões de comunicação já estabelecidas e autenticadas; e limitação do número de sessões de autenticação concorrentes, descartando tentativas de estabelecimento de novas sessões quando superado limite estabelecido.

III. METODOLOGIA

A metodologia empregada neste trabalho visa fornecer uma estratégia para a verificação e recomendação para obter adequação de equipamentos terminais com conexão à Internet e de infraestrutura de redes de telecomunicações com o Ato 77, garantindo que esses dispositivos atendam aos requisitos de segurança estabelecidos pela regulamentação previsto para área de capacidade de mitigar ataques. O processo metodológico é dividido nas seguintes etapas:

- 1) Levantamento dos Requisitos do Ato 77 da Anatel: foi realizado um levantamento detalhado dos requisitos específicos estabelecidos pelo Ato 77. Esta etapa envolveu uma análise aprofundada do texto da regulamentação, bem como a identificação e compreensão dos padrões de segurança e privacidade exigidos para os equipamentos;
- 2) Análise dos Equipamentos de Telecomunicações Existentes: consistiu na análise detalhada das funcionalidades e configurações dos equipamentos selecionados a partir de consulta no mercado. A seleção incluiu uma variedade de modelos, principalmente aqueles utilizados por usuários com menor capacidade técnica para implementar estratégias voltadas a segurança da informação;
- 3) Identificação de Lacunas de Conformidade: com base nos requisitos estabelecidos pelo Ato 77 e na análise dos equipamentos existentes, identificaram-se as lacunas de conformidade entre os dispositivos e os padrões regulatórios. Esta fase permitiu uma compreensão clara das áreas que necessitam de melhorias para atender às exigências de segurança estabelecidas;
- 4) Desenvolvimento de Recomendações: a partir da identificação das lacunas de conformidade, foram desenvolvidas recomendações específicas para aprimorar os dispositivos de modo a garantir sua conformidade com o Ato 77. Essas recomendações, conforme descrito no início da seção, ficaram restritas a capacidade de mitigar ataques.

IV. RESULTADOS

A partir da avaliação dos dispositivos, os resultados foram sumarizados na Tabela I. É destacada a capacidade de mitigar

ataques de cada equipamento, evidenciando os mecanismos de defesa implementados, a limitação de taxa de upload, presença de filtro anti-spoofing e a mitigação de ataques de negação de serviço.

Dentre os oito equipamentos avaliados, apenas três deles tem total conformidade com a capacidade de mitigar ataques, o ASUS ZenWiFi XD6, ASUS ZenWiFi AX (XT8) e TP-Link Archer AX6000. Outros quatro equipamentos, TP-Link Archer C6, TP-Link Archer C21, Intelbras RG 1200 e D-Link DIR-846 AC1200, obtiveram uma avaliação parcial para a mitigação de ataques DoS. O equipamento APsystems ECU-R não provê conformidade com nenhum dos requisitos.

Roteadores como o TP-Link Archer C21 são amplamente adotados e facilmente encontrados no mercado brasileiro devido à sua popularidade e disponibilidade generalizada. Por outro lado, modelos como o ASUS ZenWiFi XD6, TP-Link Archer AX6000 e ASUS ZenWiFi AX (XT8) destacam-se como exemplos de dispositivos focados especificamente na segurança. Eles oferecem recursos avançados, como firewall integrado, proteção contra ataques DDoS e criptografia robusta, atendendo às necessidades daqueles que priorizam a proteção de suas redes.

Já roteadores como o Intelbras RG 1200, D-Link DIR-846 AC1200 e TP-Link Archer C6 representam opções mais intermediárias. Eles demonstram a capacidade de equilibrar a segurança com um preço mais acessível, oferecendo funcionalidades básicas de segurança, como firewall SPI, controle parental e criptografia WPA/WPA2. Esses dispositivos são adequados para usuários que buscam uma proteção sólida sem comprometer significativamente o orçamento.

Os resultados demonstram que os roteadores ASUS ZenWiFi XD6, TP-Link Archer AX6000 e ASUS ZenWiFi AX (XT8) atendem plenamente aos três requisitos de defesa. Isso significa que esses dispositivos possuem recursos robustos para limitar a taxa de upload, implementar filtros anti-spoofing e mitigar ataques de negação de serviço, proporcionando uma camada sólida de proteção para as redes domésticas ou empresariais.

Por outro lado, alguns dispositivos, como TP-Link Archer C21 e APsystems ECU-R, demonstram adequações parciais ou nulas em relação aos mecanismos de defesa avaliados. Embora alguns desses dispositivos possam oferecer alguma forma de mitigação de ataques, eles podem não ser tão eficazes quanto os dispositivos que atendem plenamente aos requisitos.

A presença de mecanismos de defesa eficazes, conforme destacado nesta tabela, desempenha um papel crucial na proteção das redes contra ameaças cibernéticas. Portanto, ao selecionar um dispositivo de rede, é essencial considerar não apenas os recursos de conectividade, mas também sua capacidade de fornecer segurança robusta contra possíveis ataques. Este aspecto é fundamental para garantir a integridade e a confiabilidade das redes domésticas ou empresariais no ambiente digital atual.

TABLE I
AVALIAÇÃO DE CONFORMIDADE COM OS REQUISITOS DE CAPACIDADE DE MITIGAR ATAQUES.

Requisito	Capacidade de Mitigar Ataques			
Equipamento	Adequação	Mecanismos de Defesa		
		Limitação de taxa de upload	Filtro anti-spoofing	Mitigação de ataques DoS
APsystems ECU-R	Não	Não	Não	Não
ASUS ZenWiFi XD6	Sim	Sim	Sim	Sim
TP-Link Archer AX6000	Sim	Sim	Sim	Sim
ASUS ZenWiFi AX (XT8)	Sim	Sim	Sim	Sim
TP-Link Archer C6	Sim	Sim	Sim	Parcial
TP-Link Archer C21	Parcial	Não	Sim	Parcial
Intelbras RG 1200	Sim	Sim	Sim	Parcial
D-Link DIR-846 AC1200	Sim	Sim	Sim	Parcial

V. CONSIDERAÇÕES FINAIS

Este artigo avaliou a conformidade de equipamentos terminais com conexão à Internet e infraestrutura de redes de telecomunicações com os requisitos do Ato 77 da Anatel. A análise incluiu critérios relacionados a capacidade de mitigação de ataques. Os resultados mostraram variações nos níveis de conformidade dos dispositivos avaliados, destacando a necessidade de aprimoramentos.

Nesse contexto, as recomendações buscam assegurar que os equipamentos atendam plenamente às exigências regulatórias, contribuindo substancialmente para a robustez da segurança cibernética no Brasil. É fundamental que a conformidade com o Ato 77 seja percebida não apenas como uma obrigação regulatória, mas como um componente essencial na proteção das infraestruturas críticas e na segurança dos usuários. Desse modo, a implementação eficiente destas diretrizes reduz vulnerabilidades e fortalece a resiliência contra ameaças cibernéticas, as quais seguem em constante evolução.

Logo, a colaboração entre reguladores, fabricantes e a comunidade de científica deve ser considerada vital para enfrentar os desafios emergentes na segurança cibernética. Adaptações contínuas às mudanças no cenário global de ameaças são essenciais para manter a integridade e a confiabilidade das redes de telecomunicações.

Os trabalhos futuros almejam realizar testes experimentais para comprovar a análise feita a partir da documentação dos equipamentos, assim como ampliar as áreas avaliadas, buscando contemplar todos os requisitos previstos no Ato 77.

AGRADECIMENTOS

O presente trabalho foi realizado com o suporte da Padtec S.A., por meio de projeto de inovação.

REFERENCES

- [1] Agência Nacional de Telecomunicações (Anatel), “Ato Nº 77, de 5 de janeiro de 2021: Estabelece requisitos de segurança cibernética para equipamentos de telecomunicações,” 2021, Publicado no Diário Oficial da União em 6 de janeiro de 2021.
- [2] Presidência da República, “Decreto Nº 10.222, de 5 de fevereiro de 2020: Aprova a Estratégia Nacional de Segurança Cibernética,” 2020, Gabinete de Segurança Institucional da Presidência da República.
- [3] European Parliament and Council, “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation),” 2016, Official Journal of the European Union, L 119, 4 May 2016.
- [4] Presidência da República, “Lei Nº 9.472, de 16 de julho de 1997: Dispõe sobre a organização dos serviços de telecomunicações, a criação e o funcionamento de um órgão regulador e outros aspectos institucionais,” 1997, Publicado no Diário Oficial da União em 17 de julho de 1997.
- [5] Agência Nacional de Telecomunicações (Anatel), “Regulamento para Avaliação da Conformidade e Homologação de Produtos para Telecomunicações,” 2021, Publicado no Diário Oficial da União em 13 de dezembro de 2019.
- [6] Presidência da República, “Lei Nº 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais,” 2018, Publicado no Diário Oficial da União em 15 de agosto de 2018.
- [7] OWASP Foundation, “Security by Design Principles,” 2016, Disponível em: https://wiki.owasp.org/index.php/Security_by_Design_Principles. Acesso em: 31 jul 24.

Estudo Experimental de Técnicas de Exploração de Canais Colaterais

Acsa Laiane*, Artur Braga*, Diego Pereira[†], Victor Pimentel[†], Ramon Fontes*, Augusto Neto* and Roger Immich*

*Universidade Federal do Rio Grande do Norte, Natal, RN

[†]Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, Parnamirim, RN

Abstract—Ataques de canais colaterais representam uma ameaça crescente à segurança criptográfica, explorando vulnerabilidades físicas sem necessidade de alterações nos sistemas ou equipamentos. Este artigo apresenta um estudo experimental focado em técnicas de exploração de canais colaterais, alinhado com as diretrizes estabelecidas pela norma ISO/IEC 17825, que aborda métodos de teste para mitigar ataques não invasivos em módulos criptográficos. A pesquisa detalha a implementação e os resultados de uma série de experimentos práticos destinados a explorar e avaliar a eficácia de diferentes técnicas de análise de canais colaterais, como análise de consumo de potência e análise eletromagnética. Para cada técnica foi realizada uma prova de conceito em ambiente controlado para identificar vulnerabilidades e avaliar a eficácia de possíveis contramedidas. O estudo não apenas demonstra como essas técnicas podem ser utilizadas para comprometer a segurança de sistemas criptográficos, mas também como podem ser efetivamente mitigadas através das práticas de segurança sugeridas. Os resultados demonstraram que a eficácia dos ataques está fortemente associada a qualidade dos traços capturados a partir da observação do canal colateral. Além disso, nos dois cenários propostos foi possível obter parte das chaves criptográficas.

Index Terms—Módulos criptográficos, Segurança Não-Invasiva, Ataques de canal colateral, Mitigação de ataques, ISO/IEC 17825

I. INTRODUÇÃO

A segurança dos sistemas criptográficos enfrenta desafios crescentes à medida que as ameaças cibernéticas se tornam cada vez mais sofisticadas. Neste cenário, os ataques não invasivos, também conhecidos como ataques de canais colaterais (*Side-Channel Attacks*), emergem como uma ameaça significativa devido à sua capacidade de explorar vulnerabilidades sem alterações físicas nos sistemas ou equipamentos. Esses ataques não derivam de falhas nos algoritmos ou módulos criptográficos propriamente ditos, mas aproveitam-se das características físicas resultantes do comportamento operacional dos sistemas.

Conforme definido pela norma ISO/IEC 17825, esses ataques exploram informações coletadas através de medições físicas não intrusivas dos módulos criptográficos, destacando a necessidade urgente de entender e mitigar tais riscos. De acordo com Vidakovic (2023) [1], os ataques não

invasivos são geralmente mais acessíveis, baseando-se na observação ou manipulação de dispositivos sem causar danos físicos. Tais ataques requerem apenas equipamentos moderadamente sofisticados e permitem que invasores monitorem características analógicas como conexões de energia e emissões eletromagnéticas, acessando informações internas sem violações físicas ou evidências de uso não autorizado.

Nesse contexto, este artigo examina a segurança não invasiva sob a perspectiva da norma ISO/IEC 17825, intitulada *Testing methods for the mitigation of non-invasive attack classes against cryptographic modules*. Apresentamos uma análise detalhada de como vulnerabilidades em canais colaterais podem ser exploradas para comprometer a integridade de dispositivos, incluindo demonstrações práticas de análises de potência, eletromagnéticas e temporais. A principal contribuição deste estudo é avançar o estado da arte, fornecendo a pesquisadores da área de segurança não invasiva um conjunto de estratégias, ferramentas e diferentes perspectivas.

II. EXPLORAÇÃO DE CANAIS COLATERAIS

Sob o conceito dos canais colaterais, tratamos da exploração de vulnerabilidades que não estão relacionadas a falhas na implementação lógica de um sistema criptográfico, por exemplo. Neste sentido, a origem das vulnerabilidades não advém de problemas no *software* ou algoritmos de criptografia em si. Em vez disso, considera-se o funcionamento global do módulo, desde a construção do algoritmo até padrões de comportamento das rotinas executadas [2].

A ISO 17825 esclarece que ataques não invasivos exploram vieses latentes nas quantidades físicas medidas de forma não invasiva sobre ou ao redor do módulo [3]. Esses vieses são induzidos e dependem da informação secreta que é alvo do ataque. Nessa perspectiva, análises sobre consumo de energia, emissões de ondas eletromagnéticas e até sobre a emissão de sons podem ser realizadas para extrair informações visando ações maliciosas.

Com isso, observamos uma ampla variedade de ataques por canais colaterais, que podem ou não ser do tipo de caixa preta. Os ataques de canal colateral permitem que o invasor monitore as características analógicas das conexões de alimentação e interface, além de qualquer radiação eletromagnética do dispositivo durante a operação normal do módulo, sem interação direta com o módulo atacado.

¹Instituto Metrópole Digital. Universidade Federal do Rio Grande do Norte, Natal, RN.

²Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, Parnamirim, RN. Email: diego.pereira@ifrn.edu.br

A. Classificação dos canais colaterais

Os canais colaterais microarquiteturais podem ser categorizados em dois tipos principais: **persistentes** e **efêmeros** [4]. Essa classificação é complementada por descrições adicionais que distinguem entre ataques **ativos** e **passivos**, ampliando o entendimento das diversas maneiras pelas quais esses canais podem ser explorados para comprometer a segurança de sistemas computacionais. Uma descrição dos canais é feita a seguir:

- **canais colaterais persistentes:** o atacante observa informações que persistem por um período prolongado após uma operação legítima do sistema. A informação vazada é mantida até que outro evento no sistema modifique esse estado. Ataques que exploram canais colaterais persistentes incluem técnicas como *Prime+Probe* e *Flush+Reload*, que observam e manipulam estados mantidos, como o conteúdo de memórias cache.
- **canais colaterais efêmeros:** estes canais se baseiam em vazamentos de informações que são temporários e desaparecem rapidamente após a execução da operação. Para capturar essas informações efêmeras, um atacante precisa agir simultaneamente ou em uma janela de tempo muito próxima à atividade da vítima. Exemplos incluem ataques que observam flutuações na temporização ou consumo de energia que desaparecem assim que a tarefa específica é concluída.
- **ataques de canal colateral ativos:** esses atacantes não apenas observam, mas também manipulam ativamente o sistema para criar condições favoráveis para a coleta de dados. Eles podem, por exemplo, injetar suas próprias operações para induzir um determinado estado do sistema que facilita o vazamento de informações.
- **ataques de canal colateral passivos:** limitam-se a observar os efeitos das operações normais da vítima sem intervir diretamente no funcionamento do sistema. Ataques passivos geralmente medem variações como o tempo de execução ou padrões de consumo de energia sem perturbar o estado operacional do sistema.

A exploração dos canais colaterais revela a complexidade e a variedade de ameaças que podem afetar a segurança dos sistemas. Percebe-se que os canais persistentes fornecem ao atacante a capacidade de observar mudanças duradouras que podem ser exploradas até serem alteradas por novos eventos, enquanto os canais efêmeros apresentam desafios e oportunidades devido à sua natureza transitória.

Além disso, a diferenciação entre ataques ativos, que modificam o estado do canal para facilitar a exploração, e passivos, que se limitam à observação, sublinha a necessidade de estratégias de segurança robustas que considerem tanto a detecção como a prevenção de tais explorações. Este entendimento aprofundado é essencial para o desenvolvimento de sistemas mais seguros e para a formulação de respostas eficazes às diversas ameaças cibernéticas. Nas seções seguintes, aprofundaremos nosso estudo ao examinar

especificamente os ataques por análise de potência e emissões eletromagnéticas.

B. Visão Geral da norma ISO/IEC 17825

O padrão internacional ISO/IEC 17825 especifica métricas de teste para a mitigação de ataques não invasivos [3]. Essas métricas são definidas em conformidade com requisitos de segurança para módulos criptográficos estabelecidos por mais duas normas: a ISO/IEC 19790 [5] e a ISO/IEC 24759 [6]. Dessa forma, o principal objetivo da ISO 17825 é definir métodos de teste que garantam que um módulo criptográfico implementa técnicas de mitigação e resistência a ataques não invasivos. São destacados um conjunto de ataques, divididos em análises de três tipos de canal colateral: análises de consumo de potência, análises eletromagnéticas e análises temporais [3]. Detalhes sobre análises de consumo de potência e análises eletromagnéticas são apresentados nas seções seguintes.

C. Análise de Consumo de Potência

A Análise de Potência, como um tipo de ataque de canal colateral, explora a variação no consumo de energia de dispositivos eletrônicos, particularmente durante a execução de operações criptográficas. Diferente dos ataques que exploram falhas lógicas ou de implementação, os ataques de análise de potência baseiam-se nas características físicas e operacionais do *hardware*. Segundo a norma ISO 17825, esses ataques não invasivos utilizam variações mensuráveis e não invasivas em grandezas físicas, como o consumo de energia, para extrair informações sensíveis sem alterar fisicamente o dispositivo, o que torna esses ataques ainda mais complexos quanto a sua mitigação [3].

Ataques de análise de potência tem a capacidade de revelar informações secretas analisando a variação do consumo de energia a partir diferentes entradas de dados ou operações executadas pelo dispositivo [7]. Foi demonstrado que, ao monitorar o consumo de energia de dispositivos enquanto eles executam algoritmos criptográficos, é possível inferir chaves secretas com uma quantidade surpreendentemente pequena de dados. Esta descoberta fundamentou a área de pesquisa sobre ataques de canal colateral e estimulou o desenvolvimento de uma diversidade de técnicas de mitigação para proteger dispositivos contra tais vulnerabilidades.

Esse campo evoluiu significativamente nas últimas décadas, expandindo-se desde a Análise de Potência Simples (*Simple Power Analysis* - SPA) até técnicas mais sofisticadas como a Análise de Potência Diferencial (*Differential Power Analysis* - DPA) e a Análise de Potência por Correlação (*Correlation Power Analysis* - CPA) [8]. Esses métodos não apenas exploram as propriedades físicas do *hardware*, mas também se aprofundam nas flutuações mínimas de energia para extrair informações sensíveis que idealmente deveriam estar seguras.

Em geral, os ataques SPA, DPA e CPA se diferenciam, principalmente, pela maneira de como os dados são coletados e analisados. Os dados coletados são armazenados em arquivos chamados de traços de energia. Os traços são registros

detalhados das variações de consumo de energia de um dispositivo ao longo do tempo, especialmente durante a execução de operações criptográficas. Eles são fundamentais para a realização do ataque.

Para capturar traços de energia, geralmente, são usados osciloscópios ou outros dispositivos de medição de alta precisão conectados ao dispositivo alvo. Estes equipamentos medem as flutuações de corrente ou tensão que ocorrem quando o dispositivo executa determinadas operações. Desse modo, ao examinar como o consumo de energia altera com diferentes entradas para processos criptográficos, os analistas podem inferir informações sobre os processos internos do dispositivo, incluindo a extração de chaves criptográficas. A seguir é feito um aprofundamento nos três tipos de ataques de análise de potência.

- **Análise de Potência Simples:** SPA envolve observar diretamente o traço do consumo de energia de um dispositivo para identificar padrões ou operações específicas. Por exemplo, um atacante pode ser capaz de deduzir a chave de criptografia de um algoritmo observando os picos de consumo de energia correspondentes a certas operações de chave.
- **Análise de Potência Diferencial:** DPA é mais sofisticada e envolve a coleta de um grande número de traçados de energia enquanto o dispositivo executa a mesma operação (como criptografar dados) com diferentes entradas. O atacante então usa análise estatística para correlacionar as variações no consumo de energia com os dados específicos sendo processados, permitindo a extração de informações secretas, como chaves criptográficas.
- **Análise de Potência de Correlação:** uma variante do DPA, a CPA usa técnicas estatísticas para correlacionar o consumo de energia observado com as previsões teóricas baseadas em modelos de consumo de energia do dispositivo. Isso pode permitir que ataques sejam mais eficientes e eficazes, mesmo com menor quantidade de dados.

D. Análises Eletromagnéticas

De forma geral, análises eletromagnéticas consistem em uma técnica de canal colateral que examina as variações nos campos eletromagnéticos emitidos por um dispositivo durante a execução de operações criptográficas [1]. Desse modo, utilizando sondas específicas, um atacante faz aferições desses campos e analisa as variações para identificar padrões que revelem dados sensíveis, por exemplo, chaves criptográficas. Assim como as análises de potência, essa técnica pode comprometer a segurança do sistema, permitindo a extração de informações protegidas.

É importante destacar que estudos demonstram a viabilidade de ataques de análise eletromagnética (*ElectroMagnetic Analysis* - EMA) contra o algoritmo *Advanced Encryption Standard* (AES) executado em celulares [9]. Os resultados mostraram que é possível extrair chaves criptográficas de dispositivos móveis utilizando apenas as emissões eletromagnéticas geradas durante a execução do AES.

Os ataques EMA são técnicas de canal colateral que exploram as variações nos campos eletromagnéticos emitidos por dispositivos eletrônicos durante a execução de operações criptográficas. Esses ataques são classificados principalmente em dois tipos: SEMA (*Simple Electromagnetic Analysis*) e DEMA (*Differential Electromagnetic Analysis*). Os principais pontos de destaque para esse tipo de ataque são:

- exploram emissões eletromagnéticas de dispositivos criptográficos;
- os vazamentos incluem informações sobre operações internas, como execução de algoritmos de criptografia;
- ataques dessa natureza podem ser realizados de forma local, utilizando sondas próximas ao dispositivo alvo;
- as variações nos campos eletromagnéticos durante operações criptográficas são uma fonte de vazamento de informações;
- analisando padrões eletromagnéticos, um laboratório de testes pode recuperar chaves secretas ou outros dados sensíveis.

Nesse contexto, SEMA trata-se de uma técnica de análise direta dos padrões eletromagnéticos emitidos por um dispositivo. Um exemplo típico de uso do SEMA seria medir o campo eletromagnético de um dispositivo enquanto ele executa um algoritmo de criptografia, como AES. Observando mudanças significativas nos padrões do campo eletromagnético, um atacante pode inferir informações sobre as chaves usadas, comprometendo a segurança do dispositivo. As principais características são:

- utiliza observações diretas, sem a necessidade de métodos estatísticos complexos;
- foca em picos ou mudanças evidentes nos sinais eletromagnéticos durante operações criptográficas;
- funciona melhor em dispositivos com pouca ou nenhuma proteção contra ataques de canal colateral.

O DEMA é uma técnica mais sofisticada, ela faz uso de métodos estatísticos para analisar variações mínimas nos campos eletromagnéticos emitidos por um dispositivo. O objetivo principal é identificar informações criptográficas sensíveis através da análise diferencial de várias amostras de medições eletromagnéticas. Um exemplo de DEMA seria coletar várias medições eletromagnéticas enquanto um dispositivo executa operações criptográficas com diferentes dados de entrada. Analisando as diferenças entre essas aferições, um atacante pode detectar correlações que revelam informações sobre as chaves criptográficas, comprometendo a segurança do dispositivo. As principais características são:

- utiliza a diferença entre múltiplas medições para identificar correlações e padrões sutis;
- é capaz de extrair informações detalhadas mesmo em dispositivos com proteção básica;
- requer um processamento de dados mais complexo e equipamento mais avançado.

III. METODOLOGIA

A partir dos objetivos do trabalho, nesta seção serão apresentadas propostas de cenários e ferramentas para

construção de experimentos na temática de segurança não invasiva. Desse modo, devido a restrição de páginas do artigo, serão investigadas apenas análises de consumo de potência e análises eletromagnéticas. Para execução dos ataques de análise de consumo de potência e análise eletromagnética, foram utilizadas as ferramentas Jlsca [10], ASCAD [11] e ChipWhisperer [12].

IV. RESULTADOS

A. Resultados dos Ataques de Análise de Potência

Os ataques de análise de potência foram realizados utilizando a ferramenta Jlsca para analisar os traços de energia de um dispositivo durante a execução de operações criptográficas. Os traços de energia utilizados neste trabalho foram obtidos a partir do *dataset* ASCAD.

1) *Resultados Positivos*: Os resultados mostraram que foi possível recuperar informações sensíveis, como a chave criptográfica, com uma correlação significativa entre os traços de energia capturados e os modelos teóricos. As fases do ataque permitiram identificar as partes da chave com alta precisão, conforme ilustrado nas Figuras 1 e 2. É possível observar que na Fase 6 (Fig. 2), o valor correto estava em uma posição mais baixa no ranking, sugerindo que ajustes nos parâmetros de análise são necessários.

```
Jlsca running in Julia version: 1.4.1, 1 processes/1 workers/1 threads per worker

DPA parameters
attack: Sha1 input
T xor: false
analysis: Incremental CPA
leakages: HW.bit0.bit1.bit2.bit3.bit4.bit5.bit6.bit7
maximization: abs global max
combination: +
data at: 1
phases: [1, 2, 3, 4]
known key: 685d79d6007d6cb9437d9393c6eacdefb69a0a

phase: 1 / 14, #targets 1

Attacking columns 1:5 out of 5 columns (run 1 out of 1)
Running "Incremental correlation" on trace range 1:1:500, 1 data passes, 2 sample passes
Incremental CPA on range 1:5 produced (5, 2304) correlation matrix
Results @ 500 rows, 5 cols (500 rows consumed)
target: 1, phase: 1, #candidates 256
rank: 1, candidate: 0xaf, score: 0.914699 @ 4
rank: 2, candidate: 0x2f, score: 0.910536 @ 4
rank: 3, candidate: 0xb7, score: 0.863415 @ 2
rank: 4, candidate: 0xb1, score: 0.857900 @ 1
rank: 8, correct: 0x47, score: 0.845734 @ 4
recovered key material: af
```

Fig. 1. Resultado da Fase 1.

```
phase: 6 / 14, #targets 1
phase input: 47cfa7307e

Attacking columns 1:5 out of 5 columns (run 1 out of 1)
Running "Incremental correlation" on trace range 1:1:500, 1 data passes, 2 sample passes
Incremental CPA on range 1:5 produced (5, 2304) correlation matrix
Results @ 500 rows, 5 cols (500 rows consumed)
target: 1, phase: 6, #candidates 256
rank: 1, candidate: 0x4a, score: 0.917065 @ 1
rank: 2, candidate: 0x16, score: 0.897856 @ 2
rank: 3, candidate: 0x96, score: 0.884578 @ 2
rank: 4, candidate: 0x14, score: 0.867130 @ 2
rank: 101, correct: 0x01, score: 0.744942 @ 2
recovered key material: 4a
```

Fig. 2. Resultado da Fase 6.

Ao final do processo de análise foi obtido ranqueamento para todas as fases do ataque. Para as fases no qual o resultado final da correlação foi ranqueado mais próximo da posição 1,

indica maior probabilidade da estimativa de parte da chave criptográfica estar correta. Os resultados para cada fase estão dispostos na Fig. 3.

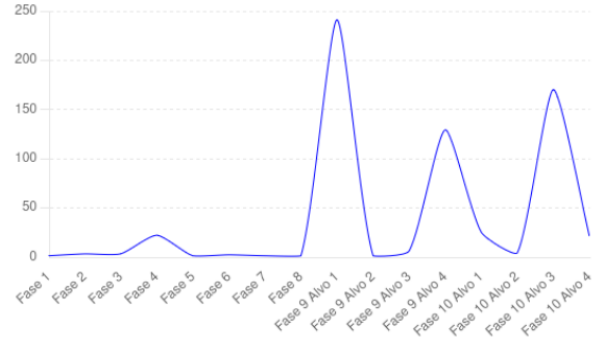


Fig. 3. Análise CPA com boa correlação.

2) *Resultados Negativos*: Em alguns casos, a correlação obtida foi baixa, indicando dificuldades em distinguir a chave correta de outras hipóteses. Isso pode ser atribuído a ruídos nos traços capturados ou limitações intrínsecas ao algoritmo de ataque utilizado. Nesses cenários, a recuperação de informações sensíveis, como a chave privada, não foi totalmente bem-sucedida, sugerindo a necessidade de ajustes nos parâmetros de análise e melhorias na qualidade dos dados capturados. Exemplo de análise mal sucedida é apresentada na Fig. 4.

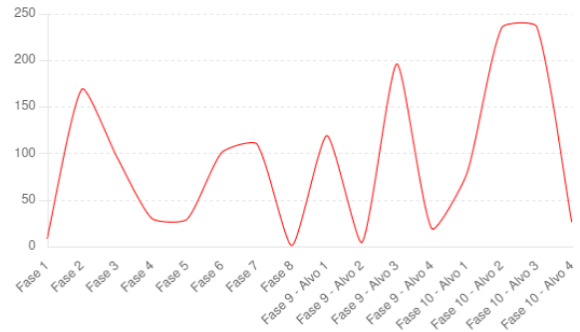


Fig. 4. Análise CPA com correlação insatisfatória.

B. Resultados dos Ataques de Análise Eletromagnética

Os ataques de análise eletromagnética foram conduzidos utilizando a ferramenta ChipWhisperer, analisando as emissões eletromagnéticas durante a execução de algoritmos criptográficos a partir de arquivos de traço disponíveis na ASCAD.

1) *Resultados Positivos*: Os ataques de DEMA conseguiram identificar a chave criptográfica correta em várias tentativas. A correlação máxima observada nas análises estatísticas mostrou que as emissões capturadas continham informações suficientes para deduzir partes críticas das chaves criptográficas.

2) *Resultados Negativos:* Apesar das tentativas bem-sucedidas, houve situações em que a correlação máxima obtida foi baixa, limitando a eficácia do ataque. A baixa correlação sugere que a presença de ruídos e a complexidade do ambiente de medição podem interferir na precisão dos resultados. Nesses casos, não foi possível obter informações sensíveis de forma consistente, destacando a importância de um ambiente de medição controlado e de técnicas avançadas de filtragem de ruídos.

A Fig. 5 ilustra os resultados obtidos por meio do ataque DEMA, mais especificamente uma análise de correlação de matriz de deslocamento. Este método foi utilizado na tentativa de identificar a chave criptográfica em uso.

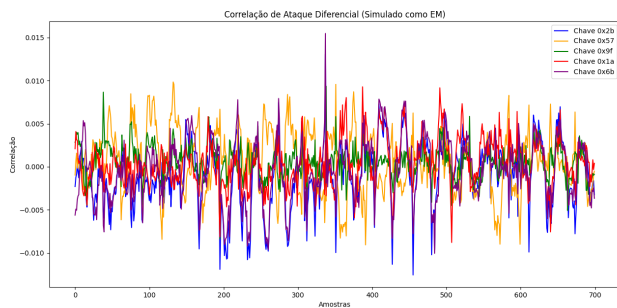


Fig. 5. Resultado da correlação do DEMA

C. Discussão dos Resultados

Os resultados dos ataques de análise de potência e eletromagnética demonstram a viabilidade dessas técnicas em comprometer a segurança de dispositivos criptográficos. No entanto, a eficácia dos ataques depende fortemente da qualidade dos traços capturados e da configuração do ambiente de medição. A baixa correlação observada em alguns experimentos ressalta a necessidade de aprimorar as técnicas de captura e análise para superar as limitações impostas por ruídos e variabilidades no ambiente.

As estratégias de mitigação, como mascaramento de dados, equilíbrio do consumo de energia, e blindagem eletromagnética, são essenciais para proteger contra essas ameaças. Implementar essas contramedidas pode reduzir significativamente o risco de ataques bem-sucedidos, reforçando a segurança dos sistemas criptográficos.

V. CONSIDERAÇÕES FINAIS

À medida em que surgem novas ameaças cibernéticas, a segurança de módulos criptográficos se vê diante da necessidade de adotar mecanismos cada vez mais robustos de prevenção e mitigação a ataques. Neste contexto, os ataques de canais colaterais se apresentam como um desafio significativo, aproveitando-se da exploração de comportamentos operacionais de sistemas para comprometer sua segurança sem a necessidade de alterações físicas nos dispositivos.

O presente trabalho apresentou uma contextualização dos ataques de canal colateral, categorizados como ataques não invasivos. O estudo, que teve como base o padrão internacional ISO 17825, propôs-se a entender os diferentes tipos de canal colateral e a forma como são empregados para atingir a segurança de dispositivos criptográficos. Apresentamos estudos realizados sobre análises de consumo de potência e eletromagnética, o que envolveu a realização de demonstrações práticas de como se dão ataques de ambas as naturezas e de que formas a mitigação a eles pode ser abordada.

Os resultados obtidos com os ataques realizados demonstram a viabilidade dessas técnicas em comprometer a segurança de módulos criptográficos. No entanto, observamos que a eficácia desses ataques depende fortemente da qualidade dos traços capturados e das configurações do ambiente de medição. A baixa correlação observada em alguns experimentos ressalta a necessidade de aprimorar as técnicas de captura e análise para superar as limitações impostas por ruídos e variabilidades no ambiente.

AGRADECIMENTOS

O presente trabalho foi realizado com o suporte da Padtec S.A., por meio de projeto de inovação.

REFERENCES

- [1] M. Vidaković and D. Vinko, "Hardware-based methods for electronic device protection against invasive and non-invasive attacks," *Electronics*, vol. 12, no. 21, p. 4507, 2023.
- [2] TEHRANIPOOR, Mohammad; WANG, Cliff, *Introduction to Hardware Security and Trust*. University Of Connecticut: Springer, 2011.
- [3] "Information technology — Security techniques — Testing methods for the mitigation of non-invasive attack classes against cryptographic modules," International Organization for Standardization, Standard, Jan. 2024.
- [4] W. T. N. Peter Deutsch and M. Yan, "Metior: A comprehensive model to evaluate obfuscating side-channel defense schemes," *ISCA '23*, 2023.
- [5] "Information technology — Security techniques — Security requirements for cryptographic modules," International Organization for Standardization, Standard, Dec. 2020.
- [6] "Information technology — Security techniques — Test requirements for cryptographic modules," International Organization for Standardization, Standard, Jun. 2017.
- [7] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Advances in Cryptology—CRYPTO'99: 19th Annual International Cryptology Conference Santa Barbara, California, USA, August 15–19, 1999 Proceedings 19*. Springer, 1999, pp. 388–397.
- [8] M. Randolph and W. Diehl, "Power side-channel attack analysis: A review of 20 years of study for the layman," *Cryptography*, vol. 4, no. 2, p. 15, 2020.
- [9] D. Aboukassimi, M. Agoyan, L. Freund, J. Fournier, B. Robisson, and A. Tria, "Electromagnetic analysis (ema) of software aes on java mobile phones," in *2011 IEEE International Workshop on Information Forensics and Security*. IEEE, 2011, pp. 1–6.
- [10] C.-B. Breunese and I. Kizhvatov, "Jlsc. side-channel toolkit in julia," *Side-Channel Toolkit in Julia*, 2019.
- [11] R. Benadjila, E. Prouff, R. Strullu, E. Cagli, and C. Dumas, "Deep learning for side-channel analysis and introduction to ascad database," *Journal of Cryptographic Engineering*, vol. 10, no. 2, pp. 163–188, 2020.
- [12] C. O'flynn and Z. Chen, "Chipwhisperer: An open-source platform for hardware embedded security research," in *Constructive Side-Channel Analysis and Secure Design: 5th International Workshop, COSADE 2014, Paris, France, April 13–15, 2014. Revised Selected Papers 5*. Springer, 2014, pp. 243–260.

Uma Estratégia para Mitigação de Ataques DoS em Gateways IoT

Guilherme Aurélio Ribeiro Rocha

Laboratório de Redes e Sistemas Computacionais (NOCS)

Instituto Federal do Rio Grande do Norte (IFRN)

Parnamirim, Brasil

guilherme.aurelio@escolar.ifrn.edu.br

Diego da Silva Pereira

Laboratório de Redes e Sistemas Computacionais (NOCS)

Instituto Federal do Rio Grande do Norte (IFRN)

Parnamirim, Brasil

diego.pereira@ifrn.edu.br

Abstract—Este artigo aborda a mitigação de ataques de negação de serviço (DoS) em gateways IoT, utilizando o Snort como ferramenta de detecção e prevenção de intrusões. A implementação do Snort busca atender aos requisitos de segurança cibernética do Ato 77 da Anatel. O estudo compara a eficácia do Snort nos modos IDS (Detecção de Intrusão) e IPS (Prevenção de Intrusão) através de resultados experimentais que demonstram sua capacidade em proteger gateways IoT contra ataques DoS. A análise comparativa destaca as vantagens e desvantagens de cada modo, auxiliando na escolha da configuração ideal para diferentes cenários.

Index Terms—Gateway IoT, Ato 77 da Anatel, Segurança Cibernética, Ataques DoS, Snort

I. INTRODUÇÃO

Uma plataforma de computação segura é essencial para prover serviços de gateway IoT com segurança e confiabilidade. É importante que ela seja capaz de lidar com as ameaças à segurança e garantir a integridade dos dados e a confidencialidade das informações.

Os gateways IoT desempenham um papel crucial na segurança dos sistemas. Eles são responsáveis por fornecer conectividade entre os dispositivos IoT e a infraestrutura de rede, além de desempenharem funções como agregação, protocolos de comunicação e gerenciamento de segurança [1]. Portanto, compreender seus requisitos de segurança é fundamental para garantir a proteção e o funcionamento adequado desses sistemas.

Nesse sentido, o Ato 77 da Anatel, publicado em janeiro de 2021, estabelece requisitos de segurança cibernética para equipamentos para telecomunicações, incluindo os gateways IoT. Essa regulamentação visa garantir que as plataformas estejam em conformidade com os padrões de segurança e privacidade estabelecidos, a fim de proteger os usuários e os dados transmitidos pelos dispositivos.

Uma plataforma de computação segura homologada pelo Ato 77 deve atender a todos os requisitos estabelecidos pela regulamentação, incluindo autenticação, criptografia, controle de acesso, gerenciamento de chaves, entre outros. Além disso, ela deve ser capaz de lidar com ameaças à segurança e garantir a resiliência operacional em tempos de incerteza. O uso de uma plataforma homologada pelo Ato 77 pode aumentar a confiança dos usuários na segurança dos sistemas IoT e garantir a proteção dos dados transmitidos pelos dispositivos.

Neste artigo, apresentaremos uma estratégia para a mitigação de ataques de negação de serviço (DoS) em gateways IoT, utilizando o Snort como ferramenta de detecção e prevenção de intrusões (IDS/IPS). O objetivo é demonstrar como a implementação do Snort pode atender a um dos requisitos de segurança do Ato 77, proporcionando uma solução eficaz para proteger os gateways IoT contra ataques DoS. Discutiremos a configuração e os modos de operação do Snort, além de apresentar resultados experimentais que evidenciam a eficácia dessa abordagem. Por fim, avaliaremos os benefícios e desafios dessa solução no contexto de segurança cibernética para dispositivos IoT.

II. TRABALHOS RELACIONADOS

No artigo “A systematic literature review on IoT gateways” [2], é apresentada uma revisão sistemática da literatura sobre gateways IoT, abrangendo artigos publicados entre 2011 e 2021. O estudo categoriza os gateways IoT em básicos e inteligentes, com subdivisões adicionais em gateways passivos, semi-automatizados e totalmente automatizados. A análise abrange critérios como tipos de gateways, requisitos, ferramentas/plataformas, abordagem adotada, avaliação e domínio de aplicação. Identifica lacunas de pesquisa e sugere perspectivas futuras para orientar pesquisadores em investigações avançadas.

Outra importante contribuição é o artigo “Secure IoT Integration in Daily Lives: A Review” [3], que aborda a integração de IoT na vida cotidiana e os desafios de segurança associados. O estudo discute a vulnerabilidade das integrações de IoT a ameaças e ataques, destacando preocupações como autorização, privacidade, controle de acesso, configuração do sistema, verificação, gerenciamento e armazenamento de informações. Este trabalho ressalta a necessidade de melhorar a segurança das implementações de IoT para proteger os dados dos usuários e garantir a integridade e confidencialidade das informações.

Além disso, o artigo “Developing Secure IoT Services: A Security-Oriented Review of IoT Platforms” [4] analisa as plataformas de IoT mais proeminentes do mercado sob a perspectiva da segurança, utilizando cinco critérios comuns detalhados em subcritérios. Este estudo fornece uma comparação aprofundada das características de segurança

das plataformas, auxiliando pesquisadores e desenvolvedores na seleção da plataforma mais adequada para implementar serviços de IoT seguros. O artigo também inclui a implementação de um serviço de monitoramento de voos comerciais em duas das plataformas analisadas, oferecendo diretrizes detalhadas para desenvolvedores de IoT.

III. PLATAFORMA SECURE-IOT

Como plataforma do projeto, foi escolhido o Raspberry Pi, pois era a melhor opção disponível no momento, é versátil e acessível e tem uma capacidade de desempenho e flexibilidade que permite a implementação de funcionalidades variadas, essenciais para a prestação de serviços como Gateway IoT.

O Raspberry Pi é uma plataforma de computação de baixo custo e tamanho compacto que se tornou popular entre entusiastas e desenvolvedores de projetos de IoT. Ele oferece suporte para uma ampla gama de periféricos de entrada e saída, além de possuir conectividade Wi-Fi, o que o torna ideal para interagir com uma variedade de dispositivos e ser utilizado em diversas aplicações de IoT. Sua flexibilidade e versatilidade o tornam uma escolha popular para projetos de automação residencial, monitoramento e controle [2].

Essa plataforma, como qualquer outro computador, também usa um sistema operacional (SO) e o sistema operacional "stock" é uma versão do Linux chamada Raspbian. O Linux é uma ótima combinação para o Raspberry Pi porque é gratuito e de código aberto, mantendo o preço da plataforma baixo, embora possa ser mais suscetível a ataques [2].

Entre as vantagens do Raspberry Pi estão o seu poder de processamento, capacidade de memória, conectividade, ampla disponibilidade de portas USB, sistema operacional Linux, suporte para programação em alto nível e linguagens orientadas a objetos, além de contar com uma comunidade robusta e preços competitivos em relação a outras plataformas.

A. Visão Geral do Ato 77

O Ato 77 da Anatel é um regulamento que estabelece requisitos de segurança cibernética para equipamentos de telecomunicações. Foi publicado em 5 de janeiro de 2021 e entrou em vigor em 1º de julho do mesmo ano. O Ato 77 abrange uma ampla gama de produtos, incluindo equipamentos terminais com conexão à Internet (CPE), como roteadores e modems, e dispositivos de infraestrutura de redes de telecomunicações, como switches, firewalls e gateways. Além dos dispositivos IoT tais como câmeras IP, babás eletrônicas, smart TVs, smart speakers e hubs IoT. O objetivo deste Ato é reduzir ou corrigir vulnerabilidades por meio de atualizações de software/firmware ou recomendações de configurações.

Em resumo, as orientações estabelecem que os fornecedores de equipamentos de telecomunicações são responsáveis por garantir que seus produtos sejam seguros e que atendam aos requisitos de segurança cibernética estabelecidos pela Anatel. A agência pode realizar auditorias para verificar se os produtos estão em conformidade com esses requisitos e pode tomar medidas de sanção caso sejam encontradas falhas ou vulnerabilidades.

Um ponto importante a se destacar é o conceito de security by design, que representa uma abordagem que visa incorporar a segurança desde o início do processo de desenvolvimento de um produto ou serviço. Essa abordagem é importante para garantir que os produtos sejam seguros desde o início e que não sejam necessários ajustes de segurança posteriores.

No contexto do Ato 77 da Anatel, o security by design é importante porque ajuda a garantir que os equipamentos de telecomunicações sejam seguros e que atendam aos requisitos de segurança cibernética estabelecidos pela agência. Ao incorporar a segurança desde o início do processo de desenvolvimento, os fornecedores podem evitar vulnerabilidades que possam ser exploradas por cibercriminosos.

B. Gateways IoT

Um gateway IoT é um dispositivo que atua como ponto de entrada para a rede IoT, permitindo a comunicação entre os dispositivos IoT e a infraestrutura de rede. Ele desempenha funções como agregação de dados, protocolos de comunicação e gerenciamento de segurança. A segurança do gateway IoT é essencial para proteger a integridade dos dados e a confidencialidade das informações transmitidas. [1] afirma que, se esse dispositivo não estiver mais acessível, todos os outros dispositivos conectados a ele também ficarão inacessíveis. Como resultado, os gateways são considerados ativos valiosos em termos de segurança e precisam incorporar medidas de segurança adicionais.

C. Mitigação de Ataques DoS

A mitigação de ataques de negação de serviço (DoS) é uma medida essencial para garantir a segurança e a disponibilidade de gateways IoT. Ataques DoS têm como objetivo esgotar os recursos de um dispositivo ou rede, tornando-os indisponíveis para usuários legítimos. No contexto de dispositivos IoT, que muitas vezes possuem recursos limitados, a mitigação eficaz desses ataques é crítica.

A mitigação de ataques DoS está diretamente relacionada ao Ato 77 da Anatel, que estabelece requisitos de segurança cibernética para equipamentos de telecomunicações, incluindo gateways IoT. Um dos principais requisitos do Ato 77 é a capacidade de mitigar ataques, garantindo a integridade e disponibilidade dos serviços de rede.

1) *Snort*: Para atender a esses requisitos, uma solução eficaz é a implementação do Snort. O Snort é uma ferramenta de código aberto amplamente utilizada para detecção e prevenção de intrusões (IDS/IPS). Ele é capaz de analisar o tráfego de rede em tempo real e identificar padrões associados a atividades maliciosas, como ataques DoS. O Snort pode ser configurado em dois modos principais:

- **IDS (Sistema de Detecção de Intrusão)**: Neste modo, o Snort monitora o tráfego de rede e gera alertas quando detecta atividades suspeitas. Ele não interfere no tráfego, mas fornece informações valiosas que podem ser usadas para tomar ações corretivas.
- **IPS (Sistema de Prevenção de Intrusão)**: No modo IPS, o Snort não apenas detecta, mas também bloqueia

automaticamente os pacotes suspeitos, impedindo que ataques atinjam seus alvos.

Implementar o Snort como ferramenta de detecção e prevenção de intrusões em gateways IoT atende ao requisito de mitigação de ataques estabelecido pelo Ato 77. Além disso, fortalece a postura de segurança do dispositivo, protegendo-o contra tentativas de esgotamento de recursos e mantendo a confiabilidade do sistema para os usuários.

IV. METODOLOGIA

Neste estudo, foi avaliada a capacidade de mitigação de ataques de negação de serviço (DoS) em um gateway IoT utilizando um Raspberry Pi com OpenWrt e Snort configurado em dois modos distintos: IDS (Sistema de Detecção de Intrusão) e IPS (Sistema de Prevenção de Intrusão). A seguir, descrevemos detalhadamente os procedimentos realizados.

A. Configuração do Ambiente de Teste

Para conduzir os experimentos, foi montado um ambiente de teste que consistiu nos seguintes componentes:

- **Gateway IoT:** Um Raspberry Pi com OpenWrt instalado, atuando como o gateway IoT.
- **Sistema de Detecção/Prevenção de Intrusões:** Snort, configurado para operar tanto no modo IDS quanto no modo IPS.
- **Ferramenta de Ataque:** Ettercap, executado em uma máquina com Kali Linux, utilizado para gerar o tráfego de ataque (flood) simulando um ataque DoS.
- **Ferramenta de Coleta de Dados:** Data Acquisition (DAQ), para coletar os dados de tráfego de rede durante os experimentos.

B. Procedimentos Experimentais

Os procedimentos experimentais foram realizados em duas fases distintas, conforme descrito a seguir:

1) *Fase 1: Modo IDS:* Configuramos o Snort no Raspberry Pi para operar no modo IDS. Neste modo, o Snort apenas monitora o tráfego de rede e gera alertas quando detecta atividades suspeitas, sem interferir no tráfego.

Utilizando a ferramenta Ettercap no Kali Linux, realizamos um ataque de flood para simular um ataque de negação de serviço (DoS) contra o gateway IoT.

Durante o ataque, utilizamos o DAQ para coletar os dados de tráfego de rede, registrando os alertas gerados pelo Snort.

2) *Fase 2: Modo IPS:* Reconfiguramos o Snort no Raspberry Pi para operar no modo IPS. Neste modo, o Snort não apenas detecta, mas também bloqueia automaticamente os pacotes suspeitos.

Novamente, utilizamos a ferramenta Ettercap no Kali Linux para realizar um ataque de flood contra o gateway IoT.

Durante o ataque, utilizamos o DAQ para coletar os dados de tráfego de rede, registrando tanto os alertas quanto as ações de bloqueio executadas pelo Snort.

C. Coleta e Análise dos Dados

Após a execução dos experimentos, os dados coletados pelo DAQ foram analisados para comparar o desempenho do Snort nos modos IDS e IPS. Especificamente, foram avaliados os seguintes aspectos:

- Número de pacotes recebidos nos modos IDS e IPS.
- Número de pacotes analisados nos modos IDS e IPS.
- Número de pacotes permitidos nos modos IDS e IPS.
- Número de pacotes bloqueados nos modos IDS e IPS.
- Impacto do ataque de DoS na disponibilidade do gateway IoT em ambos os modos.

Os dados coletados foram então visualizados utilizando gráficos, permitindo uma análise comparativa clara da eficácia do Snort em mitigar ataques de DoS nos dois modos de operação. Com esta metodologia, buscamos demonstrar a eficácia do Snort como uma ferramenta de mitigação de ataques DoS, atendendo aos requisitos de segurança estabelecidos pelo Ato 77 da Anatel. Os resultados desta análise são apresentados na próxima seção.

V. RESULTADOS

Nesta seção, apresentamos e analisamos os dados coletados durante a execução do Snort em modos BLOCK e ALERT durante um minuto de ataque flood. Os dados fornecem uma visão comparativa do desempenho e comportamento do Snort em ambos os modos, destacando as diferenças e possíveis implicações para um gateway IoT.

TABLE I
DADOS DE TRÁFEGO NOS MODOS BLOCK E ALERT

Categoria	BLOCK	ALERT
received	470781	468937
analyzed	464076	457555
dropped	6684	11365
outstanding	6705	11382
outstanding_max	6705	11382
allow	5228	457104
whitelist	7	451
blacklist	458841	0
rx_bytes	28714834	48772917

A tabela acima apresenta os dados coletados durante a execução do Snort em modos BLOCK e ALERT durante um ataque flood.

A. Categorias

- **received:** Número total de pacotes recebidos pela interface de rede.
- **analyzed:** Número de pacotes analisados pelo Snort.
- **dropped:** Número de pacotes descartados pelo Snort.
- **outstanding:** Número de pacotes pendentes de análise.
- **outstanding_max:** Máximo de pacotes pendentes durante a execução.
- **allow:** Número de pacotes permitidos após a análise.
- **whitelist:** Número de pacotes permitidos com base em uma lista branca.
- **blacklist:** Número de pacotes bloqueados com base em uma lista negra.
- **rx_bytes:** Total de bytes recebidos.

B. Análise Comparativa

1) Pacotes Recebidos (received) e Analisados (analyzed):

O número de pacotes recebidos e analisados é semelhante entre os modos BLOCK e ALERT. No entanto, o modo BLOCK possui ligeiramente mais pacotes recebidos e analisados, indicando que mais pacotes foram processados.

2) *Pacotes Descartados (dropped)*: Observa-se que no modo ALERT, o número de pacotes descartados é significativamente maior (11,365) comparado ao modo BLOCK (6,684).

Intuitivamente, espera-se que o modo BLOCK descarte mais pacotes, já que sua função principal é bloquear pacotes indesejados de forma ativa. No entanto, a maior quantidade de pacotes descartados no modo ALERT pode ser explicada por alguns fatores:

Diferença entre Pacotes Descartados e Bloqueados

- **Pacotes Descartados (dropped)**: Estes são pacotes que foram recebidos pelo Snort, mas não foram processados completamente devido a limitações de recursos, como falta de memória ou capacidade de processamento, ou por não corresponderem a nenhuma regra específica. Isso pode acontecer em ambos os modos, mas parece ocorrer mais frequentemente no modo ALERT.
- **Pacotes Bloqueados (blacklist)**: No modo BLOCK, os pacotes que correspondem a regras específicas de bloqueio (listas negras) são diretamente bloqueados, o que não se reflete no número de pacotes descartados, mas sim na contagem de pacotes na categoria blacklist.

Carga de Processamento no Modo ALERT

- No modo ALERT, o Snort está configurado para analisar o tráfego e gerar alertas sem necessariamente bloquear pacotes. Isso pode levar a uma maior carga de processamento, especialmente durante um ataque flood, resultando em mais pacotes descartados devido à incapacidade de processar todos os pacotes recebidos em tempo hábil.

Eficiência na Aplicação de Regras

- O modo BLOCK pode ser mais eficiente na aplicação de regras de bloqueio, diretamente descartando ou bloqueando pacotes indesejados conforme correspondem às regras definidas. No modo ALERT, todos os pacotes são analisados para fins de geração de alertas, o que pode resultar em mais pacotes sendo descartados quando o sistema está sobrecarregado.

Essas observações indicam que, embora o modo BLOCK seja mais eficaz na mitigação de ataques ao bloquear pacotes específicos, o modo ALERT enfrenta desafios de desempenho durante picos de tráfego, levando a um maior número de pacotes descartados.

3) *Pacotes Pendentes (outstanding) e Máximo de Pendentes (outstanding_max)*: Tanto os pacotes pendentes quanto o máximo de pendentes são mais elevados no modo ALERT. Isto pode indicar uma maior carga de processamento no modo ALERT, onde os pacotes são apenas monitorados sem intervenção imediata.

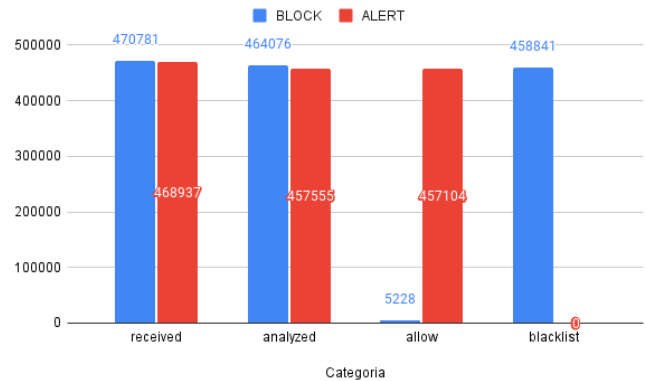


Fig. 1. Comparação do Snort em modo BLOCK e ALERT.

4) *Pacotes Permitidos (allow)*: No modo ALERT, um número muito maior de pacotes é permitido (457,104) em comparação ao modo BLOCK (5,228). Isso reflete a natureza do modo ALERT, que é monitorar e gerar alertas, mas não necessariamente bloquear os pacotes.

5) *Bytes Recebidos (rx_bytes)*: O modo ALERT tem um total de bytes recebidos (48,772,917) significativamente maior do que o modo BLOCK (28,714,834), o que pode ser resultado do processamento contínuo sem o bloqueio de pacotes.

6) *Análise dos Dados Destacados no Gráfico*: O gráfico destaca as categorias mais relevantes para comparação entre os modos BLOCK e ALERT do Snort. As categorias selecionadas são: **received**, **analyzed**, **allow**, e **blacklist**.

- **Pacotes Recebidos (received)**: O modo BLOCK processou um total de 470,781 pacotes, enquanto o modo ALERT processou 468,937 pacotes, uma diferença mínima de aproximadamente 0,39%. Isso indica uma consistência na capacidade de recebimento de pacotes entre os dois modos.
- **Pacotes Analisados (analyzed)**: O número de pacotes analisados também é similar entre os modos, com 464,076 pacotes no modo BLOCK e 457,555 no modo ALERT, resultando em uma diferença de cerca de 1,4%. Essa pequena variação sugere que ambos os modos têm uma eficiência de análise comparável.
- **Pacotes Permitidos (allow)**: Uma diferença significativa é observada nesta categoria, onde o modo ALERT permitiu 457,104 pacotes, enquanto o modo BLOCK permitiu apenas 5,228 pacotes. Isso demonstra que o modo ALERT deixa passar uma quantidade substancialmente maior de pacotes, representando cerca de 99% dos pacotes analisados. Em contraste, o modo BLOCK permite aproximadamente 1,1% dos pacotes recebidos, refletindo sua função de bloquear o tráfego indesejado.
- **Pacotes em Lista Negra (blacklist)**: O modo BLOCK bloqueou 458,841 pacotes, enquanto o modo ALERT não bloqueou nenhum pacote (0 pacotes). Isso reforça a função de cada modo: o modo BLOCK é configurado

para impedir que pacotes suspeitos cheguem ao destino, utilizando uma lista negra ativa, enquanto o modo ALERT se concentra em monitoramento e geração de alertas sem bloquear diretamente o tráfego.

7) *Implicações para um Gateway IoT*: Os dados destacados indicam implicações importantes para a implementação de um gateway IoT:

- **Segurança vs. Disponibilidade**: O modo BLOCK oferece uma camada de segurança mais robusta ao bloquear ativamente pacotes indesejados, o que é essencial para proteger dispositivos IoT de ataques maliciosos. No entanto, isso pode impactar a disponibilidade de serviços, pois pacotes legítimos podem ser erroneamente bloqueados.
- **Monitoramento e Diagnóstico**: O modo ALERT, ao permitir a passagem de mais pacotes, fornece uma visão abrangente do tráfego de rede, útil para monitoramento contínuo e diagnóstico de problemas. Contudo, a ausência de bloqueio ativo pode deixar o sistema vulnerável a ataques até que intervenções manuais sejam realizadas.
- **Uso de Recursos**: A escolha entre BLOCK e ALERT pode depender dos recursos disponíveis e dos requisitos de segurança. O modo BLOCK pode exigir menos capacidade de análise em tempo real, pois descarta pacotes potencialmente perigosos mais cedo, enquanto o modo ALERT pode demandar mais recursos para monitoramento contínuo e análise posterior.

8) *Uso da CPU durante os ataques de Flood*: Durante a execução dos testes de ataque flood, foi possível observar um comportamento distinto no uso da CPU do Raspberry Pi, dependendo do modo de operação do Snort (IDS ou IPS). Quando o Snort foi executado em modo IDS, o uso da CPU manteve-se em uma média constante de aproximadamente 55%, com um pico inicial de 55% que se estabilizou rapidamente e permaneceu constante durante todo o ataque.

Em contraste, no modo IPS, o uso da CPU apresentou um pico inicial ligeiramente mais alto, alcançando 60% nos primeiros segundos do ataque. No entanto, após cerca de 10 segundos, houve uma queda significativa no uso da CPU, que se estabilizou em torno de 20% e permaneceu nesse nível até o término do ataque.

Este comportamento distinto pode ser explicado pela natureza das operações realizadas por cada modo. No modo IDS, o Snort está configurado para analisar e gerar alertas para todo o tráfego de rede sem intervir diretamente no fluxo dos pacotes. Esse processo contínuo de monitoramento e análise resulta em um uso constante e relativamente elevado da CPU, uma vez que o sistema está continuamente processando todos os pacotes que entram na rede. Por outro lado, no modo IPS, o Snort não apenas analisa, mas também toma medidas ativas para bloquear pacotes indesejados. O pico inicial de 60% pode ser atribuído ao processamento intensivo necessário para identificar e bloquear pacotes potencialmente maliciosos no início do ataque. Após este processo inicial, a redução acentuada no uso da CPU sugere que o Snort

conseguiu estabelecer um estado estável de mitigação do ataque, reduzindo assim a carga de processamento, já que muitos pacotes foram descartados ou bloqueados antes de demandarem processamento adicional.

Essa diferença no uso da CPU entre os modos IDS e IPS é significativa para a implementação em dispositivos IoT com recursos limitados, como o Raspberry Pi, uma vez que a escolha do modo de operação pode impactar diretamente o desempenho do sistema e a eficácia na proteção contra ataques.

VI. CONCLUSÕES

Este estudo investigou a eficácia do Snort como ferramenta de mitigação de ataques DoS em gateways IoT, atendendo aos requisitos de segurança estabelecidos pelo Ato 77 da Anatel. Implementando o Snort em modos IDS (ALERT) e IPS (BLOCK), pudemos observar diferenças significativas no desempenho e comportamento do sistema durante um ataque flood. Os resultados demonstram que o modo BLOCK é mais eficaz na mitigação de ataques ao bloquear pacotes indesejados, garantindo uma proteção imediata ao gateway IoT. No entanto, essa abordagem pode impactar a disponibilidade do sistema, uma vez que alguns pacotes legítimos podem ser bloqueados erroneamente. Por outro lado, o modo ALERT permite um monitoramento detalhado do tráfego de rede sem bloquear pacotes, o que pode ser útil para diagnósticos e análises contínuas. No entanto, essa abordagem deixa o sistema mais vulnerável durante um ataque ativo, devido à ausência de bloqueio automático de pacotes suspeitos.

Portanto, a escolha entre os modos BLOCK e ALERT deve ser baseada nas necessidades específicas de segurança e disponibilidade do sistema. Para garantir uma proteção abrangente, uma abordagem híbrida que combine monitoramento detalhado com mecanismos de bloqueio ativos pode ser ideal para gateways IoT, atendendo às exigências do Ato 77 e protegendo contra ameaças cibernéticas. Este estudo contribui para a compreensão da implementação do Snort em ambientes IoT, destacando a importância de uma configuração adequada para garantir a segurança e a resiliência dos sistemas.

AGRADECIMENTOS

Agradecimento ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

REFERENCES

- [1] LINS, Fernando A. Aires; VIEIRA, Marco. Security requirements and solutions for iot gateways: A comprehensive study. *IEEE Internet of Things Journal*, v. 8, n. 11, p. 8667-8679, 2020.
- [2] BENIWAL, Gunjan; SINGHROVA, Anita. A systematic literature review on IoT gateways. *Journal of King Saud University-Computer and Information Sciences*, v. 34, n. 10, p. 9541-9563, 2022.
- [3] KHALID, Lawchak Fadhil; AMEEN, Siddeeq Y. Secure Iot integration in daily lives: A review. *Journal of Information Technology and Informatics*, v. 1, n. 1, p. 6-12, 2021.
- [4] DÍAZ LÓPEZ, Daniel et al. Developing secure IoT services: A security-oriented review of IoT platforms. *Symmetry*, v. 10, n. 12, p. 669, 2018.
- [5] VUJOVIĆ, Vladimir; MAKSIMOVIĆ, Mirjana. Raspberry Pi as a Sensor Web node for home automation. *Computers & Electrical Engineering*, v. 44, p. 153-171, 2015.

Aplicabilidade de Inteligência Artificial em Testes de Penetração: Um Mapeamento Sistemático

Lucas Ferreira Chagas Júnior

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Universidade Tecnológica Federal do Paraná (UTFPR)

Ponta Grossa – PR – Brasil

lucasjunior@alunos.utfpr.edu.br

Lourival Aparecido de Góis

Departamento de Computação

Universidade Tecnológica Federal do Paraná (UTFPR)

Ponta Grossa – PR – Brasil

gois@utfpr.edu.br

Abstract—Este artigo apresenta um mapeamento sistemático que discorre sobre a aplicabilidade da Inteligência Artificial (IA) em testes de penetração. O objetivo é identificar as fases e tarefas do pentest onde a IA pode ser útil seguindo a metodologia do Penetration Testing Execution Standard (PTES). Além disso, o estudo mostrou como os autores na literatura validaram a eficiência e a eficácia da IA em testes de penetração. A revisão incluiu trabalhos publicados nos últimos cinco anos em repositórios científicos. Os resultados mostram que é possível obter êxito ao aplicar Inteligência Artificial em testes de intrusão; no entanto, há possibilidade de melhorar e otimizar técnicas de IA nesse campo.

Index Terms—Inteligência Artificial, Testes de Penetração e Cibersegurança

I. INTRODUÇÃO

A crescente digitalização dos processos de transmissão de dados na internet transforma a segurança da informação em uma prioridade essencial no ambiente digital. Com o avanço constante da tecnologia, é importante manter a proteção e garantir a integridade desses dados. Neste contexto, um dos campos importantes da segurança da informação são os testes de penetração que desempenham um papel crucial na mitigação prévia de vulnerabilidades que possam ser exploradas por um cibercriminoso [1].

A segurança nunca é absoluta. Uma lição importante é que um atacante determinado sempre terá uma vantagem, e, com poucas exceções, quanto maior a empresa, mais vulnerável ela se torna. Isso ocorre porque há mais sistemas para monitorar, mais pontos de entrada e saída, e as fronteiras entre as unidades de negócios podem se tornar confusas, resultando em um maior número de usuários. No entanto, isso não significa que se deve perder a esperança; o conceito de “segurança através da conformidade” não é suficiente por si só para garantir a proteção [2].

Visando resolver questões de segurança, Weidman (2014) define que testes de invasão ou pentesting envolvem a simulação de ataques reais para avaliar os riscos associados a potenciais brechas de segurança. Em um teste de invasão, os pentesters não só identificam vulnerabilidades que poderiam ser usadas pelos invasores, mas também exploram essas vulnerabilidades sempre que possível para avaliar o que os invasores poderiam obter após uma exploração bem-sucedida das falhas.

O Penetration Testing Execution Standard (PTES) é uma metodologia reconhecida na comunidade de segurança da informação que é utilizada para a realização de testes de penetração. Ela fornece um conjunto de diretrizes que cobrem todas as etapas dos testes de intrusão, desde o planejamento inicial até a exploração pós-exploração e geração de relatórios. O PTES visa padronizar os procedimentos de pentesting, assegurando que todas as áreas críticas de segurança sejam devidamente testadas e que seja possível correção de resultados e correção dos problemas.

Como estes testes requerem amplo conhecimento técnico em diversas áreas de tecnologia da informação, é importante destacar que esses processos vêm sendo aprimorados com técnicas de Inteligência Artificial, conforme mostram os estudos de Confido et al. (2022) e Hilário et al. (2024).

O objetivo deste artigo foi mapear como as ferramentas de Inteligência Artificial podem ser aplicadas nas fases do pentest baseado na metodologia do PTES (Penetration Testing Execution Standard) e mostrar como os autores validaram sua eficiência e eficácia.

O mapeamento abrange estudos dos últimos cinco anos presentes em repositórios como IEEE Xplore, ACM Digital Library, Scopus e ScienceDirect, seguindo a metodologia de Kitchenham e Charters (2007).

Este artigo foi estruturado da seguinte maneira: A Seção II apresenta a estrutura organizacional dos testes de intrusão. A Seção III apresenta um mapeamento detalhado da literatura sobre a aplicabilidade das abordagens de IA em testes de penetração, abordando os principais avanços estudados. A Seção IV descreve a metodologia utilizada para o mapeamento sistemático, incluindo os critérios de inclusão e exclusão e os procedimentos de busca nos repositórios selecionados. A Seção V discute os resultados obtidos, respondendo às questões de pesquisa deste artigo. Finalmente, a Seção VI apresenta as conclusões do estudo, ressaltando a eficácia das técnicas de IA em melhorar os testes de penetração e sugerindo direções futuras para pesquisa nesta área.

II. TESTE DE PENETRAÇÃO

Os testes de penetração podem ser definidos como uma tentativa legal e autorizada de localizar e explorar com sucesso

sistemas de computador com o propósito de tornar estes ambientes mais seguros. O processo inclui a busca por vulnerabilidades, bem como a realização de ataques de prova de conceito para demonstrar que as vulnerabilidades são reais. Um teste de penetração adequado sempre termina com recomendações específicas para resolver e corrigir os problemas que foram descobertos durante o teste. De modo geral, este processo é utilizado para ajudar a proteger computadores e redes contra futuros ataques. A ideia geral é encontrar problemas de segurança usando as mesmas ferramentas e técnicas que um atacante. Essas descobertas podem ser mitigadas antes que um hacker real as explore [3].

Como define o estudo de Engebretson (2013), os testes de penetração também são conhecidos como: Pentesting, PT, Hacking, Ethical hacking, White hat hacking, Offensive security e Red teaming. Consoante a isso, Weidman (2014) categoriza os testes de intrusão em três principais classificações:

- **Black-box:** O profissional não possui informações prévias sobre o sistema alvo.
- **White-box:** O profissional tem acesso completo à estrutura interna do sistema.
- **Gray-box:** O profissional tem acesso limitado, conhecendo apenas algumas partes do sistema.

Para a realização do pentest que demanda muitas tarefas, segue-se a metodologia do PTES (Penetration Testing Execution Standard). Esse padrão, como já mencionado, é um esforço colaborativo entre especialistas e profissionais de segurança da informação, organizando o pentest em várias fases, cada uma com objetivos específicos. A seguir, são apresentadas essas fases e seus breves resumos adaptados do seu site oficial para melhor entendimento:

- **Interações pré-engajamento:** Estabelecimento das expectativas do teste, incluindo os objetivos, escopo e as regras de engajamento entre o cliente e a equipe de teste.
- **Coleta de informação:** Coleta de informações iniciais sobre o alvo através de diversas técnicas.
- **Modelagem de ameaças:** Identificação e categorização de possíveis ameaças ao sistema com base nas informações coletadas, criando um modelo de ameaças para direcionar o teste.
- **Análise de Vulnerabilidade:** Avaliação detalhada das informações coletadas para identificar vulnerabilidades específicas nos sistemas e redes alvo.
- **Exploração:** Tentativa de explorar as vulnerabilidades identificadas para comprovar sua severidade e o impacto potencial.
- **Pós-exploração:** Realização de atividades para consolidar o acesso obtido, avaliar o impacto da exploração e identificar formas de manter o acesso persistente.
- **Relatório:** Documentação detalhada dos achados do teste, incluindo as vulnerabilidades exploradas, métodos utilizados e recomendações para mitigação e melhoria da segurança.

Após detalhar os testes de intrusão e sua metodologia organizacional do PTES, a seguir será apresentada a revisão

da literatura sobre aplicabilidade da IA em penetration testing.

III. INTELIGÊNCIA ARTIFICIAL E OS TESTES DE INTRUSÃO

A Inteligência Artificial (IA) está sendo utilizada como uma ferramenta importante para aprimorar os testes de penetração, principalmente por meio de técnicas como aprendizado de máquina (ML) e aprendizado por reforço (RL). Essas técnicas estão sendo integradas nas fases e nos processos de pentest, promovendo a automação e otimização da detecção e exploração de vulnerabilidades.

Os avanços na aplicabilidade da IA em pentest têm mostrado resultados promissores em diversas áreas. Confido et al. (2022) demonstraram que a integração de técnicas de ML, especialmente o aprendizado por reforço (RL), em frameworks de pentest pode automatizar a identificação de caminhos de ataque ótimos e priorizar vulnerabilidades com base na criticidade. Consoante a isso, com abordagem diferente, Hilário et al. (2024) destacaram o uso de IA generativa (GenAI) para criar novos dados e padrões, auxiliando na identificação de vulnerabilidades com a API do Chat GPT (Shell GPT) visando aprimorar as tarefas do pentest.

A eficiência e a redução da necessidade de intervenção humana são outras áreas onde a IA tem demonstrado um impacto significativo. Kasim et al. (2020) comprovaram que a utilização de competidores adversariais de IA pode automatizar tanto os aspectos de ataque quanto de defesa nos testes de intrusão, reduzindo significativamente a necessidade de intervenção humana. Além disso, com a revisão sistemática de McKinnel et al. (2019) foi possível identificar que a combinação de RL com abordagens tradicionais pode aumentar a eficiência dos testes, economizando tempo e recursos ao automatizar ações repetitivas e complexas.

A adaptação contínua às evoluções tecnológicas é importante para manter a eficácia dos pentests. Saber et al. (2023) observaram que, embora a IA tenha demonstrado utilidade significativa em testes de invasão, é necessário acompanhar os avanços tecnológicos e a evolução dessas tecnologias.

IV. METODOLOGIA DE PESQUISA

A fim de realizar uma busca sobre o uso da Inteligência Artificial aplicada a Testes de Penetração, adotou-se a Metodologia de Mapeamento Sistemático de Kitchenham e Charters [6]. Em consonância ao modelo adotado, a pesquisa foi realizada considerando os últimos cinco anos visando abranger artigos mais recentes sobre o estudo.

A. Questões de Pesquisa e Repositórios de Busca

Para obter êxito ao objetivo proposto no Artigo, foram definidas duas principais questões de Pesquisa: (Q1) Como as fases de Testes de Intrusão são utilizadas com auxílio da IA? (Q2) Qual é o impacto da aplicação de IA em Testes de Invasão em termos de eficácia e eficiência?

Para responder essas questões e obter resultados relevantes, foram selecionados os principais repositórios relevantes para a Área de Ciber Segurança e Inteligência Artificial: IEEE Xplore, ACM Digital Library, Scopus e ScienceDirect.

B. Estratégia de Busca e Critérios de Inclusão e Exclusão

Utilizou-se palavras-chave específicas para construir strings de busca eficazes nos repositórios já mencionados. As palavras-chave incluíam "Artificial Intelligence", "AI", "Penetration Testing", "Invasion Tests" e "Cyber Security". Logo, a string de busca ficou com a seguinte estrutura: ("Artificial Intelligence" OR "AI") AND ("Penetration Testing" OR "Invasion Tests") AND "Cyber Security".

Aplicou-se a string nos três diferentes repositórios com foco em publicações entre 2019 e 2024. Após aplicar, obteve-se os resultados conforme a Tabela I.

TABLE I
ARTIGOS OBTIDOS APÓS APLICAR A STRING DE BUSCA NOS REPOSITÓRIOS

Base de Dados	Resultado de Artigos
IEEE Xplore	30
ACM Digital Library	79
Scopus	22
ScienceDirect	173
Total	304

É importante entender que esses resultados são preliminares, pois os artigos resultantes ainda serão submetidos a critérios de inclusão e exclusão para garantir a relevância ao estudo de cada artigo escolhido.

Para obter os artigos mais relevantes, foram definidos critérios de Inclusão e Exclusão conforme a Tabela II.

TABLE II
CRITÉRIOS DE INCLUSÃO E EXCLUSÃO PARA SELEÇÃO DOS ARTIGOS

Inclusão	Artigos que incluem "Teste de Intrusão" e "Inteligência Artificial" juntos no título
	Artigos que possuem relevância ao foco principal do trabalho: Utilizar a Inteligência Artificial para os Testes de Intrusão
	Artigos gratuitos ou de acesso universitário
Exclusão	Artigos não relacionados diretamente ao uso de IA em testes de invasão
	Artigos duplicados

C. Seleção dos Estudos

Após aplicar os critérios de inclusão e exclusão, seis artigos foram relevantes, representando aproximadamente 198

A qualidade dos artigos foi avaliada observando a relevância do conteúdo em relação às questões de pesquisa. Durante a leitura dos estudos primários, um artigo foi retirado por não ter relação direta com as questões de pesquisa, totalizando cinco artigos para análise. Estes artigos foram estudados em detalhes. Este processo de seleção está demonstrado na Figura 1.

V. RESULTADOS

Após realizar a seleção dos estudos, foram estudados os cinco artigos, os quais são destacados conforme a Tabela III.

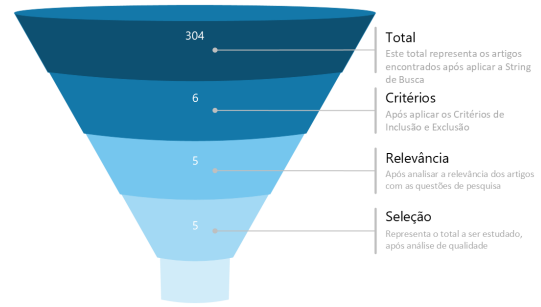


Fig. 1. Seleção dos Estudos

TABLE III
ARTIGOS ESTUDADOS

Autor(es)	Título	Ano
McKinnel et al.	A systematic literature review on AI in pentesting	2019
Kasim et al.	Cybersecurity as a Tic-Tac-Toe Game	2020
Confido et al.	Reinforcing Penetration Testing Using AI	2022
Saber et al.	Automated Penetration Testing: A Systematic Review	2023
Hilario et al.	Generative AI for pentesting	2024

A. Q1. Como as fases de Testes de Intrusão são utilizadas com auxílio da IA?

Para responder esta pergunta, foi realizada uma análise nos cinco estudos selecionados, porém apenas dois do total tiveram relevância direta com a questão de pesquisa: Confido et al. (2022) e Hilário et al. (2024).

O estudo de Confido et al. (2022) se concentra na aplicação de técnicas de Aprendizado por Reforço (Reinforcement Learning, RL) para melhorar a eficiência e a eficácia dos testes de penetração automatizados utilizando o PenBox, um framework desenvolvido pela Agência Espacial Europeia (ESA). O objetivo principal é automatizar e otimizar o processo de pentesting, permitindo que um agente de RL aprenda e execute ataques cibernéticos de forma autônoma, minimizando a necessidade de intervenção humana.

Com a abordagem de RL, um script keyboard agent.py imita a ação de um pentester/atacante real que executa a ação passo a passo e escolhe manualmente qual será a próxima ação. O experimento realizado pelos autores Confido et al. (2022) aborda as fases principais do pentest conforme denominadas pelos autores, que incluem:

- Coleta de informações: No experimento, o PenBox automatiza a coleta de dados utilizando ferramentas de varredura para obter informações detalhadas sobre o alvo. O RL ajudou na tomada de decisões.
- Análise de Vulnerabilidades: A análise de vulnerabilidades com o PenBox é feita de forma automatizada, e ferramentas como Nmap e Wireshark são utilizadas só para a descoberta de vulnerabilidades, enquanto o RL ajuda a priorizar e validar essas vulnerabilidades com

base na criticidade.

- Exploração: O PenBox, utilizando RL, automatiza a seleção da sequência de ações mais eficazes para explorar as vulnerabilidades. Ferramentas como o Metasploit Framework são utilizadas para comprometer o sistema alvo. O RL otimiza o caminho de ataque, reduzindo o número de ações necessárias para alcançar posições estratégicas na rede.
- Pós-Exploração: O RL foi utilizado para avaliar o nível de comprometimento do sistema e manter o acesso ao sistema. Isso inclui técnicas de pilhagem de informações e configuração de métodos de persistência para acesso contínuo.

Ao concluir o experimento, Confido et al. (2022) afirmam que os melhores resultados são alcançados com um número grande de etapas de treinamento e reduzindo drasticamente as etapas de exploração. Essa estratégia permite que o agente atinja o equilíbrio correto entre explorar novas ações e focar naquelas que já se mostraram lucrativas.

Com uma abordagem diferente, o estudo Hilario et al. (2024) aborda o uso da GenAI, especificamente a GenAI, que é um subcampo da inteligência artificial que se concentra na criação de novos dados, padrões ou modelos com base em dados existentes; abrange várias técnicas, incluindo aprendizado profundo, processamento de linguagem natural (NLP) e visão computacional. DALL-E, MidJourney, Stable Diffusion, Google Bard, GitHub Copilot, Bing AI Chat e Microsoft 365 Copilot estão entre alguns dos nomes mais proeminentes em GenAI hoje, mas talvez o mais popular seja o GPT da OpenAI, uma série de Modelos de Linguagem de Grande Escala (LLMs) multimodais, atualmente em sua quarta iteração [5].

Para conectar os estudos às fases e à realização do pentest em suas fases, foi utilizada a ferramenta Shell GPT, uma ferramenta de interface de linha de comando (CLI) baseada em Python que utiliza a API do Chat GPT para responder a perguntas, gerar comandos de shell, snippets de código e documentação foi integrada ao ambiente de pentesting. Usar a API do Chat GPT para se conectar com ferramentas utilizadas em pentesting como Nmap, Nessus e OpenVAS envolve o uso de Python ou outra linguagem de script para criar uma interação entre o Chat GPT e cada ferramenta.

O experimento dois utilizado vai das fases conforme denominadas pelo autor de Reconhecimento até Exploração de um alvo. Será explanado um resumo da aplicabilidade em cada fase do Pentest:

- Coleta de informações: A identificação de informações iniciais do alvo, o uso da IA Shell GPT foi utilizado junto com o Nmap para obter o IP (Internet Protocol) da máquina local e sondar a rede, facilitando identificar hosts ativos na rede e seus principais serviços.
- Análise de Vulnerabilidades: Ao coletar informações detalhadas sobre cada serviço que havia nos hosts ativos na rede, foi realizado o uso da ferramenta Shell GPT para auxiliar no login com FTP anonymous, leitura de arquivos importantes e análise de código fonte da aplicação.

- Exploração: Só foi possível obter acesso inicial ao alvo com o auxílio da ferramenta que automatiza processos de varredura no serviço Wordpress e automatização de varreduras de diretórios com o Gobuster, identificando vulnerabilidades e usuários adicionais.
- Pós-Exploração: Para escalar privilégios, o Shell GPT auxiliou na decodificação de mensagens, localização e configuração de chaves privadas OpenSSH e na execução de comandos para obter acesso root, automatizando o processo de exploração de vulnerabilidades.
- Relatório: O formato do relatório pode variar amplamente para cada alvo e para cada cliente. Chat GPT pode ser usado para personalizar o relatório com base nas especificações do cliente, na natureza e nas conclusões do teste. Em termos de qualidade, pode produzir um resultado preciso, o relatório mais completo e refinado com base na capacidade do Chat GPT de verificar erros, inconsistências e identificar áreas que precisam de esclarecimento [5].

Com esta análise, é possível identificar que as fases dos Testes de Penetração podem ser auxiliadas tecnicamente com abordagens diferentes de inteligência artificial. A Tabela IV resume a aplicação de Inteligência Artificial nas fases de Testes de Penetração (PTES) conforme os estudos de Confido et al. (2022) e Hilário et al. (2024). A tabela mostra se e como cada fase do pentesting foi relevante e utilizada nos experimentos desses estudos.

TABLE IV
IA APLICADA ÀS FASES DOS TESTES DE PENETRAÇÃO

Fases do Pentesting	Confido et al. (2022)	Hilário et al. (2024)
Interações pré-engajamento	Não relevante	Não relevante
Coleta de informação	Utilização do RL com PenBox	Utilização com Shell GPT e ferramentas de pentest
Modelagem de ameaças	Não relevante	Não relevante
Análise de Vulnerabilidade	Automatização com RL e ferramentas	Utilização com Shell GPT e ferramentas de pentest
Exploração	RL para otimizar a sequência de exploração	Utilização com Shell GPT e ferramentas de pentest
Pós-exploração	RL para manter acesso e pilhagem de informações	Utilização com Shell GPT e ferramentas de pentest
Relatório	Não realizado	GenAI e LLMs auxiliaram na escrita e melhoria do relatório

B. Q2. Qual é o impacto da aplicação de IA em testes de invasão em termos de eficácia e eficiência?

Para responder a esta questão de pesquisa, foi possível extrair respostas dos cinco estudos selecionados: Saber et al. (2023), Kasim et al. (2020), Confido et al. (2022), McKinnel et al. (2019) e Hilário et al. (2024).

O estudo Kasim et al. (2020) provou que ao realizar o experimento “A trained Adversarial AI Competitor” que automatiza testes de intrusão em um espaço restrito (um jogo da velha) para automatizar testes de intrusão utilizando Inteligência Artificial, foi possível identificar que a IA bem treinada é eficaz em ataque e defesa, demonstrando eficácia e eficiência. Consoante ao experimento, também foi realizado um estudo através do estudo de Confido et al. (2022) de um experimento da aplicabilidade direta de ML/IA em testes de penetração automatizados que reduziu a interação humana e otimizou o número de ações. O artigo estuda a eficácia de combinar RL com abordagens tradicionais para melhorar ainda mais a eficiência e eficácia dos testes de penetração, economizando tempo e recursos.

Considerando o estudo McKinnel et al. (2019), é possível afirmar que, embora a utilidade da IA em testes de penetração e avaliação de vulnerabilidades tenha sido demonstrada em estudos existentes, precisamos acompanhar os avanços tecnológicos e a evolução das técnicas adversárias para evitar detecção.

Outra abordagem de AI é a GenAI estudada por Hilário et al. (2024). Foi estudado que, através do uso da GenAI, a eficiência e a eficácia dos esforços de pentesting podem ser significativamente aprimoradas, o que resulta em sistemas mais robustos e seguros. Os LLMs podem analisar rapidamente grandes quantidades de dados e gerar cenários de teste com base em vários parâmetros, agilizando o processo de teste e economizando tempo valioso para os profissionais de segurança.

Portanto, a importância desta questão de pesquisa, considerando o estudo Saber et al. (2023), ajuda a concluir que o pentest é mais preciso quando há o uso de automatização e algoritmos.

A Tabela V resume os métodos de validação e os resultados obtidos em todos os estudos já mencionados que analisaram o impacto da aplicação de técnicas de Inteligência Artificial em testes de invasão.

VI. CONCLUSÃO

Conforme avaliado, sabe-se que a segurança da informação é um campo da tecnologia da informação importante e as técnicas de pentest podem ser utilizadas para prevenir atacantes mal intencionados. Com isso, pode-se usar a IA como uma aliada para atingir êxito na mitigação desses processos.

A análise dos estudos permitiu concluir que a aplicação de Inteligência Artificial nas fases dos Testes de Penetração aumenta consideravelmente a eficiência e a eficácia dos testes. Tecnologias como, por exemplo, Aprendizado por Reforço e GenAI, como demonstrados nas questões de pesquisa, são

TABLE V
DUAS ABORDAGENS DISTINTAS DE IA APLICADA AO PENTEST

Estudo	Método de Validação	Resultados
McKinnel et al. (2019)	Meta-análise de estudos empíricos	Algoritmos baseados em PDDL performam melhor em planejamento de ataques. Algoritmos genéticos mostram melhorias significativas na eficiência. Escalabilidade ainda é um desafio.
Kasim et al. (2020)	Sistema de IA adversarial treinado	IA mostrou-se quase impossível de vencer por humanos; agentes de IA bem treinados exploram vulnerabilidades eficazmente.
Confido et al. (2022)	Uso de RL para comprometer a rede simulada	Redução significativa no número de ações humanas necessárias.
Saber et al. (2023)	Revisão sistemática de técnicas automatizadas	Eficiência melhorada com automação; desafios em casos específicos e sensibilidade.
Hilário et al. (2024)	Utilização de GenAI em pentesting	Chat GPT agilizou o processo, aumentou a precisão das detecções e melhorou a qualidade dos relatórios.

eficazes na automação de processos relevantes que percorrem as fases do pentest: Coleta de informação, Modelagem de ameaças, Análise de Vulnerabilidade, Exploração e Pós-exploração. Portanto, foi possível diminuir a interação humana e otimizar as ações necessárias para comprometer sistemas vulneráveis.

Para o avanço deste estudo, futuras pesquisas devem ser realizadas para explorar novas técnicas de IA visando melhorar e otimizar processos para mitigação de novas vulnerabilidades que irão surgir. É importante desenvolver ferramentas e algoritmos para que a IA possa auxiliar profissionais que mitigam estes problemas cibernéticos. Além disso, pode-se destacar que deve-se realizar a aplicabilidade da IA em diversos cenários considerando a diversidade de novas tecnologias e a complexidade da tramitação de dados.

REFERENCES

- [1] G. Weidman, *Testes de Invasão*. No Starch Press, 2014.
- [2] W. Allsopp, *Advanced Penetration Testing: Hacking the World's Most Secure Networks*. Wiley, 1ª ed., 2017.
- [3] P. Engebretson, *The Basics of Hacking and Penetration Testing*, 2ª ed. Elsevier, 2013.
- [4] A. Confido, E. V. Ntagiou, M. Wallum, “Reinforcing Penetration Testing Using AI,” in *Proc. IEEE*, 2022.
- [5] E. Hilario, S. Azam, J. Sundaram, K. I. Mohammed, B. Shanmugam, “Generative AI for Pentesting: The Good, The Bad, The Ugly,” in *Proc. Scopus*, 2024.
- [6] B. Kitchenham, S. Charters, “Guidelines for Performing Systematic Literature Reviews in Software Engineering,” Keele University and Durham University Joint Report, Tech. Rep. EBSE-2007-01, 2007.
- [7] S. Kasim, S. S. Samadi, N. Valliani, L. Watkins, N. K. Wong, A. Rubin, “Cybersecurity as a Tic-Tac-Toe Game Using Autonomous Forwards (Attacking) And Backwards (Defending) Penetration Testing in a Cyber Adversarial Artificial Intelligence System,” in *IEEE Xplore*, 2020.
- [8] V. Saber, D. ElSayad, A. M. Bahaa-Eldin, Z. T. Fayed, “Automated Penetration Testing: A Systematic Review,” in *IEEE Xplore*, 2023.

- [9] D. R. McKinnel, T. Dargahi, A. Dehghantanha, K. K. R. Choo, "A Systematic Literature Review and Meta-Analysis on Artificial Intelligence in Penetration Testing and Vulnerability Assessment," *ScienceDirect*, 2019.

Aplicação de Técnicas de Agrupamento na Análise de Tráfegos de Rede

Luis Marinho
Instituto Metrópole Digital
UFRN
Natal, Brasil
fhelipe.marinho@ufrn.br

Marcos Pinheiro
Departamento de Informática e
Matemática Aplicada, UFRN
Natal, Brasil
marcos.madruga@ufrn.br

Roger Immich
Instituto Metrópole Digital
UFRN
Natal, Brasil
roger@imd.ufrn.br

Abstract—O uso de redes em ambientes corporativos é essencial para a comunicação, compartilhamento de recursos e execução de aplicações. Monitorar e gerenciar a rede garante performance, segurança e disponibilidade. Diferentes perfis de uso de tráfego, como acesso a vídeos, arquivos e videomonitoramento, exigem ajustes específicos na alocação de recursos e medidas de segurança, otimizando a eficiência e proteção da rede. Nesse trabalho foram aplicadas técnicas de aprendizado de máquina não supervisionado a um *dataset* contendo informações coletadas das interfaces dos ativos de rede, com o objetivo de determinar o perfil dos usuários da rede nos setores acadêmicos, administrativo e de aula da Universidade Federal do Rio Grande do Norte (UFRN). Foram encontradas particularidades no uso da rede como os equipamentos da rede cabeada de setores acadêmicos que consomem mais recursos da rede que outras máquinas e que o período vespertino possui a maior quantidade de máquinas conectadas, principalmente na rede Wi-Fi. Foi observado um comportamento anômalo no perfil de tráfego de algumas máquinas de videomonitoramento, indicando um possível erro de configuração em algumas interfaces de alguns ativos de rede.

Index Terms—Machine Intelligence, Computer Networks, Performance Analysis, Networking

I. INTRODUÇÃO

O uso de rede refere-se à maneira como os dispositivos e usuários interagem e trocam informações em um sistema interconectado [1]. Os dados de rede representam todas as informações geradas a partir dessas interações. Isso inclui o tráfego de dados, padrões de acesso, tipos de dispositivos conectados, tempos de resposta e volumes de transferência.

Esses dados são cruciais porque fornecem perspectivas sobre o funcionamento e o desempenho da rede. Eles permitem identificar gargalos, prever falhas, otimizar o uso de recursos e garantir a segurança. Além disso, a análise dos dados de rede pode revelar padrões de comportamento dos usuários, auxiliar na detecção de atividades anômalas e contribuir para a tomada de decisões informadas sobre expansões e atualizações de infraestrutura [2], [3].

Através da análise de dados de rede, é possível melhorar a eficiência operacional, reduzir custos, aumentar a satisfação dos usuários e proteger informações sensíveis contra ameaças cibernéticas. Esses dados são, portanto, fundamentais para o gerenciamento e a manutenção eficaz de qualquer rede de comunicação moderna.

A ciência de dados pode ser utilizada para facilitar a análise das informações de rede. Ela é uma área de estudo multidisciplinar que utiliza diversas tecnologias, modelos e metodologias para extrair informações de dados, sejam eles estruturados ou não. Uma das técnicas da ciência de dados é o aprendizado de máquina (AM), que é compreendido um conjunto de métodos utilizados para a identificação automática de padrões por um algoritmo, que podem ser utilizados para a tomada de decisões ou realização de ações.

A depender dos dados coletados, a análise pode ser feita de três formas: supervisionada, não supervisionada e por reforço. Uma das técnicas não supervisionadas é o agrupamento (*clustering*), que tem como objetivo encontrar uma estrutura de clusters (grupos) nos dados em que os objetos pertencentes a cada cluster compartilham alguma característica ou propriedade relevante para o domínio do problema em estudo.

II. MOTIVAÇÃO

A rede de dados da UFRN possui a característica de ser heterogênea, tanto na diversidade dos equipamentos que transmitem os dados quanto na variedade de serviços e clientes que a utilizam diariamente. A saída do tráfego de rede é centralizada no campus central, mas os dados vêm de todos os Campi da UFRN. São três Campi no interior do estado, e pelo menos uma dúzia de setores e prédios externos, atendendo desde setores administrativos, hospitais, residências universitárias, museus, entre outros.

A quantidade de usuários atendidos cresce a cada prédio construído e a cada setor coberto com infraestrutura de rede sem fio corporativa. Atualmente uma média de 10.000 usuários simultâneos utilizam a rede sem fio corporativa da UFRN. Esse tráfego sem fio considera clientes visitantes, discentes e servidores, além de uma demanda crescente vinda dos equipamentos de Internet das Coisas (IoT). Nos últimos anos foi realizada uma migração dos serviços de telefonia e videomonitoramento, saindo dos modelos analógicos para digitais, aumentando a demanda pelos serviços de infraestrutura de rede.

Tendo essa diversidade em mente, evidenciasse a necessidade de se conhecer melhor os clientes que utilizam a rede, entender os padrões de uso e as possibilidades de melhorias, tendo em vista aprimorar a qualidade do serviço prestado à

comunidade acadêmica. Com isso, será possível desenvolver estratégias mais eficazes e personalizadas que atendam às demandas específicas dos usuários.

Esse trabalho propõe a utilização de técnicas de aprendizado de máquina não supervisionado aplicadas em um *dataset* contendo dados coletados dos equipamentos de rede da UFRN. O objetivo é encontrar relações entre os tipos de tráfego na rede e os diversos tipos de usuários e setores da universidade. Isso possibilita, entre outros, dimensionar corretamente enlaces de rádio e links, principalmente para prédios externos aos campi da UFRN.

III. CONCEITOS RELACIONADOS

Informação é um conjunto de dados que, ao ser processados adequadamente, se tornam compreensíveis e significativos, criando padrões e acionando significados na mente das pessoas [4]. A ciência de dados é o ramo do conhecimento que desenvolve e aplica essas técnicas.

Essas técnicas são aplicadas em bases de dados (*datasets*) coletadas com dados relevantes para o objetivo desejado. Essas bases são fundamentais durante o processo de aprendizagem, cada algoritmo utiliza os exemplos apresentados no conjunto de treinamento para tomar decisões conforme a tarefa em questão [5].

Uma das formas utilizadas pela ciência de dados é a utilização de técnicas de agrupamento (clusterização) dos dados. O objetivo de uma técnica de agrupamento é identificar uma estrutura de *clusters*, onde os objetos em cada *clusters* compartilham características ou propriedades relevantes para o domínio do problema, ou seja, são de alguma forma similares. Algumas definições comuns para cluster são: *cluster* bem separado, aonde os pontos de determinado agrupamento estão mais próximos entre si do que de outros agrupamentos e *Cluster* baseado em centro, aonde os pontos de um agrupamento estão mais próximos ao centro do seu grupo do que o centro de outros agrupamentos [6].

Alguns dos algoritmos de agrupamento mais utilizados são: K-Means, *Expectation Maximization* (EM) e o algoritmo hierárquico aglomerativo. O K-Means é um algoritmo de agrupamento não hierárquico que utiliza iterações para dividir um conjunto de dados em *k* clusters, procurando minimizar a distância dos elementos em relação aos *k* centros definidos [7]. O algoritmo EM permite aprender os parâmetros que governam a distribuição dos dados amostrais, especialmente quando algumas características estão ausentes, sendo realizado em duas etapas: E, para estimar a matriz de probabilidade, e M, para maximizar os pesos, média e matriz de covariância [8]. Já no algoritmo hierárquico aglomerativo, cada objeto individual inicialmente forma seu próprio cluster, esses grupos vão se aglutinando de acordo com suas similaridades, formando uma estrutura de árvore [9].

Para garantir que os resultados sejam significativos, os dados devem ser tratados antes da aplicação dessas técnicas. Preparar os dados para o agrupamento envolve diversos aspectos, como o pré-processamento dos dados e a escolha de uma representação adequada para ser utilizada por um algoritmo

de agrupamento [6]. São etapas do processo de preparação dos dados: desconsiderar atributos irrelevantes para o objetivo desejado (números sequenciais, identificadores, etc.); retirar objetos que contenham atributos vazios ou inválidos; substituir valores de atributos vazios ou inválidos; identificar atributos com valores discrepantes dos demais (outliers); transformar atributos categóricos (texto) em binários; normalizar dados numéricos; transformar dados numéricos em ordinais, quando for o caso; e reduzir o volume de dados (objetos ou atributos), para reduzir o esforço computacional necessário.

Terminado o processo de manipulação e categorização dos dados, é realizada a validação dos resultados. Duas técnicas comuns para validação são o índice Silhueta e o índice de Davies-Bouldin (DB). O primeiro avalia a proximidade dos objetos dentro de um cluster e a distância destes ao cluster mais próximo. Esse índice é utilizado para avaliar a qualidade de uma partição e considera tanto a adequação de cada objeto ao seu cluster quanto a qualidade geral de cada cluster, com valores que variam entre -1 e 1, sendo que uma partição ideal possui valor máximo, 1 [6]. O índice DB é utilizado para averiguar se os clusters estão bem separados e bem compactos. Quanto mais próximo de 0, melhor será a qualidade dos clusters.

IV. METODOLOGIA DOS EXPERIMENTOS

Devido a especificidade do objetivo desejado, o *dataset* foi coletado durante o período de desenvolvimento desse trabalho. Posteriormente, os dados foram tratados e as técnicas citadas no tópico anterior foram aplicadas nos dados.

A. Coleta e tratamento dos dados

A Figura 1 apresenta o cenário de coleta para cada tipo de setor selecionado. Para construir os dados, foram selecionados 30 switches localizados em setores distintos da universidade, sendo 10 em prédios administrativos, 10 em setores de aula e 10 em centros acadêmicos.

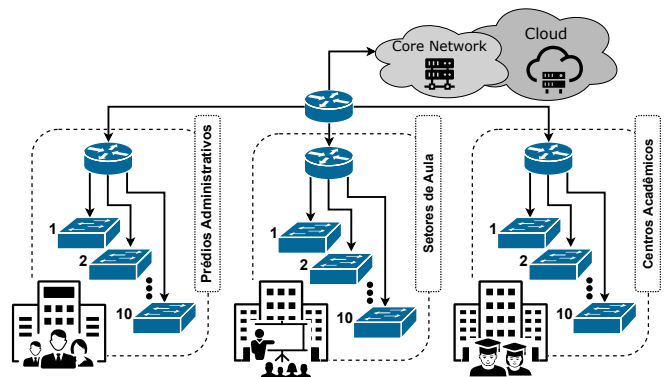


Fig. 1. Cenário de coleta. Elaboração própria.

Cada interface desses ativos foi listada, sendo excluídas as interfaces de uplink, derivação e as interfaces onde não havia nenhum equipamento conectado. Das interfaces restantes foram consultadas: a quantidade de máquinas conectadas;

foram coletadas as taxas de transferências de download e upload; o tipo de usuário conectado; o setor ao qual o switch está instalado; e o período da consulta. Foram definidos três tipos de usuários: Wi-Fi; monitoramento, utilizado para os equipamentos de segurança patrimonial; e acesso cabeado.

As coletas foram realizadas nos dias 08 e 09 de maio de 2023, a cada hora no período das 08h00 às 22h00. Com esses dados, foi definido o *dataset* inicial, contendo as seguintes colunas: Quantidade de Máquinas; Banda de Download; Banda de Upload; Tipo de Usuário; Setor; e Período.

A partir do *dataset* coletado, as colunas que continham apenas dados categóricos, entre elas: 'Tipo Usuário', 'Setor' e 'Período', foram binarizadas. Esse processo resultou na criação das seguintes colunas: Tipo Usuário Monitoramento, Tipo Usuário Wifi, Tipo Usuário Cabeado, Setor Acadêmico, Setor Administrativo, Setor Aula, Período Matutino, Período Noturno e Período Vespertino. Assim, cada categoria foi transformada em colunas binárias para facilitar a análise dos dados.

Para auxiliar na execução dos algoritmos de agrupamento, os dados numéricos foram normalizados, contemplando as colunas: 'Quantidade máquinas', 'Banda Download' e 'Banda Upload'. O *dataset* final ficou com 7.109 objetos. Cada objeto possui 12 atributos.

B. Escolha da técnica utilizada

A partir do *Dataset* final, foram aplicadas três técnicas de agrupamento: K-Means, hierárquico aglomerativo e EM. Para cada uma dessas técnicas de agrupamento foram utilizadas duas técnicas de validação: o índice silhueta e o índice de Davies-Bouldin (DB); calculando para cada índice os resultados de dois a vinte grupos.

A Figura 2 apresenta o resultado do cálculos do índices através do gráfico da silhueta. Conforme comentado anteriormente, no gráfico silhueta, valores próximos a um indicam que os grupos estão mais concisos. Nesse gráfico observa-se que a técnica de agrupamento hierárquico aglomerativo teve um resultado ligeiramente pior que as demais, enquanto que a técnica K-Means teve o melhor resultado por todo o gráfico.

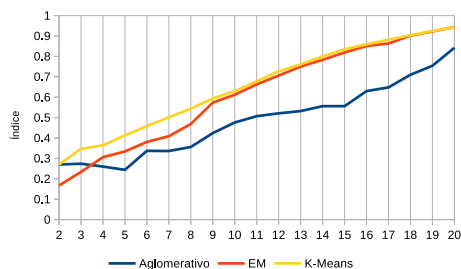


Fig. 2. Gráfico silhueta

A Figura 3 representa o gráfico gerado a partir dos cálculos do índice DB. Conforme comentado anteriormente, no gráfico DB, valores próximos a zero indicam que os grupos estão mais separados e compactos. Nesse gráfico observa-se que a técnica de agrupamento EM teve um resultado ligeiramente pior que as

demais, oscilando bastante; enquanto que a técnica K-Means obteve o melhor resultado por todo o gráfico.

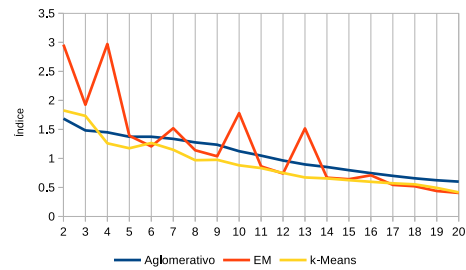


Fig. 3. Gráfico DB

Ainda no gráfico DB, verifica-se que da quinta para a sexta partições forma-se um leve cotovelo. Levando em consideração a quantidade de dados coletados, a eficiência computacional e clareza nos dados agrupados, foi escolhido o agrupamento com cinco partições para esse estudo. A distribuição da quantidade de objetos em cada grupo foi de: 1.516 objetos no Cluster0, 1.311 no Cluster1, 1.481 no Cluster2, o Cluster4 com 1.336 objetos e 1.465 no Cluster4.

V. RESULTADOS

A partir dos dados agrupados foi dado início a etapa de análise dos resultados. Adicionando a coluna com a informação dos agrupamentos no *dataset* inicial, foi possível avaliar os agrupamentos criados: Figura 4 e Figura 5.

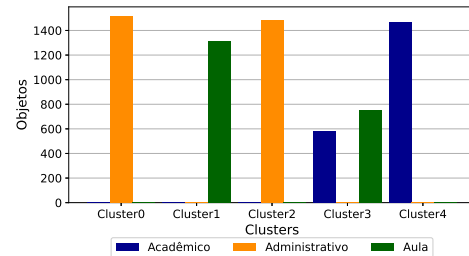


Fig. 4. Cluster por setor

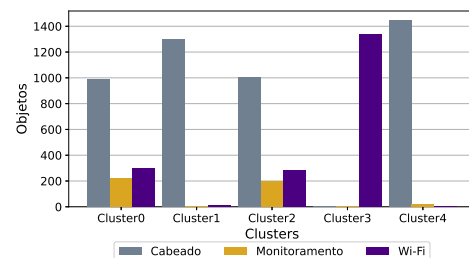


Fig. 5. Clusters por tipo de usuário

Verificou-se uma segmentação por setor, o cluster0 e cluster2 possuem apenas instâncias dos setores administrativos, o cluster1 apenas do setor de aula e o cluster4 apenas dos centros acadêmicos. O cluster 3 possui instâncias dos setores acadêmicos e de aula.

Foi percebida uma segmentação por tipo de usuário, o cluster1 e o cluster4 tiveram em quase sua totalidade máquinas cabeadas, enquanto que o cluster3 possui apenas máquinas conectadas a rede Wi-Fi. Observou-se também que as câmeras ficaram concentradas no cluster0 e cluster2.

Concentrando os dados clusterizados por setor, temos os seguintes gráficos: Figura 6, Figura 7 e Figura 8. Verificou-se que o cluster3 concentrou a maior parte do tráfego de download, seguido pelo cluster4 e cluster2. O cluster0 e o cluster1 tiveram um uso baixo da banda de download, enquanto o tráfego de upload foi concentrado no cluster0 e cluster2, seguido pelo cluster3. O cluster1 teve um uso da banda de upload bem pequeno, se comparado aos demais agrupamentos. O cluster3 concentrou a grande maioria dos equipamentos conectados a rede, seguidos pelo cluster2 e cluster0.

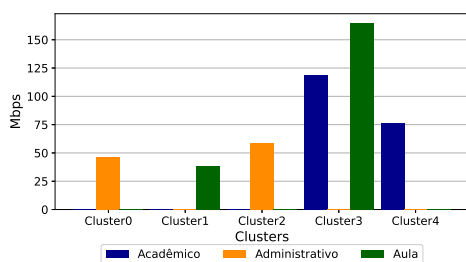


Fig. 6. Cluster por setor e somatório da banda de download

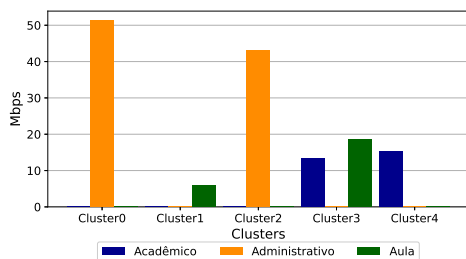


Fig. 7. Cluster por setor e somatório da banda de upload

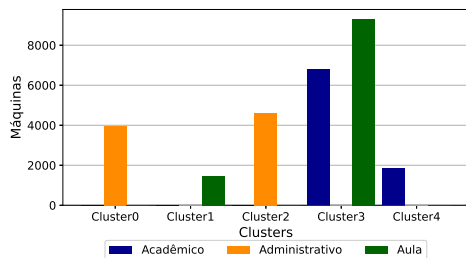


Fig. 8. Clusters por setor e somatório da quantidade de máquinas

Concentrando nos dados por tipo de usuários, temos os seguintes gráficos: Figura 9, Figura 10 e Figura 11. O cluster3 concentrou a maior parte do tráfego de download entre os dados coletados, seguido pelo cluster4, enquanto a banda de upload ficou concentrada entre o cluster0, o cluster2 e o

cluster3, seguido pelo cluster4. O cluster3 concentrou a grande maioria dos equipamentos da rede, seguido pelo cluster2; o cluster1 ficou com poucas máquinas conectadas, findando com um pouco uso tanto da banda de download quanto da banda de upload.

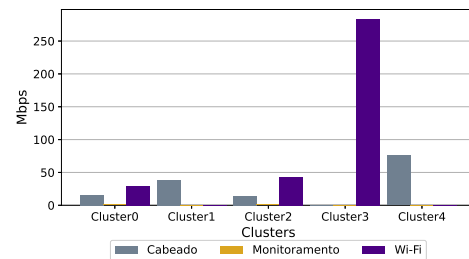


Fig. 9. Tipo de usuário e somatório da banda de download

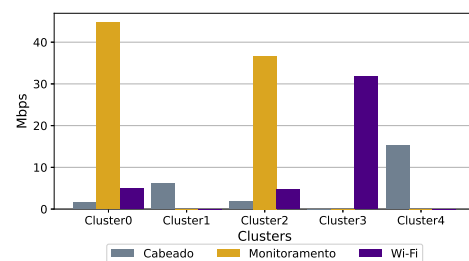


Fig. 10. Tipo de usuário e somatório da banda de upload

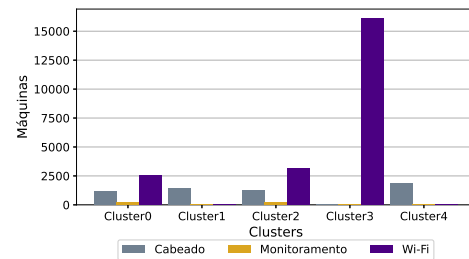


Fig. 11. Clusters por tipo de usuário e somatório da quantidade de máquinas

Concentrando nos dados clusterizados por período, temos os seguintes gráficos: Figura 12, Figura 13 e Figura 14. O cluster3 possui uma utilização bem maior da banda de download, seguido pelo cluster4 e cluster2. O cluster1 agregou máquinas que utilizam muito pouco a rede. O cluster0 e cluster2 utilizam mais a banda de upload da rede, seguido pelo cluster3. Outro detalhe interessante é que o cluster3 possui a maior parte das máquinas conectadas a rede, independente do período.

Os dados coletados neste *dataset* foram fundamentais para avaliar o uso da rede e o comportamento dos usuários em diferentes ambientes acadêmicos. Algumas aferições puderam ser observadas a partir dos dados agrupados:

- O cluster0 e o cluster2 concentraram as máquinas conectadas aos setores administrativos, além da maior parte das máquinas de monitoramento;
- O cluster1 agregou os clientes da rede cabeada dos setores de aula, sendo esses laboratórios, salas de aula e parte

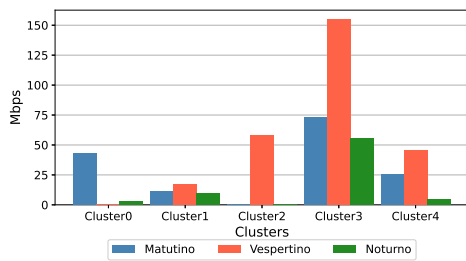


Fig. 12. Cluster por período e somatório da banda de download

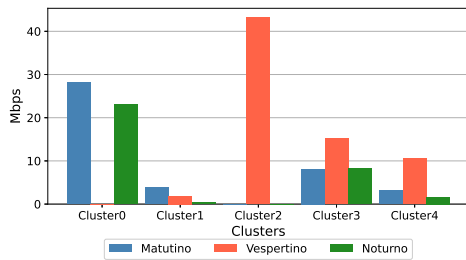


Fig. 13. Cluster por período e somatório da banda de upload

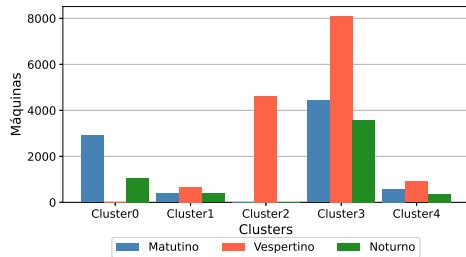


Fig. 14. Clusters por período e somatório da quantidade de máquinas

administrativa dos setores. Nesse *dataset*, essas máquinas utilizaram pouco a rede. É possível que em uma coleta mais duradoura, os resultados sejam diferentes;

- O cluster3 concentrou apenas máquinas conectadas a rede Wi-Fi dos setores de aula e centros acadêmicos, além de possuir a maioria das máquinas conectadas a rede, sendo responsável pela maioria do uso da banda de download, independente do horário;
- O cluster4 concentrou os equipamentos conectados a rede cabeadas dos centros acadêmicos, foram poucas máquinas se comparado com os outros agrupamentos, entretanto, comparando aos outros equipamentos conectados a rede cabeada, tiveram uma utilização mais intensa da rede, tanto em download quanto em upload;
- Independente do período e setor, a maior parte do tráfego de upload foi concentrado nos equipamentos de monitoramento;
- No *dataset* analisado, os equipamentos de monitoramento foram encontrados apenas nos setores administrativos, isso certamente influenciou os resultados;
- Durante o período vespertino foi utilizada a maior parte da banda de download e o período que tivemos a maior quantidade de equipamentos na rede, principalmente Wi-

Fi;

- Foi verificado que o cluster4 possui uma pequena quantidade de máquinas classificadas como monitoramento. Entretanto, essas máquinas aparentam um comportamento diferente das máquinas de monitoramento do cluster0 e cluster2, principalmente no quesito upload. Possivelmente isso indica que são equipamentos que não estão funcionando normalmente, ou são portas dos ativos de rede que estão configuradas de maneira incorreta.

VI. CONCLUSÃO

A avaliação dos dados de uso de rede é fundamental para entender padrões de tráfego, identificar potenciais vulnerabilidades e otimizar a performance da infraestrutura de rede. Este estudo propõe aplicar técnicas de aprendizado de máquina não supervisionado em um conjunto de dados coletados de uso de uma rede acadêmica em operação. O objetivo é identificar relações entre os tipos de tráfego na rede e os diferentes usuários e setores da universidade.

De maneira geral, os dados encontrados foram relevantes, apesar do *dataset* reduzido, com apenas 2 dias de coleta. Durante a análise, foram identificados diversos padrões na utilização da rede. Além disso, foi detectado um possível problema de configuração ou nos equipamentos de monitoramento.

Para dar continuidade a esse trabalho, é interessante continuar esse procedimento de coleta dos dados, seguindo a mesma metodologia aplicada. Uma possibilidade é adicionar ao *dataset* um atributo dia da semana, além de coletar também nos sábados, pelo menos das 08h00 às 12h00.

REFERENCES

- [1] D. Rodrigues, F. Santos, G. Filho, A. Akabane, R. Cabral, R. Immich, W. Junior, F. Cunha, D. Guidoni, T. Silva, D. Rosario, E. Cerqueira, A. Loureiro, and L. Villas, *Computação Urbana da Teoria à Prática: Fundamentos, Aplicações e Desafios*. Minicurso do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC, 2019).
- [2] H. Cesario, G. Girão, A. Riker, B. Dalmazo, and R. Immich, "Arquitetura para gerenciamento de dispositivos através de assistentes virtuais comandados por voz," in *Anais Do VI Workshop De Computação Urbana*, 2022, pp. 1–14.
- [3] M. Fiorenza, D. Kreutz, R. Mansilha, D. Macedo, E. Feitosa, and R. Immich, "Representação e aplicação de políticas de segurança em firewalls de redes híbridas," in *Anais do XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. Porto Alegre, RS, Brasil: SBC, 2021, pp. 490–503. [Online]. Available: <https://sol.sbc.org.br/index.php/sbrc/article/view/16742>
- [4] F. Ceci, *Business Intelligence, Santa Catarina*. Editora PalhoçaUnisulVirtual, 2012.
- [5] F. Ceschin, L. S. Oliveira, and A. Grégio. (2019) Aprendizado de máquina para segurança: Algoritmos e aplicações, sociedade brasileira de computação. Disponível em. [Online]. Available: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/85/373/637-1>
- [6] K. Faceli, A. C. Lorena, J. Gama, and A. de Carvalho, *Inteligência Artificial: Uma abordagem de aprendizagem de máquina*. Rio de Janeiro: Grupo Editorial Nacional, 2011.
- [7] L. F. Palma and A. de, *dados: k-médias*. Universidade Federal do Recôncavo da Bahia Centro de Ciências Exatas e Tecnológicas, 2018.
- [8] H. Macedo, E. Torres, M. Chella, and L. Matos, "Otimização do algoritmo expectation maximization para o modelo de misturas gaussianas usando cuda," *Revista de Sistemas e Computação-RSC*, vol. 7, no. 1, 2017.
- [9] M. V. A. de Araújo et al. (2021) Métodos de clustering em aprendizado de máquinas não supervisionado, niterói. Disponível em. [Online]. Available: <https://app.uff.br/riuff/handle/1/26201>

Estudo comparativo de algoritmos meta-heurísticos na solução de posicionamento de UAV-BS

1st Davi Moraes

Instituto de Tecnologia (ITEC)

Universidade Federal do Pará (UFPA)

Belém-PA, Brasil

davi.moraes@itec.ufpa.br

2nd Wilson Silva

Instituto de Tecnologia (ITEC)

Universidade Federal do Pará (UFPA)

Belém-PA, Brasil

wrss10@gmail.com

3rd Diego Cardoso

Departamento de Engenharia Elétrica

Universidade Federal do Pará (UFPA)

Belém-PA, Brasil

diego@ufpa.br

Abstract—Veículos Aéreos Não Tripulados servindo como Estações Base UAV-BS (*Unmanned Aerial Vehicle Base Stations*) são capazes de restabelecer a conectividade dos usuários em cenários de desastres naturais. O posicionamento de UAV-BS apresenta desafios significativos. É necessário adaptar-se dinamicamente à densidade de usuários e aos requisitos de qualidade de serviço. Propõe-se neste trabalho a avaliação de cinco algoritmos meta-heurísticos (ABC, ACO, AG, PSO e TLBO) para otimizar o posicionamento de UAV-BS, considerando a cobertura dos usuários, qualidade do sinal e custo computacional. Para avaliação do desempenho foram utilizados modelos comparativos que indicaram vantagens do TLBO e PSO.

Index Terms—UAV, Estação base, Algoritmos Meta-heurísticos.

Abstract—Unmanned Aerial Vehicles as Base Stations (UAV-BS) are capable of reestablishing user connectivity in natural disaster scenarios. Optimizing the positioning of UAV-BS presents significant challenges, including identifying the most suitable metaheuristic for implementation. This study proposes the evaluation of five metaheuristic algorithms (ABC, ACO, GA, PSO, and TLBO) to optimize UAV-BS positioning considering user coverage, signal quality and computational cost. Comparative models were used to evaluate the performance of the algorithms indicating superior advantages of TLBO and PSO.

Index Terms—UAV, Base stations, metaheuristic algorithms.

I. INTRODUÇÃO

Veículos Aéreos Não Tripulados UAV (do inglês, *Unmanned Aerial Vehicle*) têm emergido como uma solução promissora devido à sua flexibilidade e capacidade de implantação rápida, atuando como Estações Base BS (do inglês, *Base Station*) para restabelecer a conectividade em cenários de desastres naturais. A otimização do posicionamento de UAV-BS, no entanto, apresenta desafios significativos. É necessário adaptar-se dinamicamente às mudanças nos cenários, à densidade de usuários e aos requisitos de qualidade de serviço (QoS), contribuindo de forma relevante para redes móveis modernas e futuras.

Algoritmos meta-heurísticos são propostos na literatura para abordar o posicionamento de UAV-BS, incluindo a otimização por Colônia de Abelhas ABC (do inglês *Artificial Bee Colony Algorithm*), a Otimização por Colônia de Formigas ACO (do inglês *Ant Colony Optimization Algorithm*), o Algoritmo Genético AG, a Otimização por Enxame de Partículas PSO (do inglês *Particle Swarm Optimization*) e

a Otimização Baseada em Ensino-Aprendizagem TLBO (do inglês *Teaching-Learning Based Optimization*).

Apesar dos avanços significativos no uso de algoritmos meta-heurísticos para otimização do posicionamento do UAV-BS, ainda há lacunas na literatura quanto à escolha da meta-heurística mais adequada. Esta pesquisa visa preencher essa lacuna, comparando o desempenho de diferentes meta-heurísticas para identificar a mais eficiente.

Nesse sentido, este estudo se propõe a implementar e avaliar métodos de otimização para o posicionamento de UAV-BS. Especificamente, visa: (1) Comparar e avaliar diferentes algoritmos meta-heurísticos (AG, PSO, ACO, ABC e TLBO) para determinar a posição ótima ou sub-ótima dos UAV-BS; (2) Analisar a correlação entre diferentes métricas de desempenho, como quantidade de usuários conectados, vazão e tempo de execução; (3) Identificar a meta-heurística mais eficiente em termos de qualidade de serviço e custo computacional, fornecendo recomendações práticas para a implementação.

A principal contribuição deste artigo é fornecer uma análise comparativa do desempenho de cinco meta-heurísticas na otimização do posicionamento de UAV-BS. Este estudo amplia o conhecimento teórico sobre algoritmos de otimização, destacando a importância de selecionar o algoritmo adequado para maximizar a cobertura de usuários e a qualidade do sinal, minimizando o tempo de resposta e o custo computacional.

O restante do documento está organizado da seguinte forma: a Seção II apresenta os trabalhos correlatos. A Seção III descreve metodologia utilizada. A Seção IV discute os resultados. Finalmente, a Seção V conclui o trabalho.

II. TRABALHOS CORRELATOS

O posicionamento eficaz dos UAVs para maximizar a cobertura de usuários é um desafio complexo abordado por diversos estudos. Em [1], foi proposto um algoritmo híbrido para a colocação de UAVs em cidades inteligentes, destacando a eficácia das abordagens meta-heurísticas. Já [2] introduziu um algoritmo baseado em meta-heurística para maximizar a cobertura com o número mínimo de UAVs.

Em [3] os autores demonstraram a relevância no uso do PSO propondo um algoritmo para otimizar a cobertura de estações base em UAVs, enquanto [4] focou na maximização da cobertura do usuário através da otimização conjunta do

posicionamento do UAV e do fator de compensação de perda de percurso.

Para a seleção de locais dos UAV-BS, [5] propuseram um método baseado no algoritmo espiral, integrando ACO para o planejamento ideal. Adicionalmente, [6] demonstraram a aplicabilidade do ACO em domínios contínuos para problemas de otimização.

Em [7], foi apresentada a utilização do ACO para otimizar o caminho em computação de borda móvel, enquanto [8] empregaram o ACO para minimizar o custo do caminho dos UAVs em Redes de Sensores sem Fio WSN (do inglês, *Wireless Sensor Networks*).

A eficiência dos algoritmos genéticos foi destacada em [9], que otimizaram parâmetros para um controlador adaptativo direto no controle de atitude de UAV-BS. Em [10], foi apresentado um método para maximizar a satisfação do usuário através de algoritmos genéticos e PSO para posicionar múltiplos UAV-BS.

O algoritmo ABC foi utilizado no posicionamento de UAVs como estações base em [12], propondo um método híbrido de otimização multiobjetivo para redes 5G. Em [13], foi apresentado um método eficiente para reconfigurar formações de UAVs com ABC, considerando restrições como prevenção de colisões e estabilidade da formação.

O TLBO foi aplicado no posicionamento de UAVs em [14], otimizando tarefas de emergência de múltiplos UAVs, melhorando a eficiência da convergência e garantindo a conclusão eficiente das missões.

III. METODOLOGIA

Esta seção fornece uma descrição detalhada do ambiente simulado no qual as soluções de meta-heurísticas foram avaliadas, bem como as métricas utilizadas e a forma de coleta.

A. Ambiente de Simulação

A pesquisa é conduzida em um ambiente de simulação que modela a propagação do sinal e a interferência em uma área de 1000m x 1000m, considerando uma frequência de 2.4GHz. Para a construção do cenário, foi utilizado Python, na versão 3.2.12. Para análise e visualização dos dados, foram utilizadas as versões mais atualizadas das bibliotecas NumPy (1.26.0) e Matplotlib (3.9).

No cenário construído os usuários são distribuídos de forma aleatória. Entretanto para uma análise comparativa entre os diferentes tipos de algoritmos meta-heurísticos, é atribuído uma semente na aleatoriedade da posição dos usuários, utilizando o módulo Random do Python, com o objetivo de criar cenários de busca idênticos para todos os algoritmos.

B. Planejamento de posicionamento

Ao planejar o posicionamento de UAV-BS, a propagação eletromagnética apresenta vários desafios, que deve considerar a comunicação entre UAV-BS e os usuários terrestres levando em conta alguns conceitos e fatores que afetam essa comunicação.

Para calcular a distância dos usuários até os UAV-BS é utilizada a equação da distância euclidiana em um plano bidimensional, amplamente compreendida, que calcula a distância entre dois pontos (x, y) e (x_i, y_i) [15].

C. Modelo de Propagação

Foi utilizado o modelo de perdas no caminho *Air-To-Ground* (A2G) [16], que leva em consideração L_{LoS} como a linha de visão (LoS) e η_{LoS} como a não linha de visão (NLoS) entre o UAV-BS e o usuário no solo. Onde LoS e NLoS são perdas de propagação no espaço livre e dependem do ambiente, conforme [17]. O modelo é expresso como:

$$PL_{LoS}(dB) = 20 \log_{10} \left(\frac{4\pi f_c d_{ij}}{c} \right) + \eta_{LoS} \quad (1)$$

onde: f_c é uma variável que representa a frequência portadora, d_{ij} é a distância em metros, e c é uma constante que representa a velocidade da luz [16].

D. Relação Sinal-Interferência mais Ruído

Assumimos que a Relação Sinal-Interferência mais Ruído (SINR, do inglês *Signal-to-Interference-Plus-Noise Ratio*) é utilizada para fornecer limites superiores teóricos para a capacidade do canal em sistemas de comunicação sem fio. A SINR representa o cenário onde o ruído de fundo e a intensidade da interferência de outras transmissões simultâneas também são consideradas.

Neste experimento consideramos que, durante o período de tempo em que o UAV-BS está transmitindo dados aos usuários terrestres, o UAV-BS mantém uma altitude e velocidade constante. Assumimos que possui menos obstáculos para que o UAV-BS opere eficientemente em alturas menores, 50 metros, aproveitando melhor a visibilidade direta [18]. Estimamos que cada nó transmite com a mesma potência, para permitir comparações diretas entre os diferentes algoritmos, que é o foco do trabalho. A fórmula para o cálculo do SINR [19]

$$SINR = \frac{P}{I + N} \quad (2)$$

onde:

- P é a potência do sinal de interesse.
- I é a potência de interferência de outros sinais na rede.
- N é a potência do ruído.

E. Cálculo de Shannon

Para calcular a taxa de transmissão ou rendimento da rede, que indica a taxa de bits transmitidos em um intervalo de tempo, é utilizada a fórmula de Shannon [20]. A taxa de usuário alcançável é expressa como:

$$C = B \log_2 \left(1 + \frac{P_u G}{L N_0 B} \right) \quad (3)$$

onde B é a largura de banda alocada por usuário, P_u é a potência transmitida pelo UAV-BS, G é o ganho da antena direcional, L representa a perda de percurso definida em (1) e N_0 é a densidade espectral de potência do ruído [21].

F. Avaliação Comparativa dos Algoritmos

Para o cenário de avaliação, foi utilizado um computador equipado com um processador Intel(R) Core(TM) i5-8350U CPU 1.70GHz - 1.90GHz, com 8GB de memória RAM e uma placa de vídeo Intel (R) UHD Graphics 620.

Foram testadas variações de 2 a 6 UAV-BS nos experimentos, executando cada algoritmo 30 vezes para cada configuração. Essa abordagem permitiu avaliar a convergência e a capacidade de resiliência dos algoritmos em diferentes condições.

1) *Teste de Significância Estatística*: Para avaliar o desempenho de diferentes algoritmos foi utilizado o teste de significância estatística, especificamente a Análise de Variância [22] que tem como objetivo verificar a existência de diferenças significativas entre os algoritmos para cada métrica avaliada: quantidade de Usuários Conectados UC, SINR, Vazão e Tempo de Execução TE.

2) *Cálculo de superioridade*: A fórmula utilizada para calcular a superioridade percentual de uma meta-heurística sobre a outra, é dada por:

$$S(\%) = \left(\frac{\theta - \iota}{\iota} \right) \times 100 \quad (4)$$

onde: $S(\%)$ é a superioridade de um algoritmo em relação ao outro, medida em percentual, θ é o Valor da métrica de uma Meta-Heurística; ι é Valor da métrica de uma Meta-Heurística Comparada.

G. Algoritmos de Meta-Heurísticos

Neste estudo, buscou-se algoritmos inspirados em diferentes fenômenos naturais (abelhas, formigas, evolução genética, comportamento social de partículas, e processos de ensino-aprendizagem), garantindo uma diversidade de estratégias e mecanismos de solução que cobrem amplamente o espectro de possíveis soluções. Ao final de cada meta-heurística busca-se otimizar o posicionamentos dos UAV-BS de forma a maximizar a cobertura e a qualidade do sinal para os usuários.

Os algoritmos meta-heurísticos utilizados foram: O ABC que simula o comportamento forrageiro das abelhas para encontrar soluções ótimas, possuindo uma estrutura simples, com poucos parâmetros para ajuste, o que facilita sua implementação e reduz a carga computacional [23]. O ACO que baseia-se no comportamento de procura de caminho das formigas que utilizam o feromônio para otimização de planejamento de caminho, excelente para resolver problemas combinatórios e possui adaptabilidade a mudanças no ambiente, crucial para cenários onde a topologia e a demanda de rede podem variar [24]. O AG que utiliza conceitos de seleção natural e genética para evoluir soluções ao longo de várias gerações, atuando com flexibilidade e robustez em explorar grandes espaços de busca, sendo eficaz em evitar mínimos locais através de operações genéticas [25]. O PSO que inspira-se no comportamento social de enxames como os pássaros e peixes para encontrar soluções ótimas, explorando o espaço de busca globalmente, usando uma população de partículas que ajustam suas posições com base em suas próprias experiências

e nas dos vizinhos [26]. O TLBO que é um algoritmo baseado no processo de ensino e aprendizagem em sala de aula onde os "alunos" (soluções) aprendem do "professor" (melhor solução) e entre si, balanceando a exploração sem necessidade de muitos parâmetros ajustáveis, o que facilita sua implementação e uso [27].

A Tabela I apresenta os parâmetros da aplicação dos algoritmos para resolver o problema.

TABLE I
PARÂMETROS DOS ALGORITMOS DE META-HEURÍSTICA

Algoritmo	Parâmetro	Valor
ABC	Máximo de iterações	100
	Limite	200
ACO	Taxa de evaporação	0.1
	Feromônios (α)	1
	Informação heurística (β)	1
	Depósito de feromônio	1
	Feromônio inicial	0.5
	Máximo de iterações	100
AG	Número de gerações	250
	Tamanho da população	250
	Taxa de mutação	0.1
PSO	Número de partículas	200
	Máximo de iterações	10
	Peso da Inércia	0.5
	Peso Cognitivo	0.5
	Peso Social	0.8
TLBO	Máximo de iterações	400
	Taxa de aprendizado	1.0

Os parâmetros gerais do cenário são apresentados na Tabela II.

TABLE II
TABELA DE PARÂMETROS DA SIMULAÇÃO.

Parâmetro	Valor
Número de Usuários	100
Número de Pontos UAV-BS	2 - 6
Tamanho do espaço de busca	1000m x 1000m
Potência do Sinal	-80dBm
Potência de Ruído	38dBm
Frequência	2.4GHz
SINR Desejado	25dB
Altura do UAV-BS	50m
Altura antena receptora	1,2m

H. Função Fitness

A função de fitness deste problema avalia quão bem os pontos de acesso (UAV-BS) estão distribuídos para maximizar o número de usuários conectados com um SINR desejado. É definida como:

$$\text{Fitness} = \sum_i \begin{cases} 1 & \text{se } \text{SINR}_i \geq \text{SINR Objetivo} \\ 0 & \text{caso contrario} \end{cases} \quad (5)$$

onde i é o usuário avaliado.

Esta formulação contabiliza o número de usuários cujo SINR do usuário avaliado é maior que o SINR objetivo.

I. Procedimento de Coleta de Dados

1. Inicialização das posições aleatórias dos UAV-BS e usuários no espaço de busca;
2. Cálculo da SINR para cada par UAV-BS/usuário;
3. Execução dos algoritmos de meta-heurísticas para otimizar as posições dos UAV-BS;
4. Registro das posições ótimas ou subótimas encontradas, percentual de usuários conectados, SINR alcançado, vazão (Mbps) e custo computacional (tempo de execução do algoritmo).

IV. RESULTADOS

Na Tabela III são apresentados os resultados das estatísticas F e valor P estabelecidos no teste de significância estatística:

TABLE III
RESULTADOS DA SIGNIFICÂNCIA ESTATÍSTICA

Métrica	Estatística F	Valor P
Usuários Conectados	14,41	$2,46 \times 10^{-11}$
SINR	1585,89	0,0
Vazão	159,11	$2,21 \times 10^{-98}$
Tempo de execução	85,13	$1,49 \times 10^{-59}$

Os testes de significância estatística demonstraram que o Valor P, iguais ou próximos a zero, confirmam a hipótese de que a escolha do algoritmo impacta diretamente nos resultados, pois quanto mais próximo de zero, maior é a significância estatística da avaliação.

Na análise dos dados, o desempenho dos diferentes algoritmos de otimização é representado na Figura 1, demonstrando o desempenho quanto à conexão dos usuários para as diferentes quantidades de UAV-BS.

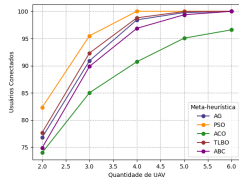


Fig. 1. Desempenho dos Algoritmos para Usuários Conectados por UAV-BS

Revelou-se que a quantidade de usuários conectados, obtidos por meio da métrica fitness, varia significativamente entre os algoritmos, com o TLBO e o PSO frequentemente conectando mais usuários em comparação aos outros algoritmos.

Percebe-se que o algoritmo PSO apresenta melhores resultados na busca por soluções ótimas quando se trata de usuários conectados, pois com 4 UAV-BS atinge a conexão máxima de 100% dos usuários e mantendo-se estável. Os algoritmos AG, ABC e TLBO mostram uma melhoria consistente no número de usuários conectados à medida que a quantidade de UAV-BS aumenta, porém alcançam a conexão total somente com 6 UAV-BS, apesar de que com 5 UAV-BS já indicam um alto nível de eficiência. Todavia o algoritmo ACO não converge

muito bem na cobertura dos usuários, apresentando o menor desempenho dentre os algoritmos analisados.

A análise comparativa de superioridade de um algoritmo em relação aos outros algoritmos, considerando 4 UAV-BS, está apresentada na matriz comparativa da Figura 2:

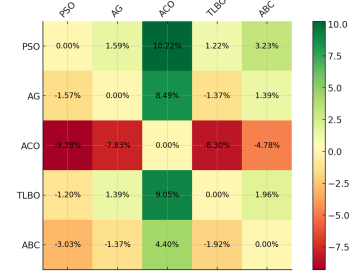


Fig. 2. Matriz comparativa de superioridade dos algoritmos com 4 UAV-BS.

O eixo X e Y representam respectivamente θ e ι no cálculo de S(%), onde o PSO apresenta uma superioridade de 10,22% frente o algoritmo ACO.

No gráfico da Figura 3 a vazão foi consistentemente alta para todos os algoritmos, com valores em torno de 97 a 100 Mbps. O AG e o ABC, destacaram-se com vazões ligeiramente superiores, especialmente com 6 UAV-BS. Percebe-se que o PSO, no momento que tem 4 UAV-BS, onde consegue atingir a cobertura de todos os usuários, seus resultados possuem uma menor variância com relação a média, ou seja, os resultados estão quase na mesma vazão e buscando um percentil menos ampliado. Entretanto, quando se analisa com 5 UAV-BS, a variância dos resultados da vazão do PSO aumenta, levando a identificar que o aumento de UAV-BS no ambiente gera uma maior interferência de sinal, sem gerar melhoria significativa nas métricas analisadas. Enquanto que o TLBO atinge seu melhor desempenho na vazão quando consegue cobrir todos os seus usuários com 5 UAV-BS, diminuindo sua variância na vazão, assumindo o mesmo comportamento do PSO.

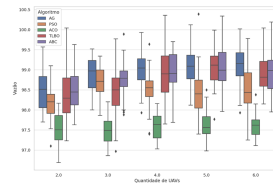


Fig. 3. Desempenho de cada algoritmo em termos de vazão

O tempo de execução aumentou conforme estava aumentando a quantidade de UAV-BS, entretanto a partir do momento em que os algoritmos PSO, TLBO e ABC começaram a atingir o ponto de parada na cobertura de todos os usuários, conforme apresentado na Figura 1, ocorreu a diminuição do tempo de execução. Os algoritmos AG e ACO não conseguiram ter a mesma convergência e continuaram a aumentar o tempo de execução com o aumento da quantidade de UAV-BS. O TLBO e o PSO mostraram tempos de execução relativamente baixos em comparação com os outros algoritmos, sugerindo uma eficiência computacional superior.

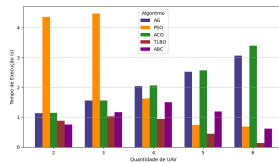


Fig. 4. Desempenho dos algoritmos em razão do tempo por UAV-BS

Os resultados apontam para as seguintes características:

AG: Desempenho sólido em vazão, com perca no desempenho na métrica de tempo de execução.

ABC: Melhor escolha para qualidade de sinal (SINR) e vazão, com boas médias de usuários conectados.

ACO: Desempenho equilibrado mas com maior variabilidade e tempos de execução mais altos.

PSO: Bom desempenho geral, mas com maior variabilidade e tempos de execução mais altos.

TLBO: Melhor escolha para um desempenho geral consistente, com excelente média em todas as métricas e tempos de execução rápidos.

V. CONCLUSÕES

Neste artigo, foram comparados cinco algoritmos meta-heurísticos para otimizar o posicionamento de UAV-BS, comparando seus desempenhos na cobertura de usuários distribuídos no espaço de busca. Os resultados experimentais indicaram que o TLBO e o PSO superaram os demais algoritmos nas principais métricas de velocidade de convergência e cobertura dos usuários, com TLBO demonstrando uma velocidade de convergência de 0,64s enquanto que o PSO tem velocidade média de 2,36s. Entretanto o PSO alcança uma cobertura completa dos usuários com 4 UAV-BS, com superioridade de 2% em relação ao TLBO e 8% em relação ao ACO. As análises mostraram que TLBO e PSO competem de forma próxima em termos de qualidade de serviço e custo computacional. Estes achados contribuem para a generalização das abordagens de posicionamento de UAV-BS garantindo uma diversidade de estratégias e mecanismos de soluções, bem como abrem caminho para futuras pesquisas, incluindo a aplicação de TLBO e PSO em problemas otimização multiobjetivo.

REFERENCES

- [1] Saadi, Amylia Ait, et al. "A hybrid improved manta ray foraging optimization with Tabu search algorithm for solving the UAV placement problem in smart cities." *IEEE Access*, vol. 11, pp. 24315–24342, 2023.
- [2] T. Kimura and M. Ogura, "Distributed collaborative 3D-deployment of UAV base stations for on-demand coverage," *IEEE INFOCOM 2020-IEEE conference on computer communications*, pp. 1748–1757, 2020.
- [3] X. Chen, W. Chen, X. Tang, L. Yang, and L. Zhou, "PSO-VFA: A hybrid intelligent algorithm for coverage optimization of UAV-mounted base stations," *Internet Technology Journal*, vol. 23, no. 3, pp. 487–495, 2022.
- [4] N. Safwat, I. Hafez, and F. Newagy, "3D placement of a new tethered UAV to UAV relay system for coverage maximization," *Electronics*, vol. 11, no. 3, pp. 385, 2022.
- [5] Bian, Q., Xu, D., Kang, K., & Liu, J. Uav base station site selection based on spiral algorithm in complex environment. *Journal of Physics Conference Series*, vol. 2185, no. 1, pp. 12080, 2022.
- [6] K. Socha and M. Dorigo, "Ant colony optimization for continuous domains," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1155–1173, 2008.
- [7] M. Mousa and M. Hussein, "Efficient UAV-based mobile edge computing using differential evolution and ant colony optimization," *PeerJ Computer Science*, vol. 8, e870, 2022.
- [8] Li, Y.; Meng, X.; Ye, F.; Jiang, T.; Li, Y. "Path Planning Based on Clustering and Improved ACO in UAV-assisted Wireless Sensor Network". *IEEE USNC-CNC-URSI North American Radio Science Meeting (Joint with AP-S Symposium)*, pp.57–58, 2020.
- [9] Carvalho, L. J. C., Prates, M. d. O., & Campos, V. C. d. S. "Utilização de algoritmo genético para estimação e otimização de parâmetros de um controlador adaptativo direto usado no controle da atitude de um drone quadrotor." *Congresso Brasileiro De Inteligência Computacional*, 2021.
- [10] Haghighi, Hassan, , Sadati, SH, Dehghan, SM e Karimi, J. "Hybrid form of particle swarm optimization and genetic algorithm for optimal path planning in coverage mission by cooperated unmanned aerial vehicles." *Journal of Aerospace Technology and Management*, vol. 12, pp. e4320, 2020
- [11] Shahin, H., Shaaban, M. F., Ismail, M. H., Mourad, H. A. M., & Khattab, A. "A Multi-Objective Situational Awareness Approach for Distribution Networks Using Drones." *IEEE Access*, 2023.
- [12] Zeng, Qingwen. "Optimization of Millimeter-Wave Base Station Deployment in 5G Networks." *Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2022.
- [13] Yang, Zipeng, et al. "Reconfiguration for UAV Formation: A Novel Method Based on Modified Artificial Bee Colony Algorithm." *Drones*, vol. 7, no. 10, pp.595, 2023.
- [14] F. Oliveira, M. Luis, and S. Sargento, "Machine learning for the dynamic positioning of UAVs for extended connectivity," *Sensors*, vol. 21, no. 13, pp. 4618, 2021.
- [15] KHAN, Shah Khalid. "Mathematical framework for 5G-UAV relay." *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 3, pp. e4194, 2021.
- [16] Yan, Chaoping, et al. "A comprehensive survey on UAV communication channel modeling." *IEEE Access*, vol. 7, pp. 107769–107792, 2019.
- [17] Zhao, K.; Razaq, M.M.; Li, K.; Peng, L.; Ho, P.-H. "3D Deployment of UAVs for Communications under Multiple Eavesdroppers." *In Proceedings of the 2022 International Conference on Networking and Network Applications (NaNA)*, IEEE pp. 437–441, 2022.
- [18] Shoer, I., Gunturk, B. K., Ates, H. F., & Baykas, T. "Altitude Optimization of UAV Base Stations from Satellite Images Using Deep Neural Network." *arXiv preprint arXiv:2112.14551*, 2021.
- [19] Jiang, Feng, and A. Lee Swindlehurst. "Optimization of UAV heading for the ground-to-air uplink." *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 5, pp. 993–1005, 2012.
- [20] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
- [21] GUO, Zinan; HU, Bo; CHEN, Shanzhi. "A Decision for Throughput Optimization in UAV-Enabled Emergency Outdoor-Indoor Fairness Communication." *Drones*, vol. 7, no. 7, pp. 460, 2023.
- [22] Terenciani, Marcelo Figueiredo, et al. "Desenvolvimento de uma plataforma web para sensoriamento remoto com VANT", 2023.
- [23] Tan, Li, et al. "Multi-UAV path planning based on IB-ABC with restricted planned arrival sequence." *Robotica*, vol. 41 no. 4, pp. 1244–1257, 2023.
- [24] Konatowski, Stanisław, and Piotr Pawłowski. "Ant colony optimization algorithm for UAV path planning." *14th international conference on advanced trends in radioelectronics, telecommunications and computer engineering (TCSET)*. IEEE, pp. 177–182, 2018.
- [25] Anosri, Siwakorn, et al. "A comparative study of state-of-the-art meta-heuristics for solving many-objective optimization problems of fixed wing unmanned aerial vehicle conceptual design." *Archives of Computational Methods in Engineering*, vol. 30, no. 6, pp. 3657–3671, 2023.
- [26] Sánchez-García, Jesús, Daniel Gutierrez Reina, and S. L. Toral. "A distributed PSO-based exploration algorithm for a UAV network assisting a disaster scenario." *Future Generation Computer Systems*, vol. 90, pp. 129–148, 2019.
- [27] Ghambari, Soheila, et al. "A comparative study of meta-heuristic algorithms for solving UAV path planning." *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 174–181, 2018.

Sensor Virtual de Pressão Baseado em Rede Neural de Memória de Curto e Longo Prazo

Luiz Felipe Rodrigues Gonçalves
Departamento de Engenharia Elétrica - DEE
Universidade Federal da Paraíba - UFPB
João Pessoa, Brasil
luiz.goncalves@estudante.cear.ufpb.br

Lucas Dantas de Oliveira
Departamento de Engenharia Elétrica - DEE
Universidade Federal da Paraíba - UFPB
João Pessoa, Brasil
lucasd.oliveira@estudante.cear.ufpb.br

Petronio Ferreira de Melo
Companhia de Água e Esgotos da Paraíba-
CAGEPA
João Pessoa, Brasil
petronioferreira@cagepa.pb.gov.br

Juan Moises Mauricio Villanueva
Departamento de Engenharia Elétrica - DEE
Universidade Federal da Paraíba - UFPB
João Pessoa, Brasil
jmauricio@cear.ufpb.br

Fernanda Maria Lima Fernandes
Departamento de Engenharia Elétrica - DEE
Universidade Federal da Paraíba - UFPB
João Pessoa, Brasil
fernanda.fernandes@estudante.cear.ufpb.br

Resumo—Este artigo apresenta o desenvolvimento de um sensor virtual de pressão baseado no modelo de rede neural de memória de curto e longo prazo a partir de dados de vazão e nível de uma rede de abastecimento de água localizada no interior da Paraíba. O objetivo é melhorar os índices de macromedição e reduzir perdas de água. Os resultados mostram que o sensor virtual apresenta um desempenho comparável aos sensores físicos, com uma margem de erro de 3,67%.

Palavras-Chave—Sensor Virtual, Memória de Curto e Longo Prazo, Macromedição.

I. INTRODUÇÃO

A Lei nº 14.026/2020, que instituiu o novo marco legal do saneamento, gerou um impacto significativo nas empresas de distribuição de água e esgoto ao estabelecer um novo contexto regulatório e competitivo para o setor. Um dos focos do marco legal é a redução do índice de perdas de água e elevação dos índices de macromedição. Segundo o Sistema Nacional de Informações sobre Saneamento (SNIS) o índice médio de perdas das companhias é de aproximadamente 40%, enquanto que apenas 70% de toda a água é macromedida.

No entanto, é comum que as companhias de água foquem o monitoramento em regiões metropolitanas devido ao alto custo dos sensores, sobretudo macromedidores de vazão. A medição indireta se torna uma alternativa, pois permite que a vazão seja calculada a partir de valores de pressão da rede de abastecimento ou nível do reservatório [1].

Este artigo descreve um sensor virtual para medição indireta de pressão de distribuição baseado em rede neural de memória de curto e longo prazo (LSTM - *Long Short Term Memory*) a partir de dados de vazão e nível de um reservatório adquiridos pela Companhia de Água e Esgotos da Paraíba

(CAGEPA) na cidade de Salgado de São Félix através de um *broker* (intermediário para comunicação de dispositivos) de protocolo MQTT e monitoramento em nuvem. O artigo está organizado da seguinte forma: A Seção II introduz o conceito de sensor virtual e a arquitetura LSTM. A Seção III apresenta a metodologia utilizada no desenvolvimento do sensor. A Seção IV apresenta os resultados obtidos e o sistema monitorado em tempo real. Por fim, a Seção V apresenta as conclusões.

II. REDE NEURAL DE MEMÓRIA DE CURTO E LONGO PRAZO

Um sensor virtual (*Soft Sensor*) é um modelo computacional que simula, em um ambiente virtual, o funcionamento de um sensor real. Existem diversas metodologias para sua criação, sendo a mais comum a utilização de redes neurais artificiais [2].

As redes neurais artificiais têm sido amplamente utilizadas para a previsão de séries temporais devido à sua capacidade de reconhecer padrões nos dados [3].

Uma variação da rede neural recorrente é o modelo LSTM (Long Short-Term Memory). Este modelo se baseia no processo de esquecimento da mente humana para otimizar a utilização dos dados na previsão de novos valores. Assim, os dados recentes (memória de curto prazo) são avaliados com base nas características previamente apresentadas (memória de longo prazo), permitindo que a célula LSTM esqueça características pouco influentes ou desnecessárias [4].

A Figura 1 ilustra a estrutura de uma célula LSTM.

A célula é dividida da seguinte forma [4]: a porta de esquecimento (*forget gate*) determina que informações do estado anterior S_{t-1} devem ser esquecidas. O valor de f_t é

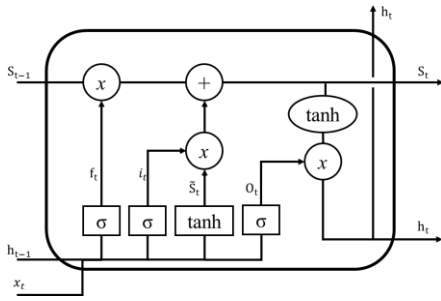


Figura 1. Estrutura da célula LSTM.

calculado pela função sigmoide a partir do valor da entrada atual x_t , o estado oculto anterior h_{t-1} e um *bias* b .

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (1)$$

A porta de atualização i_t , ou entrada (*input gate*), decide quais novas informações serão armazenadas no estado da célula. O estado $\tilde{S}_t$ é chamado de estado candidato e calculado pela função tangente hiperbólica, sendo limitado entre -1 e 1.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (2)$$

$$\tilde{S}_t = \tanh(W_{xs}x_t + W_{hs}h_{t-1} + b_s) \quad (3)$$

O estado atual é definido através do produto de Hadamard entre a porta de esquecimento e o estado candidato.

$$S_t = f_t \odot S_{t-1} + i_t \odot \tilde{S}_t \quad (4)$$

A porta de saída (*output gate*) o_t decide qual parte do estado da célula será utilizado como próximo estado oculto h_t

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(S_t) \quad (6)$$

Em que W_x e W_h representam os pesos de cada entrada e estado para os cálculos das portas de esquecimento, entrada, estado atual e saída.

III. METODOLOGIA

O sensor virtual foi treinado com dados preexistentes fornecidos pela CAGEPA. A Figura 2 ilustra a disposição dos sensores em relação à cidade. Os dados são aquisitados a cada 6 minutos através de placas ESP-8266, foi utilizada uma base de dados de aproximadamente 1 ano.

A primeira etapa consistiu no pré-processamento para correção de valores espúrios decorrentes de perda de comunicação. A partir de uma análise de correlação entre as variáveis, foram escolhidos como parâmetros de entrada os valores de vazão, provenientes de um macromedidor ultrassônico, e de nível, obtidos de uma sonda hidroestática em uma estação elevatória localizada fora da cidade. Para a saída,

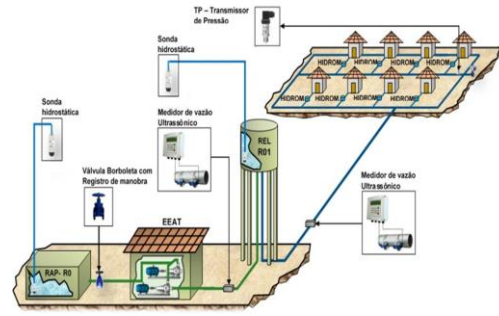


Figura 2. Disposição dos sensores físicos para validação.

foram empregados dados de um sensor de pressão localizado na cidade.

Após as etapas de treinamento e teste da rede neural, o sensor foi implantado em uma máquina virtual da AWS (*Amazon Web Services*).

Para a validação do sistema, os resultados foram comparados com outras topologias como perceptron multicamadas (MLP - *Multilayer Perceptron*) e rede neural convolucional (CNN - *Convolutional Neural Network*) com diferentes quantidades de neurônios. A CAGEPA disponibilizou acesso a um *broker* MQTT, que transmite continuamente os dados dos sensores para o sistema SCADA-LTS. Dessa forma, foi possível avaliar a eficiência de previsão do LSTM em relação a novos dados.

IV. RESULTADOS

A Figura 3 ilustra os resultados obtidos. Os valores são medidos em metros de coluna de água (MCA).



Figura 3. Apresentação dos dados do sensor real e do sensor virtual na plataforma Grafana.

A Tabela I ilustra as taxas de erro médio quadrático (MSE - *Mean Squared Error*) e erro médio absoluto percentual (MAPE - *Mean Absolute Percentage Error*) entre as diferentes topologias utilizadas.

O modelo LSTM apresenta um erro de 3,67%, o menor quando comparado com os outros modelos, o que demonstra sua eficiência na previsão de dados de séries temporais.

Tabela I
ERRO ENCONTRADO NAS TOPOLOGIAS

Topologia	MSE	MAPE (%)
MLP	1,51	6,36
CNN	1,75	5,70
LSTM	0,92	3,67

V. CONCLUSÃO

Este artigo apresentou a construção de um sensor virtual de pressão baseado em uma topologia LSTM, desenvolvido em parceria com a empresa CAGEPA. O sensor apresentou um erro de 3,67% em relação aos dados do sensor real. Este erro mostrou-se menor quando comparado com outras topologias como MLP e CNN. Atualmente, ele está implantado em uma máquina virtual da AWS, recebendo novos dados de entrada via protocolo MQTT.

Para a continuidade do trabalho, recomenda-se avaliar o desempenho do sensor em outras localidades e utilizando como entrada apenas os dados de nível, uma vez que este é frequentemente o único parâmetro disponível nas regiões mais afastadas dos centros urbanos.

REFERÊNCIAS

- [1] T. K. S. Flores et al. "Indirect Feedback Measurement of Flow in a Water Pumping Network Employing Artificial Intelligence". *Sensors* 2021, 21, 75.
- [2] R. P. G. Lima et al. "Development of a Soft Sensor for Flow Estimation in Water Supply Systems Using Artificial Neural Networks". *Sensors* 2022, 22, 3084.
- [3] R. Palmitessa et al. "Soft sensing of water depth in combined sewers using LSTM neural networks with missing observations". *Journal of Hydro-Environment Research* 38 (2021): 106-116.
- [4] Yanlin He et al. "An Improved Industrial Process Soft Sensor Based on LSTM". *2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE, 2023. p. 1750-1755.

Analysis of LoRa Modulation Performance Under Epidemic Interference Conditions

Felipe Pinheiro Correia
Computing Coordination - COINFO
IFSERTAO-PE
Petrópolis-PE, Brazil
felipe.correia@ifsertao-pe.edu.br

Marcelo Sampaio de Alencar
Communication Systems Department
UFRN
Natal-RN, Brazil
malencar@iecom.org.br

Abstract—This study evaluates the performance of LoRa modulation in digital communication systems by analyzing the Bit Error Rate (BER) versus Signal-to-Noise Ratio (SNR) under various channel conditions, including additive white Gaussian noise (AWGN), Rayleigh fading, and epidemic interference characterized by a lognormal distribution. The methodology involves generating and transmitting signals through simulated channels, then processing the received signals with a down-chirp signal and Discrete Fourier Transform (DFT) to estimate transmitted symbols. Results show distinct variations in BER across scenarios, with AWGN exhibiting the best performance and epidemic interference causing the highest BER at lower SNRs, highlighting the need for effective interference mitigation strategies to ensure reliable communication in real-world environments.

Index Terms—*IoT, digital communication systems, impairment effects*

I. INTRODUCTION

The Internet of Things (IoT) encompasses autonomous devices that sense, communicate, and actuate without human intervention, finding applications across various sectors like industry, agriculture, and smart cities [1]. Key challenges for large-scale IoT networks include long-range connectivity and energy efficiency, which are addressed by low power wide area network (LPWAN) technologies such as LoRaWAN, NB-IoT, and SigFox [2]. These technologies offer long battery life and wide coverage at low bit rates [3].

In IoT networks using LoRa communication, the assumption of stationarity simplifies the analysis of noise and interference. However, interference often varies with network activity, making it non-stationary, similar to observations in cellular systems [4]. Traditional analytical formulations are less effective for non-stationary processes as demonstrated in [5].

Interference in this type of network can significantly impact communication due to limited spectrum frequency [6]. Managing spectrum scarcity through multiple coverage areas enables frequency reuse but also leads to potential interference between areas. Traditional models assume stationary interference, but dynamic systems with rapidly changing interference levels require non-stationary modeling. Epidemic interference, caused by many devices accessing the system simultaneously, further complicates the scenario, degrading receiver performance.

The objective of this work is to evaluate the performance of a digital communication system, specifically employing LoRa modulation, by analyzing the Bit Error Rate (BER) versus Signal-to-Noise Ratio (SNR) under various channel conditions. This includes modeling and simulating the effects of additive white Gaussian noise (AWGN), Rayleigh fading, and epidemic interference characterized by a lognormal distribution. By comparing multiple scenarios, the study aims to understand the impact of different types of interference and fading on the communication system's reliability and efficiency, and to highlight the importance of effective interference mitigation strategies in achieving optimal performance in real-world environments.

The methodology involves generating transmitted signals using LoRa modulation and simulating their transmission through various channel conditions to evaluate the Bit Error Rate (BER) against Signal-to-Noise Ratio (SNR). Scenario 1 considers only additive white Gaussian noise (AWGN), while Scenario 2 incorporates Rayleigh fading, modeled as complex Gaussian random variables. Scenarios 3, 4, and 5 introduce epidemic interference, characterized by a lognormal distribution with different levels of initial interference power, to the received signal. For each scenario, the received signal, comprising the transmitted signal, noise, fading, and interference, is processed using a down-chirp signal and transformed via the Discrete Fourier Transform (DFT) to estimate the transmitted symbols. The BER is calculated by comparing the estimated symbols to the original transmitted bits across a range of SNR values, allowing for a comprehensive analysis of the performance of the system under different interference and fading conditions.

The results demonstrate distinct variations in Bit Error Rate (BER) performance across the different scenarios. Scenario 1, which includes only additive white Gaussian noise (AWGN), shows a rapid decrease in BER with increasing Signal-to-Noise Ratio (SNR), indicating strong performance in noise-limited conditions. Scenario 2, incorporating Rayleigh fading, exhibits a more gradual BER decline, highlighting the adverse impact of multipath propagation. Scenarios 3, 4, and 5, which account for epidemic interference modeled by a lognormal distribution, reveal higher BERs at lower SNRs, reflecting the significant challenge posed by interference. Among these, Scenario 5

displays the highest BER, suggesting severe interference conditions. Despite the convergence of BER at higher SNRs, these scenarios emphasize the critical need for robust interference mitigation strategies to ensure reliable communication system performance.

Section I elaborates on LoRa modulation and epidemic interference, detailing the mathematical foundations and models used. The Performance Analysis Method (Section II) describes the simulation setup, including different scenarios with various noise and interference conditions. Section III section presents the findings, illustrating how different types of interference affect the Bit Error Rate (BER). Finally, the Final Remarks (Section IV) summarize the conclusions and suggest future research directions to enhance the reliability and efficiency of LoRa-based communication systems in real-world environments.

II. THEORETICAL BACKGROUND

A. LoRa Modulation

LoRa modulation [7] leverages the bandwidth B of a channel, transmitting samples at intervals $T = \frac{1}{B}$. Symbols s , transmitted every $T_s = 2^{\text{SF}} \cdot T$, in which SF is the Spreading Factor typically ranging from 7 to 12, are real numbers constructed from SF binary digits as

$$s(\mathbf{w}) = \sum_{h=0}^{\text{SF}-1} \mathbf{w}_h \cdot 2^h. \quad (1)$$

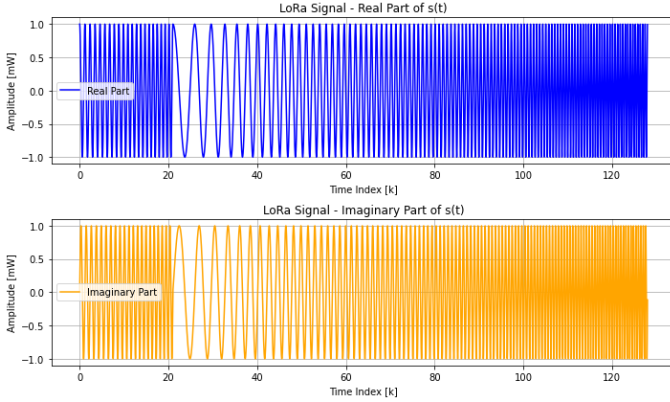


Fig. 1. Real and Imaginary parts of one oversampled LoRa chirp signal.

The symbol s takes values from 0 to $2^{\text{SF}} - 1$ and is used to form a chirp waveform, which has a frequency that increases linearly with time index k . The transmitted waveform for a given symbol is represented as

$$c(k) = \frac{1}{\sqrt{2^{\text{SF}}}} \exp[j2\pi \left(\frac{(s(nT_s) + k) \bmod 2^{\text{SF}}}{2^{\text{SF}}} \right) \frac{k}{2^{\text{SF}}}], \quad (2)$$

in which k ranges from 0 to $2^{\text{SF}} - 1$. The modulated signal is a chirp waveform with its frequency shifted by the symbol s . Signals in this domain exhibit periodic frequency representation with period B .

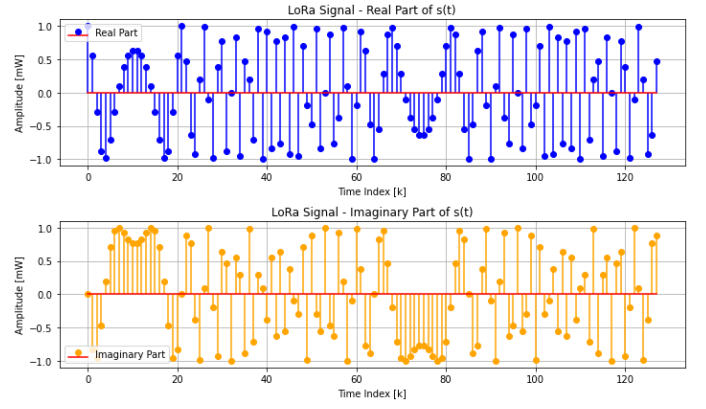


Fig. 2. Real and Imaginary parts of one LoRa chirp signal for discrete k values.

In order to detect the LoRa symbol, the received signal $r(k)$ is multiplied sample by sample by a down-chirp signal and transformed using the Discrete Fourier Transform (DFT). The output of the DFT provides the maximum projection value, indicating the best estimate of the transmitted symbol $s(\mathbf{w})$.

The down-chirp signal ($d(t)$) is given by

$$d(t) = \exp \left(-j \cdot 2\pi \cdot \frac{k^2}{2^{\text{SF}}} \right) \quad (3)$$

The product of the received signal and the down-chirp is transformed via DFT to identify the peak, which corresponds to the estimated symbol,

$$\text{DFT}(r(t) \cdot d(t)) \rightarrow \arg\max \quad (4)$$

The estimated symbol is then converted back to a binary sequence, and the bit errors are counted by comparing the estimated binary sequence to the original. Then, the BER for each SNR value is calculated.

This projection method leverages the periodicity and properties of the chirp signals, providing an optimal and computationally efficient approach to demodulating LoRa signals in the presence of AWGN.

B. Epidemic Interference

In IoT networks interference can become problematic when many devices transmit simultaneously. In such scenarios, interference can no longer be considered stationary; instead, it becomes a non-stationary stochastic process [5]. This means that the interference levels and their statistical properties change over time, complicating traditional analysis methods.

Traditional correlation analysis is inadequate for dynamic interference in IoT networks. Instead, stochastic differentiation can be used to model this phenomenon. This approach allows for formulating the interference problem using stochastic differential equations, providing a more accurate representation of the interference dynamics in IoT networks.

Recent studies have modeled epidemic interference in communication systems as a non-stationary stochastic process using stochastic integration [5]. These studies found that the

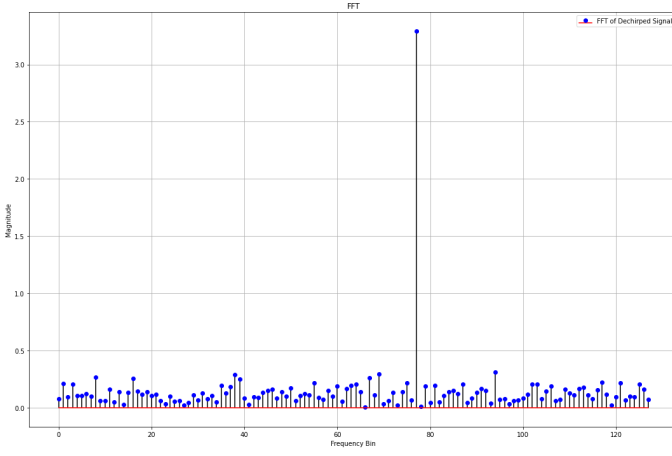


Fig. 3. DFT output indicating the best estimate of the transmitted symbol.

interference power follows a Lognormal distribution. The probability density function (PDF) of this distribution is given by

$$p_W(w) = \frac{1}{w\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(w) - \ln(P_0)}{\sigma}\right)^2\right], \quad (5)$$

in which w is the observed interference power from IoT devices, σ is the standard deviation of the logarithm of the interference power and P_0 represents the initial observed power at the beginning of the observation interval.

The Lognormal distribution indicates that the logarithm of the interference power is normally distributed. By understanding and predicting the behavior of interference with this model, IoT network designers can better manage and mitigate the impact of non-stationary interference, leading to more robust and efficient communication systems as can be observed in Section III.

C. Bit Error Rate

Bit Error Rate (BER) is a performance metric that quantifies the reliability and accuracy of data transmission. It is defined as the ratio of the number of bit errors to the total number of bits transmitted over a communication channel during a specified period. Mathematically, it is expressed as

$$\text{BER} = \frac{N_{\text{errors}}}{N_{\text{total}}}, \quad (6)$$

in which N_{errors} represents the number of incorrectly received bits and N_{total} denotes the total number of bits transmitted. A lower BER indicates a more reliable communication system, essential for applications requiring high data integrity.

To obtain the BER through simulation, a digital communication system can be modeled, and data transmission can be emulated over the simulated channel. For example, if a system transmits 10^6 bits and during the simulation, 500 bits are found to be received incorrectly, the BER can be calculated using the formula $\text{BER} = \frac{500}{10^6} = 5 \times 10^{-4}$. This process involves generating a known bit sequence, transmitting it through the

simulated communication channel, and comparing the received sequence to the original to count the bit errors.

D. Rayleigh Fading

Rayleigh fading is a statistical model used to describe the effect of a propagation environment on the signal amplitude and phase in wireless communication systems, particularly in scenarios where there is no direct line of sight between the transmitter and receiver.

The amplitude of the signal affected by Rayleigh fading follows a Rayleigh distribution, which arises when the signal experiences multiple scattering paths with different phase shifts. The amplitude $|h|$ is given by:

$$|h| = \sqrt{(X^2 + Y^2)} \quad (7)$$

in which X and Y are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and equal variance. The probability density function (PDF) of the Rayleigh distribution for the amplitude $|h|$ is

$$p_{|h|}(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \quad (8)$$

for $x \geq 0$, in which σ is the scale parameter related to the average power of the received signal.

The overall complex channel coefficient h in Rayleigh fading is represented as

$$h = X + jY \quad (9)$$

in which X and Y are i.i.d. Gaussian random variables with zero mean and variance σ^2 , and j is the imaginary unit. The magnitude $|h|$ follows the Rayleigh distribution, and the phase θ follows the uniform distribution.

This model effectively captures the random nature of signal variations due to multipath propagation in rich scattering environments, making it essential for analyzing and designing robust wireless communication systems.

III. PERFORMANCE ANALYSIS METHOD

The analysis of BER versus SNR was performed across multiple scenarios. These formulations encompass the generation of the transmitted signal, the modeling of the channel effects (including Rayleigh fading and epidemic interference), and the calculation of BER for each scenario as described in Table I.

In Scenario 1, the transmitted signal undergoes the effect of noise. Noise ($n(t)$) is added to the signal, with noise power (P_N) determined by the desired SNR

$$P_N = \frac{P_W}{\text{SNR}} \quad (10)$$

The noise is generated as a complex Gaussian random variable with zero mean and variance $\frac{P_N}{2}$ for both real and imaginary parts so that

$$n(t) = \sqrt{\frac{P_N}{2}} (N(0, 1) + jN(0, 1)) \quad (11)$$

TABLE I
SUMMARY OF CHANNEL CONFIGURATIONS FOR EACH SCENARIO

Scenario	Channel Effects	Noise Power (P_N)	Initial Interference Power (P_0)
1	AWGN	$\frac{P_W}{SNR}$	-
2	Rayleigh Fading + AWGN	$\frac{P_W}{SNR}$	-
3	Rayleigh Fading + AWGN + Epidemic Interference	$\frac{P_W}{SNR}$	0.01 W
4	Rayleigh Fading + AWGN + Epidemic Interference	$\frac{P_W}{SNR}$	0.015 W
5	Rayleigh Fading + AWGN + Epidemic Interference	$\frac{P_W}{SNR}$	0.02 W

The received signal ($r(t)$) combines these effects,

$$r(t) = c(t) + n(t). \quad (12)$$

The faded signal (s_f) is the product of the transmitted chirp signal and the Rayleigh fading coefficients

$$s_f(t) = h \cdot c(t) \quad (13)$$

The received signal ($r(t)$) in Scenario 2 is, therefore, given by

$$r(t) = s_f(t) + n(t) \quad (14)$$

The received signal ($r(t)$) in the presence of epidemic interference is modeled considering the lognormal distribution of the interference power for Scenarios 3, 4, and 5. This interference is characterized by the lognormal distribution as described in Equation 5. Given this, the received signal ($r(t)$) is formulated as

$$r(t) = s_f(t) + n(t) + I(t) \quad (15)$$

in which $I(t)$ is the epidemic interference, which is lognormally distributed.

In this formulation, the epidemic interference $I(t)$ introduces variability and randomness in the received signal, reflecting real-world conditions where interference power can significantly fluctuate. This approach captures the impact of both fading, noise, and interference, allowing for a comprehensive analysis of the system performance.

IV. SIMULATION RESULTS

The results from the analysis of Bit Error Rate (BER) versus Signal-to-Noise Ratio (SNR) across multiple scenarios reveal significant insights into the performance of the communication system under different channel conditions. Scenario 1, which considers only additive white Gaussian noise (AWGN), demonstrates a steep decline in BER as SNR increases, indicating that the system performs well in noise-limited environments. Scenario 2, incorporating Rayleigh fading, shows a more gradual decline in BER with increasing SNR compared to Scenario 1. This highlights the impact of multipath propagation, where fading causes signal degradation,

thereby requiring higher SNR values to achieve similar BER performance.

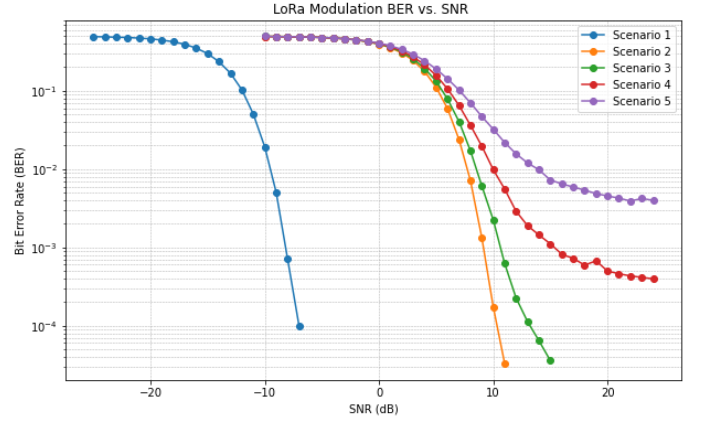


Fig. 4. Bit Error Rate (BER) versus Signal-to-Noise Ratio (SNR) across multiple scenarios.

Scenarios 3, 4, and 5 introduce epidemic interference modeled by a lognormal distribution, adding another layer of complexity to the received signal. These scenarios exhibit higher BER as the initial interference power, P_0 , increases, underscoring the significant impact of interference on signal quality. Among these, Scenario 5 shows the worst performance, indicating severe interference conditions. As SNR increases, the BER in these scenarios begins to converge, but the presence of interference continues to affect the system's reliability. This comprehensive analysis demonstrates that while increasing SNR can mitigate noise and fading effects, managing epidemic interference is crucial for optimizing communication system performance in real-world environments. The results emphasize the necessity of robust interference mitigation strategies to enhance the overall reliability and efficiency of the system.

V. FINAL REMARKS

This study provides a comprehensive analysis of the performance of LoRa modulation under various interference and fading conditions, specifically focusing on the impact of AWGN, Rayleigh fading, and epidemic interference modeled by a lognormal distribution. The results indicate that while LoRa modulation performs well in noise-limited environments, as evidenced by the steep decline in Bit Error Rate (BER) with increasing Signal-to-Noise Ratio (SNR) in Scenario 1, the presence of Rayleigh fading and epidemic interference significantly degrades performance. Rayleigh fading introduces multipath propagation effects, requiring higher SNR values to achieve similar BER performance, while epidemic interference, particularly in Scenarios 3, 4, and 5, leads to higher BERs at lower SNRs, underscoring the critical impact of dynamic and non-stationary interference on signal quality.

The findings emphasize the necessity of developing robust interference mitigation strategies to enhance the reliability and efficiency of LoRa-based communication systems, especially in environments with high interference levels typical of IoT

networks. Future work should explore advanced techniques such as adaptive modulation and coding, interference-aware routing, and machine learning-based interference prediction and management. Additionally, experimental validation in real-world settings will be essential to confirm the simulation results and further refine the strategies for optimizing LoRa performance under diverse and challenging conditions. By addressing these challenges, the robustness and scalability of IoT networks can be significantly improved, enabling more reliable and efficient communication in various applications.

REFERENCES

- [1] L. Lei, Y. Tan, K. Zheng, S. Liu, K. Zhang, and X. Shen, "Deep reinforcement learning for autonomous internet of things: Model, applications and challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1722–1760, 2020.
- [2] E. U. Ogbodo, A. M. Abu-Mahfouz, and A. M. Kurien, "A survey on 5g and lpwan-iot for improved smart cities and remote area applications: From the aspect of architecture and security," *Sensors*, vol. 22, no. 16, p. 6313, 2022.
- [3] M. Jouhari, N. Saeed, M.-S. Alouini, and E. M. Amhoud, "A survey on scalable lorawan for massive iot: Recent advances, potentials, and challenges," *IEEE Communications Surveys & Tutorials*, 2023.
- [4] J. E. Cordeiro Jr, M. S. Alencar, M. V. Yashina, and A. G. Tatashev, "Effect of epidemic interference on the performance of m-ask, m-psk and m-qam modulation schemes," *Journal of Communications Software and Systems*, vol. 17, no. 4, pp. 358–365, 2021.
- [5] M. S. Alencar, "Epidemic interference in a cellular system," in *Proceedings of the 17th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, 2020, pp. 81–84.
- [6] A. Petroni and M. Biagi, "Interference mitigation and decoding through gateway diversity in lorawan," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9068–9081, 2022.
- [7] L. Vangelista, "Frequency shift chirp modulation: The lora modulation," *IEEE signal processing letters*, vol. 24, no. 12, pp. 1818–1821, 2017.

Assessing the Impact of Jamming Attacks on LoRaWAN Devices: An Experimental Study

1st Vitória J. Baratella
IENTI

Federal University of Itajubá (UNIFEI)
Itajubá, Brazil
vitoriajacomelli@gmail.com

2nd Anderson G. de Almeida
IENTI

Federal University of Itajubá (UNIFEI)
Itajubá, Brazil
anderson.g.almeida@unifei.edu.br

3rd Caio T. B. Costa
IENTI

Federal University of Itajubá (UNIFEI)
Itajubá, Brazil
caio.tacito@unifei.edu.br

4th Décio R. M. Faria
IENTI

Federal University of Itajubá (UNIFEI)
Itajubá, Brazil
decio renno@unifei.edu.br

5th Danilo H. Spadoti
IENTI

Federal University of Itajubá (UNIFEI)
Itajubá, Brazil
spadoti@unifei.edu.br

Abstract—The exponential development of Internet of Things (IoT) devices transmitting data simultaneously has increased the unwanted signal interference. Despite typical radio frequency collisions, malicious actors can intentionally generate targeted interference, impairing network performance and packet delivery rates. This study offers an overview of the security vulnerabilities of LoRaWAN protocol and the implementation of a jammer on a LoRaWAN device transmitting sensitive environment data. The results show that a LoRaWAN device with received power of -51.2 dBm can suffer losses of 89% in its data packets due to a jamming attack with a reception power of -30.9 dBm, proving the attack's effectiveness.

I. INTRODUCTION

Extreme weather events are frequently affecting all Brazilian regions. According to [1], events that occurred in the Amazon affected other biomes, which also influenced the economy and water security in other regions of the country. In this way, the use of technology, science, and innovation in the implementation of environmental monitoring systems that aim for a sustainable future and the mitigation of damage caused by extreme events becomes relevant.

In this scenario, the Internet of Things (IoT) emerges as a prominent technology supporting the environmental monitoring. IoT refers to the fusion of computing and communication capabilities in everyday objects, facilitating interaction between the physical and digital domains through sensors and actuators [2]. However, such communication is also a challenge, once some sensing systems are placed faraway from central regions, requiring for wireless systems communications.

Moreover, to enable long-range wireless transmission for IoT devices, a collection of low-power wide area network (LPWAN) technologies has been investigated [3] offering extended connectivity over distances of kilometers [4]. Long-range (LoRa) could be interpreted as technique based on

spread spectrum modulation (SSM), developed by Semtech, and stands out as a key technology within LPWAN, offering long-range capabilities, low power consumption, and reduced data rates. LoRaWAN is the underlying protocols in LoRa, developed by LoRa Alliance [4] and aims to provide secure and uninterrupted wireless connectivity for battery-powered mobile devices and stationary end nodes in various IoT applications, with configurable frequency, bandwidth, encoding rate, and propagation factor [4]. Although LoRaWAN allows an end node to transmit messages to multiple gateways that communicate with the network server, as all wireless systems, it is subject of radio-frequency (RF) vulnerability and security limitations [5] [6] [7] [8] [9].

Regarding jamming attacks in LoRaWAN, there are numerous works that address the topic. [10] showed that due to the long air-time of LoRa messages the triggered jamming was possible and effective to blocked the signal LoRa. The triggered jamming technique was also tested in a previous work [11] sending 100 messages on the 868.1 MHz frequency. For the six SFs, and just 0.5% reached the gateway when the triggered jammer was on, corresponding 3 messages out of 600. In [12], researchers implement a jammer on a LoRa small board and investigate the effects of LoRa transmission settings on jamming performance across three scenarios. The authors discovered that simultaneous transmission of signals by the end node and jammer with the same chirp rate causes signal collisions. This collision leads to reception failures at the gateway.

Therefore, the vulnerabilities associated with wireless communication system must be studied and addressed, as the collection, interception, or failure to receive sensitive data packets can lead to large-scale impacts. Consequently, IoT devices in critical scenarios may become vulnerable and targeted by cyber attacks [13]. The main objective of this work is implementing a jammer in a LoRaWAN device to evaluate the effectiveness to maintain communication integrity

The authors thank CAPES, CNPQ -402093/2022-4, FAPEMIG, FAPESP 2021/06569-1 and the entire team at LabTel-LAIoT UNIFEI for their support to the research.

and determine the relationship between jamming power levels and the resulting packet losses in LoRaWAN transmissions.

II. JAMMING ATTACK TEST

In a real scenario, during transmission, the LoRaWAN signal can be intentionally blocked by attackers [2]. As a result, not only the data is lost, but also network throughput degrades, resulting in critical monitoring systems. By impairing the communication of critical systems, packet collisions can be considered a vulnerability in terms of security [13]. Moreover, depending on the LoRaWAN interfering parameter configurations different risk scenarios can occur. It is worth mentioning that, like all radio frequency technologies, the LoRaWAN network is subject to packet collisions or signal interference.

A typical collision situation, which does not correspond to an attack, involves parameters such as Spreading Factor (SF), Carrier Frequency (CF) and Transmission Power (TP).

LoRa can use six orthogonal spreading factors (SFs) from SF7 to SF12 to improve channel capacity and spectrum efficiency [14]. The spreading factor controls the chirp rate, and thus it controls the speed of data transmission. Lower spreading factors result in faster chirps, leading to an increased data transmission rate. For example, a signal modulated with the SF12 can travel a longer distance than a signal modulated with the SF7, but with lower bit rate.

In case the SF and timing of two LoRa end nodes overlap and the disparities in their CF and TP fall below specific thresholds, a collision may occur. Such collisions are a common occurrence LoRaWAN transmissions and are part of normal network operations [12]. Other characteristics of the network promote the occurrence of typical collisions. For example, the transmission time of LoRa messages is long due to the low bit rate, which exposes the messages to greater chances of collisions. Furthermore, LoRa messages are susceptible to interference from other synchronous LoRa messages that have a higher signal strength [11].

In contrast to proper collisions, however, malicious actors can exploit the network using LoRa Commercial-Off-The-Shelf (COTS) devices to intentionally disrupt LoRa networks [2]. These deliberate jamming attacks (collisions) generate targeted interference, significantly impairing network performance and packet delivery rates [4].

The jamming attack stands out as a prominent challenge in radiofrequency security [15]. Leveraging commercial radio devices and freely available software, attackers can execute such assaults with relative ease. Through the emission of interference signals of sufficient power, the attackers saturate the frequency band. This interference obstructs communication channels, impairing the transmission of authentic data while inundating systems with random, unauthenticated signals [2].

The gravity of such security concerns could be exemplified by real world events, such as the malicious interference with GPS systems that have impacted thousands of flights within Europe from the second semester of 2023 to the first months of 2024. Data reported from airlines reveals that approximately

46,000 aircraft have encountered GPS-related issues over the Baltic Sea. GPS jamming poses a significant security threat to airlines by disrupting crucial navigation systems, potentially leading to navigational errors and compromised flight safety. These disruptions not only underscore the vulnerability of critical communication infrastructure but also highlight the urgent need for robust cybersecurity measures to safeguard against malicious interference in vital networks [16].

This type of attack creates a blockage on the LoRaWAN network, preventing the packets sent by authorized devices from reaching their intended destination. This occurs because LoRa transmissions that use the same frequency and the same spreading factor can interfere with one another [17]. According to [?], about 99 percent of LoRa transmissions are damaged by data flooding.

In this context, the present project focuses on exploring jamming phenomena within devices employing LoRaWAN communication protocols. The primary objective is to examine various potential jamming scenarios, elucidating their implications and ramifications. The subsequent section offers a comprehensive overview of these scenarios.

Jamming attack tests were carried out in LoRa transmissions, and jamming signals at different powers in a laboratory environment were investigated. Thus, it is possible to evaluate the effectiveness of the attacks, measuring how much the jamming signals generated will interfere with LoRaWAN device's communication, generating loss of data packets.

A. Materials and Methodology

The experimentation process was developed using the following equipment:

- 1 Vector Signal Generator MG3710A (Anritsu);
- 2 Directional Antennas (Aaronia AG);
- 1 Spectrum Rider FPH Rohde&Schwarz;
- 1 LoRaWAN Gateway (RAK);
- 1 Water Level Transmission LoRaWAN Device.

The test was conducted in two phases. The first phase measured the received power of the jamming signal and the received power of the LoRaWAN device, while the second phase focused on evaluating the packet losses. It is worth highlighting that both phases were done with all the setup in the exact same position, only swapping the spectrum analyzer for the LoRaWAN Gateway on phase two. It is also noteworthy that the spreading factor used was SF12 throughout the experiment.

The first phase was executed using a signal generator to produce the jamming signal, which corresponds to a carrier wave - sinusoidal signal that sweeps a frequency range of 915.1MHz to 915.4MHz. It is necessary to point out that the bandwidth corresponding to the sweep used in the jamming signal was 200kHz, which is greater than the bandwidth of the LoRa channel being investigated. This channel has 125kHz of bandwidth, standard for the LoRaWAN protocol. This way, it is possible to guarantee that the attack will be successful across the entire bandwidth.

Figure 1 shows a diagram of the test structure.

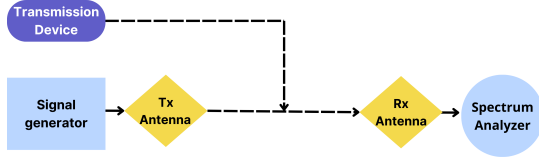


Figure 1: Phase 1 Diagram

The physical configuration of the experiment bench is shown in Fig. 2.

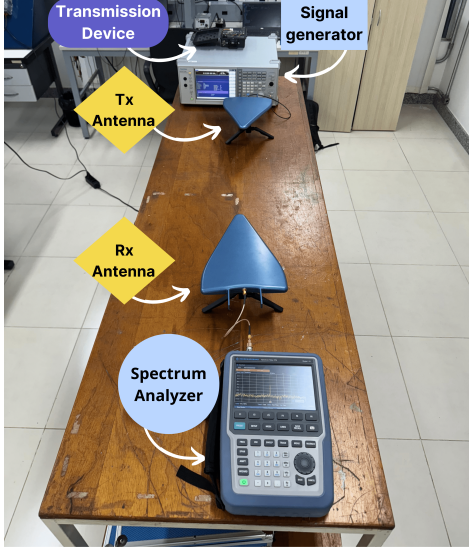


Figure 2: Phase 1 Setup

The signal generator with an attached antenna was configured to produce jamming signals at different power levels (from 0dBm to -40dBm).

The spectrum analyzer with the receiving antenna were used to measure and record the jamming signal reception power.

The same procedure was carried out with the LoRaWAN device in this first phase, that is, the device transmitted at different power levels of 5dBm and 20dBm. Thus, the power received by the spectrum analyzer with antenna in the set were recorded.

In the second phase of the experiment, the spectrum analyzer was replaced with the LoRaWAN gateway while maintaining the rest of the setup. The signal generator was configured to produce jamming signals once again, simultaneously with the LoRaWAN device transmitting packets. This configuration allowed the LoRaWAN gateway to receive the packets from the device, enabling the measurement of packet losses at each jamming power level.

Figure 3 shows a diagram of the test structure.

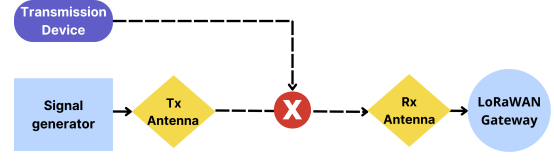


Figure 3: Phase 2 Diagram

The physical configuration of the experiment bench is shown in Fig. 4.

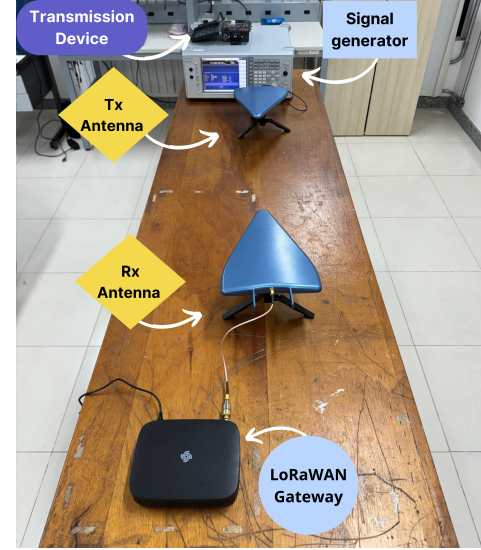


Figure 4: Phase 2 Setup

III. RESULTS

As a result of the procedures carried out in phase 1, we have a jamming signal in the spectrum analyzer as shown in Figure 5. The signal generator was configured to generate a carrier signal with a constant level that rapidly changes frequency in 2kHz steps, sweeping the 200kHz in the desired band in a cyclical manner. This method guarantees precise control of the signal level, thus enabling the assessment of the impact of blocking on the channel. Figure 5 shows the jamming signal at a single instant during its 200kHz sweep over the channel.

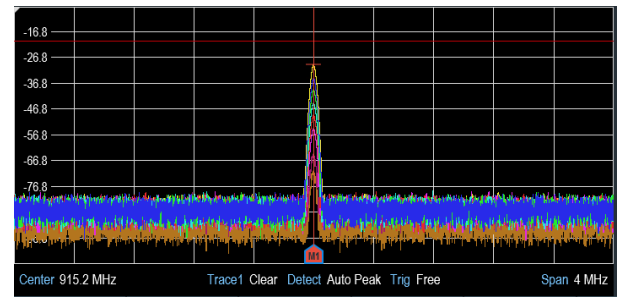


Figure 5: Received Power of Jamming Signal. Yellow: -29dBm; Blue: -30.9dBm; Green: -34.2dBm; Light Blue: -39.2dBm; Red: -49dBm; Pink: -59.4dBm; Purple: -69.2dBm; Orange: -73.5dBm.

Continuing with the results of phase 1, Fig 6 shows the graph generated by the spectrum analyzer, displaying the power levels received from the LoRaWAN device.

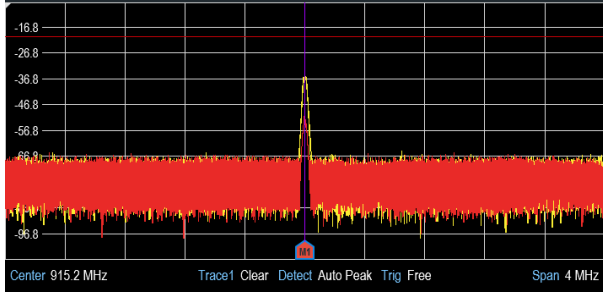


Figure 6: Received Power of LoRaWAN Device. Yellow: -32.7dBm; Red: -51.2dBm

Once phase 1 is completed, phase 2 aims to analyze what power of the jamming signal generates packet losses from the LoRaWAN device.

Initially, a control test without the jamming signal was carried out using the LoRaWAN device with received power reading of -32.7 dBm. Therefore, it was possible to determine the number of packets received by the gateway without attempting an attack. A total of 36 packets were received in a period of 5 minutes, so this result was used as a reference as being 100% reception.

After defining this parameter, phase 2 tests began. The LoRaWAN device was tested in two configurations: maximum transmit power level, and minimum transmit power level. The maximum reading received at the spectrum analyzer was -32.7 dBm, representing the highest possible interference immunity scenario, while the minimum received reading was -51.2 dBm, indicating the weakest transmit power scenario. Following this configuration, various strengths of the jamming signal were applied to analyze packet loss, as detailed in Table I below. The fourth column refers to SNR (Signal-to-Noise Ratio) reported by the LoRaWAN gateway. This parameter is the ratio of the received signal power to the noise floor, commonly used to determine the quality of the received signal [18]. As seen in the table, the SNR values were negative, which means that the signal power is less than the noise floor.

Jamming Received Power (dBm)	LoRaWAN Device Received Power (dBm)	Packet loss (%)	SNR (Signal-to-noise ratio) received by the gateway
-39.2	-32.7	0%	-0.42
	-51.2	0%	-3.92
-34.2	-32.7	0%	-0.98
	-51.2	3%	-11.63
-30.9	-32.7	0%	-2.21
	-51.2	89%	-14.35
-29	-32.7	0%	-3.68
	-51.2	94%	-16.25

Table I: Packet loss in different scenarios

The following graph on Fig. 7 shows the curves of SNR

received by the gateway versus Jamming Rx Power on both received power levels of LoRaWAN device, being blue -32.7 dBm and red -51.2 dBm.

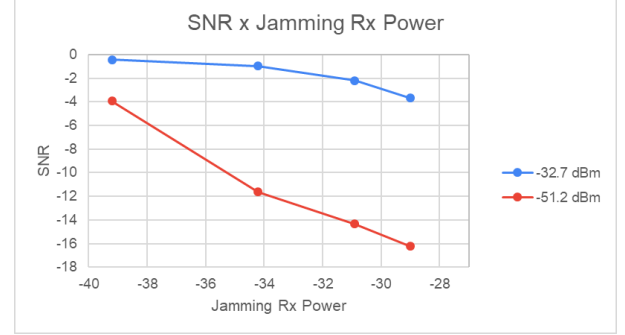


Figure 7: SNR x Jamming Rx Power

As shown in Table I, when the received power of LoRaWAN device is in -32.7 dBm, there is no packet loss, however the deterioration of the SNR is perceptible, as seen in Table 7.

When the measured transmission power from the LoRaWAN device is -51.2 dBm, the jamming significantly influences packet loss. For a jamming signal with a measured power of -34.2 dBm, the packet loss is 3%, accompanied by a signal-to-noise ratio (SNR) of -11.63 dB. However, increasing the jamming power to -30.9 dBm results in a substantial rise in packet loss, reaching 89%, with a corresponding SNR of -14.35 dB, illustrating the threshold of the gateway sensitivity.

Furthermore, at a jamming power of -29 dBm, the lost packets reach 94%, with an SNR of -16.25 dB, which constitutes and effectively denial of communication.

IV. CONCLUSIONS

Therefore, the work carried out managed to provide an overview of the security vulnerabilities presents in LoRaWAN protocol and the implementation of a jammer. It is concluded that, despite the increasing use of LoRaWAN devices to monitor various parameters, including critical environmental factors, this technology presents risk scenarios that can be exploited by malicious devices, interfering with data transmission and integrity.

The results indicate that a LoRaWAN device sending packets with a received power of -51.2 dBm suffers interference of a jamming signal with a received power of -30.9 dBm, reaching 89% of packet losses and a SNR of -14.35, highlighting a significant risk to the reliability of LoRaWAN networks.

Future work could focus on developing possible counter-measures against jamming attacks and analyzing other attack methods along with their respective effectiveness.

REFERENCES

- [1] (2023) Climate crisis: severe drought in the amazon is worsened by deforestation and fire. [Online]. Available: <https://www.wwf.org.br/?87003/Crise-climatica-seca-severa-na-Amazonia-e-agravada-por-desmatamento-e-fogo>. Accessed in 05/10/24.
- [2] N. M. G. R. Torres, "Exploring iot security vulnerabilities in lpwans: The ipcvc birac bicycle use case," *IPVC*, 2022.

- [3] (2020) Introduction to iot — what is lpwan. [Online]. Available: <https://www.leverage.com/iot-ebook/iot-lpwan>
- [4] A. G. L. R. C. N. S. Chilamkurthy, O. J. Pandey and H. N. Dai, “Low-power wide-area networks: A broad overview of its different aspects,” *IEEE Access*, vol. 10, pp. 81 926–81 959, 2022.
- [5] J. V. M. Bor and U. Roedig, “Lora for the internet of things,” *EWSN '16*, p. 361–366, 2016.
- [6] M. Bor and U. Roedig, “Do lora low-power wide-area networks scale?” *Proceedings of the 19th ACM Int. Conf. Modeling, Analysis Simulation Wireless Mob. Syst.*, pp. 59–67, 2016.
- [7] J. P. K. Mikhaylov and T. Haenninen, “Analysis of capacity and scalability of the lora low power wide area network technology,” *22nd European Wireless Conference*, pp. 1–6, 2016.
- [8] M. S. U. Raza, P. Kulkarni, “Low power wide area networks: An overview,” *IEEE Communications Surveys Tutorials*, vol. 19, p. 855–873, 2017.
- [9] O. Georgiou and U. Raza, “Low power wide area network analysis: Can lora scale?” *IEEE Wireless Communications Letters*, vol. 6, pp. 162–165, 2017.
- [10] E. A. et al., “Exploring the security vulnerabilities of lora,” *IEEE (CYBCONF)*, pp. 1–6, 2017.
- [11] —, “Selective jamming of lorawan using commodity hardware,” *MobiQuitous*, 2017.
- [12] R. G. C. S. J. Y. C. Y. Huang, C. W. Lin and S. T. Sheu, “Experimental evaluation of jamming threat in lorawan,” *IEEE VTC2019-Spring*, pp. 1–6, 2019.
- [13] S. Chacko and M. D. Job, “Security mechanisms and vulnerabilities in lpwan,” *IOP Conference Series: Materials Science and Engineering*, p. 012027, 2018.
- [14] C. K. A. Mpitiopoulos, D. Gavalas and G. Pantziou, “A survey on jamming attacks and countermeasures in wsns,” *IEEE Communications Surveys Tutorials*, vol. 11, pp. 42–56, 2009.
- [15] M. L. et al, “A colonel blotto game for anti-jamming in the internet of things,” *IEEE (GLOBECOM)*, pp. 1–6, 2015.
- [16] G. Topham. (2024) Thousands of flights to and from europe affected by suspected russian jamming. [Online]. Available: <https://www.theguardian.com/business/2024/apr/22/thousands-of-flights-to-and-from-europe-affected-by-suspected-russian-jamming>
- [17] W. M. B. Reynders and S. Pollin, “Range and coexistence analysis of long range unlicensed communication,” *ICT*, pp. 1–6, 2016.
- [18] (2023) Rssi and snr. [Online]. Available: <https://www.thethingsnetwork.org/docs/lorawan/rssi-and-snr/>. Accessed in 05/23/24.

Avaliação de Colheita de Energia como Fonte de Alimentação para Dispositivos de Baixo Consumo

1st Vidara Rodrigues Moura
Departamento da Indústria
IFCE
Fortaleza, Brasil
vidara350@hotmail.com

2nd Rejane Cavalcante Sá Rodrigues
Departamento da Indústria
IFCE
Fortaleza, Brasil
rejane_sa@ifce.edu.br

3rd Rogério Diógenes Filho
Departamento de Informática
IFCE
Fortaleza, Brasil
rogerio.diogenes@ifce.edu.br

Resumo—Este trabalho faz a análise da energia de um dispositivo de baixo consumo para chave fusíveis. O objetivo é validar a utilização da energia solar como fonte de alimentação para este equipamento. Para isso foi realizada uma análise gráfica coletada da operação do equipamento em ambiente externo, e simulada sua operação em campo para validar o produto final com o método de armazenamento de energia do dispositivo. A análise dos resultados revelou o funcionamento adequado do dispositivo após a implementação do método de colheita de energia, comprovando a eficácia da utilização da conversão de energia ambiente como forma de alimentação de circuitos de baixa corrente.

Palavras-Chave—Colheita de energia, baixo consumo, internet das coisas, células fotovoltaicas

I. INTRODUÇÃO

A colheita de energia é o processo de captura e conversão da energia ambiente em energia elétrica para armazenamento e/ou alimentação de dispositivos eletrônicos de baixo consumo, com o intuito de tornar esses equipamentos autônomos e eliminar a necessidade de fontes de energia convencionais, como as baterias. Existem diversas fontes de energia em qualquer ambiente típico, tendo como exemplo a energia solar, que pode ser colhida utilizando painéis solares fotovoltaicos que transformam a luz solar em eletricidade. A escolha do tipo de energia adequada dependerá da aplicação e do ambiente ao qual o dispositivo será inserido [1].

O desenvolvimento da colheita de energia foi impulsionado pelo crescimento da utilização de equipamentos eletrônicos autônomos [2]. Dispositivos autônomos são caracterizados por não possuir cabos para alimentação ou transmissão de dados. Esses dispositivos são desenhados para operar com baixo consumo de energia em modo de espera (*standby*), podendo ser considerados como cargas pulsadas de baixa potência [3]. Isso é alcançado através de otimizações em várias áreas, incluindo arquitetura de hardware e algoritmos de processamento eficientes [4].

O uso de baterias primárias para esse tipo de aplicação apresenta algumas desvantagens, como armazenamento de energia finito, extração ineficiente da energia armazenada [5], custos, dificuldades no processo de substituição, entre outros [6]. Outro agravante para o uso de baterias primárias, quando operam em ambientes externos e são submetidas a altas temperaturas, é o envelhecimento devido à exposição

térmica [7]. Dessa forma, a utilização de colheita de energia se faz vantajosa pois proporciona um método de alimentação em locais que não possuem fontes de energia convencionais disponíveis, oferecendo a alimentação elétrica necessária para a aplicação, além da baixa ou inexistente necessidade de manutenção [2].

A energia solar é uma das principais fontes de energia renovável devido à sua abundância e maior disponibilidade em comparação com outras fontes de energia. As células fotovoltaicas são usadas para converter energia solar em elétrica, processo conhecido como efeito fotoelétrico. A luz incidente sobre uma célula fotovoltaica pode ser refletida, absorvida ou transmitida, mas apenas a luz absorvida gera energia elétrica [8]. A eficiência dessa conversão dependerá do espectro da fonte de luz, que será afetado pelo posicionamento do dispositivo e pelas condições climáticas [9].

O IoT (*Internet of Things*) apresenta a ideia da conexão de dispositivos (coisas) à internet. Essa visão fomenta a criação de produtos de baixo custo, com restrições de tamanho e baixo armazenamento de energia, capazes de adquirir dados do contexto em que estão inseridos e transformá-los em informações pertinentes para análise e monitoramento [10]. Como forma de comunicação sem fio, os dispositivos IoT buscam utilizar métodos de comunicação que possuam baixo consumo de energia durante transmissões de longo alcance [11]. Essas características descrevem bem a tecnologia LoRa (*Long Range*), que utiliza a modulação *Chirp-spread-spectrum* (CSS) para atingir tais objetivos [12].

Um sensor monitor de ativo é um dispositivo capaz de coletar dados sobre as condições ou comportamento de um ativo específico. Esse tipo de equipamento é utilizado para o acompanhamento em tempo real ou periódico dos dados do ativo que está sendo monitorado. Essas informações podem ser sobre temperatura, movimento, vibração, entre outras variáveis do sistema. Esses sensores são comumente conectados a sistemas de aquisição de dados que permitem que os dados coletados sejam processados e analisados [13].

O objetivo deste trabalho é analisar e validar a eficácia da substituição da alimentação tradicional de um equipamento, como uma bateria primária, pela colheita de energia. Para isso, foram modificados o circuito eletrônico e o encapsulamento mecânico do equipamento de monitoramento de chave fusível

que utilizava bateria primária. O artigo está organizado da seguinte forma: a seção II descreve as propriedades do equipamento em teste, bem como suas características eletrônicas e mecânicas, a seção III detalha os testes e os resultados obtidos com o uso do circuito de colheita de energia e, por fim, a seção IV conclui o artigo.

II. DESENVOLVIMENTO

A. Atributos do dispositivo

O equipamento que utiliza as técnicas de colheita de energia como fonte de alimentação trata-se de um monitor de ativos para chaves fusíveis. Chaves fusíveis são equipamentos de proteção usualmente localizados entre transformadores de distribuição e em locais de derivação dos alimentadores. São utilizados para proteger os ativos presentes nas linhas contra defeitos de sobrecorrente, curto-circuito e em casos de falhas de aterramento à jusante. Seu funcionamento se dá através da fusão do elemento fusível após a passagem de elevadas correntes dentro de um determinado tempo [14]. Após a fusão, ocorre a formação de arcos elétricos, que são extintos pelos gases deionizantes gerados no cartucho. A expansão desses gases gera uma pressão que irá desconectar o porta-fusível do terminal superior, indicando a queima do elo [15].

Localizar exatamente onde ocorreu a fusão do elo fusível pode demandar bastante tempo se realizado em condições adversas, como baixa luminosidade [16]. A interrupção da alimentação nas linhas elétricas e os consequentes atrasos na localização e restauração do serviço geram prejuízos para as concessionárias de energia pela falta de faturamento do consumo que ocorreria em circunstâncias normais, pelas multas devido às transgressões dos indicadores de qualidade de energia e também pelas compensações financeiras aos seus consumidores [17].

Devido à difícil acessibilidade de instalação do equipamento, o aproveitamento de energia proveniente do ambiente oferece vantagens sobre formas de alimentação típicas, como baterias primárias, permitindo prolongar o tempo de operação e reduzir a necessidade de manutenções durante a vida útil do dispositivo.

B. Eletrônica

O desenvolvimento eletrônico é composto de 5 partes, a célula fotovoltaica, o circuito de alimentação, o armazenamento de energia, o processamento e comunicação e o circuito de monitoramento. Sendo apresentado no diagrama de blocos da Fig. 1.

1) *Célula Fotovoltaica*: A célula fotovoltaica é um componente capaz de converter a luz solar em energia elétrica. Seu funcionamento se dá pela luz absorvida, que transfere energia para os elétrons presentes na célula. Esses elétrons, munidos dessa energia, escapam de suas posições iniciais nos átomos do material fotovoltaico e contribuem para um fluxo elétrico, ou corrente. Células fotovoltaicas são feitas de materiais semicondutores, depositados ou dispostos em várias estruturas. Os principais materiais utilizados nesses tipos de

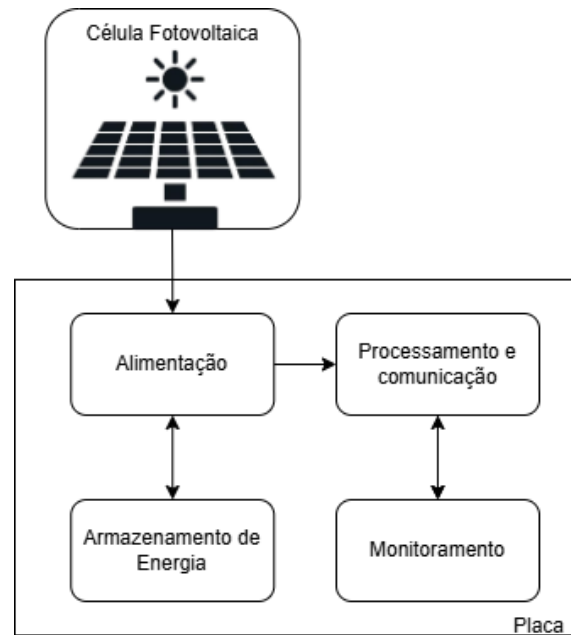


Fig. 1. Diagrama de blocos simplificado.

dispositivos são: silício, filmes finos policristalinos e filmes finos monocristalinos [8].

2) *Circuito de alimentação*: A energia absorvida pela célula fotovoltaica é direcionada para um circuito elevador de tensão. Este passo é essencial devido às variações na exposição solar ao longo do dia, que resulta em uma entrada de tensão oscilante. O conversor elevador garante uma saída estável e com níveis de tensão adequados ao sistema. Além de alimentar o sistema, essa energia é armazenada para manter o circuito operante mesmo na ausência de luz.

3) *Armazenamento de energia*: No intuito de manter o dispositivo operante na ausência de luz solar, é utilizado supercapacitores para armazenar a energia coletada durante os dias ensolarados. Os supercapacitores, também conhecidos como capacitores elétricos de dupla camada ou capacitores eletroquímicos, são dispositivos capazes de armazenar energia e costumam ser divididos em três tipos: eletroquímicos de dupla camada (EDLCs), pseudocapacitores e SCs híbridos. Este componente tem sido bastante utilizado por suportar vários ciclos de carga, possuir excelente potência específica e alta densidade de potência (rápida taxa de carga-descarga) [18].

4) *Processamento e comunicação*: Para dispositivos IoT que visam uma longa vida útil, o gerenciamento eficiente do processamento é importante para minimizar o consumo de energia e garantir a aquisição adequada da variável monitorada. O equipamento em questão apresenta um microcontrolador associado a um rádio de comunicação LoRa, que envia uma mensagem de *keep-alive* por dia. Esta mensagem tem o intuito de garantir a correta operação entre dispositivo e sistema, além de fornecer informações sobre os níveis de tensão do equipamento. No restante do tempo, o dispositivo entrará em

modo de *standby* para garantir o menor gasto de energia possível.

5) *Circuito de monitoramento*: A supervisão do ativo será realizada através de um acelerômetro. Quando ocorrer a fusão do elemento fusível, o dispositivo detectará o movimento e a alteração do ângulo em relação ao seu eixo vertical. Como forma de alerta, serão enviadas mensagens para que o sistema identifique a intercorrência, possibilitando a tomada de medidas para localização e reparo do elo.

C. Mecânica

Visando à praticidade de instalação por se tratar de um equipamento de monitoramento localizado em linhas de distribuição, houve um cuidado especial com o modo de instalação do dispositivo. O mecanismo de fixação utiliza um clipe projetado para se prender adequadamente ao cartucho do elo fusível. A peça central é injetada com PBT (Polibutileno Tereftalato) sólido, enquanto a tampa é feita de policloreto de vinila transparente, permitindo que os raios solares sejam absorvidos pela célula fotovoltaica. Na junção entre a peça central e a tampa, existe uma cavidade destinada ao uso de um anel de vedação, garantindo a estanqueidade do produto e assegurando sua proteção contra as adversidades do ambiente externo. O dispositivo é visto na Fig. 4 em branco no encaixe do elo fusível.

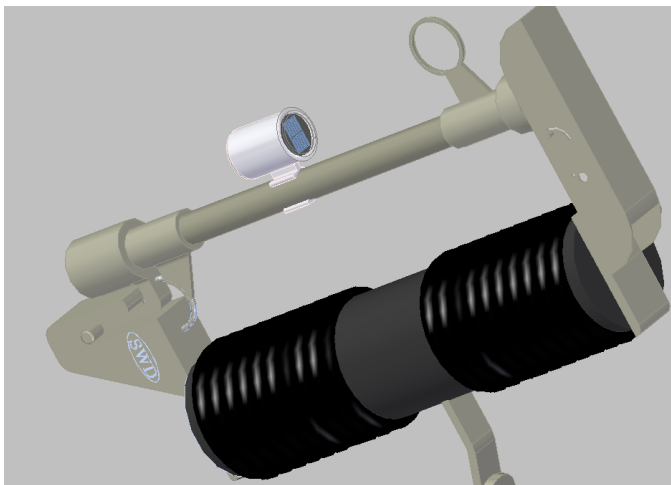


Fig. 2. Visualização 3D do encaixe do dispositivo no elo fusível.

III. TESTES E RESULTADOS

Para os testes e validação da colheita foi utilizado o equipamento visualizado na Fig. 3 Foi definido o perfil de consumo do equipamento, avaliado o padrão de carregamento e a tensão no elemento de armazenamento ao longo do tempo.

A. Perfil de consumo

Para iniciar a análise e determinar os componentes para o circuito de colheita de energia, foi fundamental obter o perfil de consumo do equipamento no qual se pretende modificar a entrada de alimentação. Isso permitiu dimensionar corretamente o elemento de armazenamento para garantir a

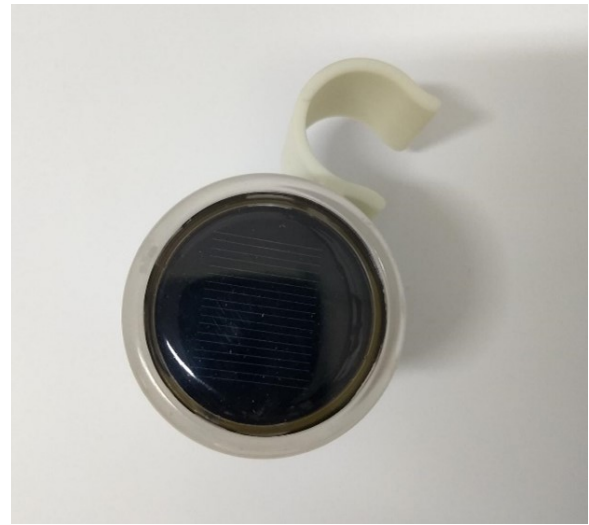


Fig. 3. Foto do dispositivo desenvolvido visto da parte de cima visualizando a célula fotovoltaica.

autonomia do dispositivo mesmo em condições climáticas adversas. Observe na Fig. 4 que, inicialmente, o dispositivo está em modo *standby*, consumindo pouca energia. Próximo ao momento do envio, o equipamento altera seu estado e inicia uma rotina de aferição dos dados a serem transmitidos. Durante a transmissão, ocorre um pico de consumo devido à comunicação LoRa. Por fim, o equipamento retorna ao uso de um nível de corrente referente a seu estado normal de operação, para processar e armazenar as informações antes de voltar ao modo *standby*.

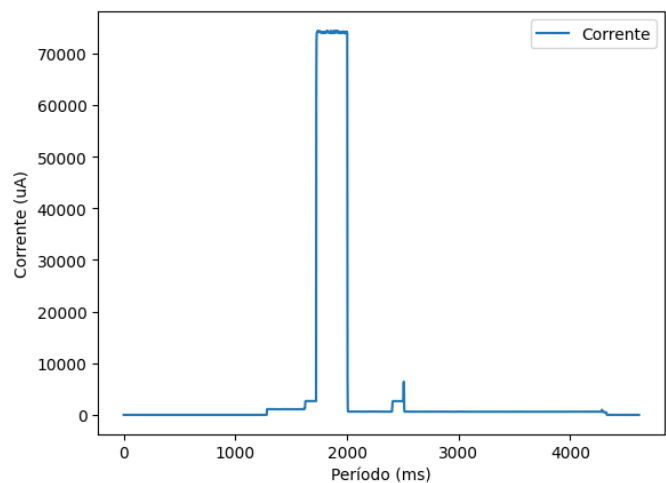


Fig. 4. Visualização ao longo do tempo do padrão de consumo durante transmissão LoRa.

B. Padrão de carregamento

O gráfico da Fig. 5 demonstra o padrão de carregamento do dispositivo, onde temos a tensão do componente armazenador (em azul) e a captura da luminosidade instantânea (em laranja). Percebe-se que a leitura da luminosidade apresenta algumas

irregularidades, sugerindo que a incidência solar no dispositivo não foi constante. Essas variações foram provocadas pela obstrução da luz solar devido às nuvens, refletindo na diminuição da energia captada pela célula solar.

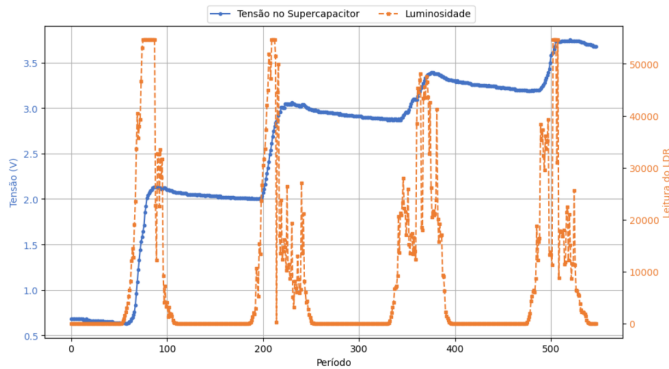


Fig. 5. Curva de carregamento do supercapacitor.

No que se refere ao carregamento do dispositivo, observou-se um aumento significativo da tensão durante o período matutino, estabilizando-se à medida que a luminosidade diminuía e apresentando uma leve descarga à noite devido ao consumo do equipamento e à ausência de alimentação de entrada. Sob as condições de luminosidade observadas, nota-se que o maior carregamento ocorreu no primeiro dia, favorecido pelas condições climáticas que permitiram um maior tempo de incidência solar no dispositivo. No total, foram necessários quatro dias para um carregamento completo.

Uma estratégia para reduzir esse período de carga foi considerar o uso de uma célula solar de área ampliada e maior eficiência, sendo assim capaz de fornecer uma corrente de carga superior para o circuito.

C. Tensão no elemento de armazenamento ao longo do tempo

Após a implementação do circuito de alimentação por colheita de energia, o equipamento foi instalado em ambiente externo ao longo de 13 dias. Esse tempo foi dividido em duas partes para melhor análise.

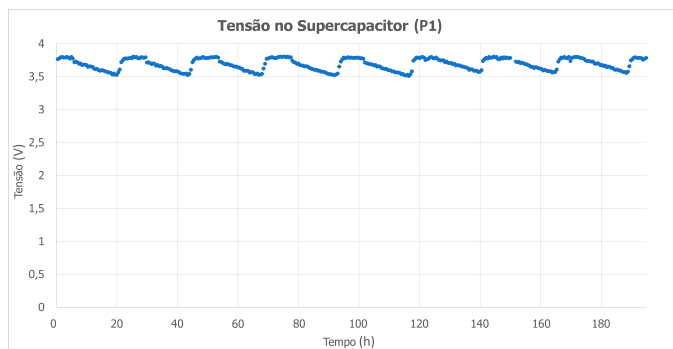


Fig. 6. Comportamento de carga e descarga do elemento armazenador ao longo do tempo - Parte 1.

No gráfico da Fig. 6, a tensão do supercapacitor é apresentada até o início do nono dia. Observa-se, no começo de cada

dia, o carregamento do elemento armazenador, estabilizando-se ao chegar à tensão nominal de carga. Com a diminuição da incidência solar sobre a célula, a tensão cai devido ao consumo do dispositivo e à ausência de alimentação de entrada. Devido ao intervalo de amostragem dos pontos do gráfico, não foi possível perceber a queda de tensão durante o envio de keep-alive em todos os dias, pois a transmissão ocorreu em momentos que havia luz solar, permitindo que os supercapacitores fossem carregados novamente. Essa queda foi visível apenas nos dias 6 a 8, onde se nota um ponto com tensão inferior durante o período de carga matinal.

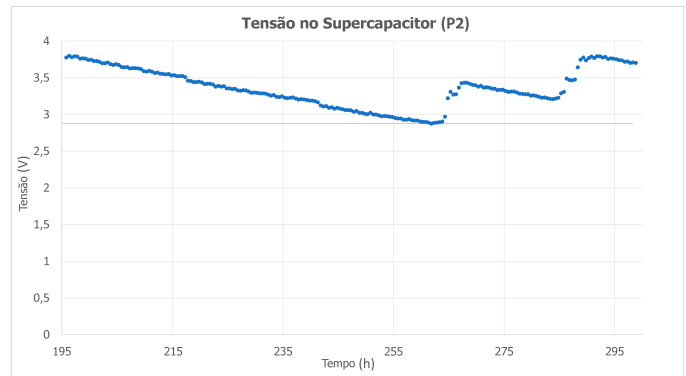


Fig. 7. Comportamento de carga e descarga do elemento armazenador ao longo do tempo - Parte 2.

O segundo gráfico se inicia na metade do nono dia até o décimo terceiro dia. Percebe-se que, durante os dias 10 e 11, não houve o carregamento do dispositivo devido às condições climáticas desfavoráveis, resultando apenas no consumo de energia. Mesmo com a falta de luz, a tensão do equipamento não caiu abaixo de 2,8V, o que foi suficiente para manter sua operação normal. Nos dias 12 e 13, percebe-se o retorno da carga até a estabilização de sua tensão nominal.

IV. CONCLUSÃO

O objetivo deste trabalho foi apresentar a aplicação de colheita de energia em um equipamento IoT de baixo consumo. Para isso, foi analisado a curva de consumo do equipamento e foram definidos os circuitos de captura e armazenamento de energia. Com essa informação, foi aferido o tempo de carregamento, e a curva apresentada poderia ser otimizada através do aprimoramento do código para reduzir o consumo do equipamento em *standby* ou adotando células solares maiores e mais eficientes. Após a execução do circuito, o equipamento foi monitorado por 13 dias, incluindo dias sem incidência solar. Apesar das condições climáticas adversas, o dispositivo se manteve operante sem atingir o seu limiar mínimo de tensão. Esse tempo de autonomia pode ser ampliado utilizando elementos armazenadores de maior capacidade, desde que compatíveis com os requisitos do produto e os custos de implementação. Assim, é possível observar que o uso de componentes de conversão de energia como forma de alimentação de dispositivos de baixo consumo é eficaz, prolongando a vida útil do equipamento sem comprometer o seu funcionamento.

Assim, o modelo de colheita foi satisfatória e validada no dispositivo.

REFERÊNCIAS

- [1] S. Beeby and N. White, *Energy Harvesting for Autonomous Systems*, Norwood, MA: Artech House, 2010, pp. 1–3.
- [2] G. Pistoia, *Energy Harvesting Systems: Principles, Modeling and Applications*. New York, NY, USA: Elsevier, 2008.
- [3] D. J. López and I. Antxustegi-etxearte Atienzar, "Powering Autonomous Sensors by RF Harvesting," M.S. thesis, Telecommunication Engineering & Management, Universitat Politècnica de Catalunya, Barcelona, Spain, Sep. 13, 2013.
- [4] C. Piguat, *Low-Power CMOS Circuits: Technology, Logic Design, and CAD Tools*. Boca Raton, FL, USA: CRC Press, 2006.
- [5] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks", *ACM Transactions on Embedded Computing Systems*, vol. 6, no. 4, pp. 393-407, Sep. 2007.
- [6] : Famitafreshi, G.; Afaqui, M.S.; Melià-Seguí, J. "A Comprehensive Review on Energy Harvesting Integration in IoT Systems from MAC Layer Perspective: Challenges and Opportunities". *Sensors* 2021, 21, 3097. <https://doi.org/10.3390/s21093097>
- [7] F. Leng, C. M. Tan, and M. Pecht, "Effect of temperature on the aging rate of Li ion battery operating above room temperature," **Scientific Reports**, vol. 5, no. 12967, 2015.
- [8] A. Khaligh and O. C. Onar, *Energy Harvesting: Solar, Wind, and Ocean Energy Conversion Systems*. Boca Raton, FL, USA: CRC Press, 2017.
- [9] P. Spies, M. Pollak, and L. Mateu, *Handbook of Energy Harvesting Power Supplies and Application*. Boca Raton, FL, USA: CRC Press, 2015.
- [10] O. Hersent, D. Boswarthick, and O. Elloumi, *The Internet of Things: Key Applications and Protocols*. Hoboken, NJ, USA: Wiley, 2012.
- [11] H. Davies, *Beginning LoRa Radio Networks with Arduino: Build Long Range, Low Power Wireless IoT Networks*. Berkeley, CA: Apress, 2020.
- [12] O. Khutsoane, A. Abu-Mahfouz, and B. Isong, "IoT devices and applications based on LoRa/LoRaWAN," in *Proc. IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, South Africa, Nov. 2017, doi: 10.1109/IECON.2017.8217061.
- [13] "Asset Monitoring Solutions," Radiant RFID Blog. [Online]. Available: <https://radiantrfid.com/blog/asset-monitoring-solutions/>. [Accessed: Jun. 12, 2024]
- [14] A. J. Pansini, in *Electrical Distribution Engineering*, 3rd ed., Lilburn, GA: Fairmont Press, 2007, pp. 382–385.
- [15] Filho João Mamede, in *Manual de equipamentos elétricos*, 4th ed., Rio de Janeiro: LTC, 2013, pp. 136–149.
- [16] Y. ASLAN and S. Türe. Location of faults in power distribution laterals using superimposed components and programmable logic controllers. *Electrical Power and Energy Systems*, vol. 33, pp. 1003-1011, 2011.
- [17] ANEEL (National Electric Energy Agency). *Electricity Distribution Procedures in the national electrical system. PRODIST Module 8 - Quality of Electric Power*. 2021.
- [18] M. Yaseen et al., "A Review of Supercapacitors: Materials Design, Modification, and Applications," *Energies*, vol. 14, no. 22, p. 7779, 2021. doi: 10.3390/en14227779.

SANDRO: Um Sistema Autônomo de Distribuição de Conteúdo Educacional para Comunidades Ribeirinhas da Amazônia

André Jordan Santos

Universidade Federal do Pará - UFPA

Belém, Brasil

andre.santos@itec.ufpa.br

Matheus Pantoja

Universidade Federal do Pará - UFPA

Belém, Brasil

matheus.pantoja@itec.ufpa.br

Albert dos Santos

Universidade Federal do Pará - UFPA

Belém, Brasil

albert.santos@itec.ufpa.br

Diego Lisboa Cardoso - UFPA

Universidade Federal do Pará

Belém, Brasil

diego@ufpa.br

Mauro Margalho

Universidade da Amazônia - UNAMA

Belém, Brasil

mauro@margalho.pro.br

Abstract—Este artigo aborda o acesso limitado à tecnologia em comunidades ribeirinhas da Amazônia. Durante a pandemia de COVID-19, a falta de conectividade dificultou a educação nessas áreas. Propomos um sistema autônomo baseado em Raspberry Pi em conjunto com o módulo RTC DS3231, projetado para distribuir conteúdo educacional sem depender de redes convencionais ou energia estável. Testes demonstraram a viabilidade da solução, com distribuição de vídeos e auto liga/desliga funcionando com sucesso. Esta abordagem promove o acesso ao conhecimento e o desenvolvimento sustentável em comunidades isoladas.

Index Terms—Raspberry Pi, RTC DS3231, Automação, Educação

I. INTRODUÇÃO

No atual cenário global, marcado por avanços tecnológicos exponenciais, as disparidades no acesso à informação e à tecnologia se tornam cada vez mais evidentes. Em regiões remotas, como as comunidades ribeirinhas da Amazônia, essas disparidades são profundamente sentidas. O isolamento geográfico, aliado à ausência de infraestrutura básica de conectividade e a um conhecimento tecnológico limitado, cria um abismo digital que impede essas comunidades de acessar o conhecimento e as oportunidades que a era digital oferece.

A realidade desafiadora dessas comunidades se tornou ainda mais evidente durante a pandemia de COVID-19, que destacou a necessidade crítica de conectividade para a educação. Enquanto o mundo se adaptava ao ensino à distância, estas comunidades encontraram-se em uma desvantagem ainda maior, evidenciando a urgência de abordar essas disparidades para garantir o direito à educação e ao desenvolvimento sustentável.

Diante do desafio de democratizar o acesso ao conhecimento em comunidades isoladas, sem infraestrutura tecnológica adequada, torna-se imperativo explorar soluções tecnológicas inovadoras e adaptáveis. Este trabalho propõe a investigação e desenvolvimento de sistemas autônomos e sustentáveis, projetados especificamente para operar em ambientes desprovidos

de infraestrutura tecnológica convencional. A proposta centra-se na criação de um sistema de baixo custo e energeticamente eficiente, baseado em Raspberry Pi em conjunto com o módulo RTC DS3231 [1], que permite a distribuição autônoma de conteúdo educacional, operando independentemente de redes de comunicação tradicionais e fontes de energia estáveis. A viabilidade e eficácia dessa abordagem são demonstradas no projeto SANDRO [2], uma iniciativa voltada para facilitar o acesso à educação em comunidades ribeirinhas.

II. TRABALHOS CORRELATOS

A pesquisa sobre sistemas autônomos de distribuição de conteúdo educacional tem avançado significativamente, com diversas abordagens sendo exploradas para superar os desafios de conectividade e gestão de energia em regiões remotas. Nesta seção, apresentamos uma revisão de alguns trabalhos relevantes nesta área, destacando suas contribuições e metodologias.

Em [3], os autores têm como objetivo discutir o uso do Raspberry Pi como uma plataforma disruptiva para a implementação de sistemas automatizados. A implementação envolveu o uso do Raspberry Pi em diversos projetos de educação e indústria, destacando sua versatilidade e baixo custo. Os resultados mostraram que o Raspberry Pi promove a inovação tecnológica em áreas com recursos limitados, facilitando a adoção de novas tecnologias em ambientes diversos.

Em [4], o objetivo foi desenvolver um sistema de automação para salas de aula utilizando o Raspberry Pi. A implementação do sistema envolveu o uso de sensores de movimento e reconhecimento facial para gerenciar a presença dos alunos e controlar o consumo de energia dos aparelhos eletrônicos. Os resultados demonstraram uma otimização significativa no uso de recursos, tornando o ambiente de aprendizagem mais eficiente e sustentável.

Em [5], os autores exploram uma abordagem baseada em IoT para automação residencial, com o objetivo de adaptar essa tecnologia para cenários educacionais em regiões remotas. A implementação incluiu a integração de sensores e módulos de comunicação com o Raspberry Pi para criar sistemas autônomos que operam independentemente de redes tradicionais. Os resultados indicaram que tais sistemas são eficazes para a distribuição autônoma de conteúdo educacional em áreas com infraestrutura limitada.

Em [6], o objetivo foi desenvolver um projeto de automação de sala de aula utilizando o Raspberry Pi para reduzir o consumo de energia e melhorar a gestão de presença dos alunos. A implementação focou na aplicação de tecnologias acessíveis e eficientes, incluindo o uso de reconhecimento facial e sensores de movimento. Os resultados mostraram uma redução significativa no consumo de energia e melhorias na gestão de recursos educacionais.

Em [7], o objetivo foi estudar o uso de IoT para monitoramento e controle de dispositivos em ambientes adversos. A implementação utilizou o módulo RTC DS3231 para gerenciamento de energia em sistemas embarcados, demonstrando a robustez e adaptabilidade desses sistemas. Os resultados confirmaram a viabilidade das tecnologias IoT para melhorar o acesso à informação e à educação em regiões remotas.

Embora os trabalhos mencionados forneçam soluções eficazes para a automação e distribuição de conteúdo em ambientes com infraestrutura limitada, a maioria deles não aborda a integração de múltiplas tecnologias para criar um sistema totalmente autônomo que pode operar independentemente de redes de comunicação tradicionais. O sistema SANDRO visa preencher essa lacuna, oferecendo uma solução integrada que combina automação, gerenciamento eficiente de energia e distribuição autônoma de conteúdo educacional, especificamente adaptada para comunidades ribeirinhas da Amazônia.

III. MATERIAIS E MÉTODOS

Nesta seção, serão detalhados o sistema proposto e os materiais utilizados no contexto do cenário adotado.

A. Modelo do Sistema

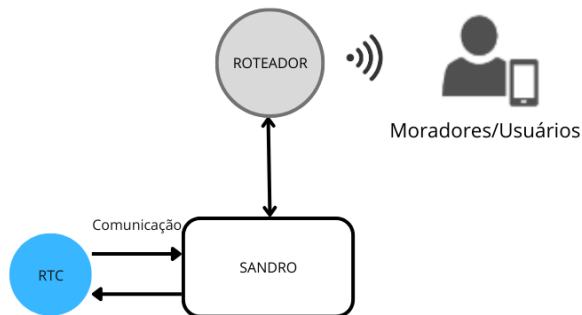


Fig. 1. Visão Geral do SANDRO

A Figura 1 apresenta uma visão geral do sistema SANDRO. O sistema funciona como um servidor offline, armazenado com conteúdos voltados à educação e também relacionados à cultura local, o sistema envia esses dados por meio de um roteador. Isso permitirá que os moradores acessem os conteúdos disponíveis através da rede Wi-Fi, proporcionando um acesso fácil e conveniente às informações e serviços oferecidos pelo sistema.

O sistema utiliza um Raspberry Pi 4 com 8 GB de memória RAM, em conjunto com um roteador externo da Aruba. O SANDRO está conectado a uma bateria alimentada por uma placa solar, garantindo autonomia energética. Além disso, o sistema inclui um módulo RTC DS3231, que opera na faixa de tensão de 3,3V e está conectado diretamente ao Raspberry Pi.

O módulo RTC possibilita a atualização do relógio do sistema, uma vez que o Raspberry Pi normalmente depende de uma conexão com a internet para essa função onde no cenário do projeto, o sistema operará totalmente offline. Com o módulo RTC, o sistema pode funcionar de forma autônoma, ligando e desligando automaticamente em horários programados. Isso garante que os moradores, muitos dos quais não têm conhecimento técnico, possam aproveitar os conteúdos oferecidos de maneira fácil e conveniente.

Para o desenvolvimento do serviço oferecido pelo sistema SANDRO, foram instalados diversos serviços de servidor no Raspberry Pi, incluindo Apache e PHP. O Apache permite o gerenciamento dos serviços HTTP disponíveis, enquanto o PHP oferece suporte à programação do lado do servidor. Além disso, foi integrado um banco de dados MySQL para armazenar os conteúdos educativos.

A funcionalidade completa do sistema SANDRO foi implementada utilizando o serviço de streaming YouPHPTube [8]. O YouPHPTube é uma plataforma de vídeos open-source que permite a personalização do serviço de streaming. Ele acessa todas as informações armazenadas no banco de dados MySQL local, permitindo que os moradores assistam aos conteúdos sem a necessidade de conexão com a internet.

Com essa configuração, o sistema SANDRO fornece um servidor offline robusto, onde os usuários podem acessar e assistir aos conteúdos educativos de maneira fácil e conveniente através da rede Wi-Fi.

B. Modelo para Automação

Para gerenciar a autonomia do sistema, foram realizadas pesquisas visando identificar uma solução prática e econômica. Nesse contexto, identificou-se o módulo RTC DS3231 como uma opção viável. Este módulo opera com uma faixa de tensão de alimentação de 3,3 a 5V e é utilizado como um relógio digital, além de contar com a funcionalidade de alarme.

No contexto do projeto, o módulo RTC é de importância crucial para o funcionamento autônomo do sistema. A função de alarme do módulo será empregada para estabelecer um cronograma de inicialização do Raspberry Pi. O alarme pode ser programado por meio de software, utilizando a comunicação I2C através dos pinos GPIO 2 e SDA (Serial

Data) do Raspberry Pi e do módulo, respectivamente, e também os pinos SCL (Serial Clock) de ambos dispositivos para o gerenciamento do clock da comunicação. O pino de SCL é utilizado como um meio de inicialização do sistema quando o Raspberry Pi já está energizado, atuando como uma chave. O pino de alarme do módulo estará em estado HIGH quando o alarme estiver configurado e mudará para o estado LOW quando o alarme for acionado, iniciando assim o sistema do Raspberry Pi.

Cabe destacar que o alarme do módulo não se reinicia automaticamente. Portanto, assim que o sistema é iniciado, o código desenvolvido permite reiniciar o alarme, garantindo seu funcionamento nos dias subsequentes. A Figura 2 mostra o módulo RTC DS3231.



Fig. 2. Módulo RTC DS3231

O programa para gerenciamento do módulo foi desenvolvido em Python utilizando as bibliotecas RPi.GPIO [9] e smbus [10]. A biblioteca smbus facilita a comunicação entre o Raspberry Pi e o RTC através da interface I2C. Inicialmente, foi criado um algoritmo para configurar o horário do RTC com base na hora atual e para determinar o momento de ligar o Raspberry Pi. A configuração do horário e do alarme do RTC é realizada enviando uma codificação BCD(Binary-Coded Decimal), exigindo a conversão de valores decimais para BCD antes do envio. O BCD é um tipo de codificação onde cada dígito é convertido para um código binário separado.

Adicionalmente, foi criado um cronograma para o desligamento seguro do sistema, visando o gerenciamento energética e prevenção de falhas no sistema operacional do Raspberry Pi. O programa foi incorporado aos arquivos de inicialização do sistema, garantindo sua execução autônoma. A Figura 3, demonstra o funcionamento do sistema automático para ligar e desligar do sistema SANDRO em conjunto com o módulo RTC DS3231.

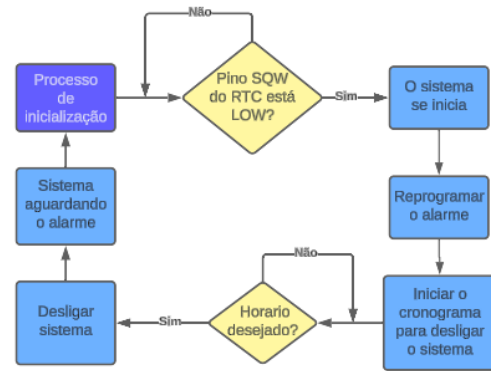


Fig. 3. Diagrama do funcionamento autônomo do sistema

IV. RESULTADOS

A Figura 4 demonstra o sistema SANDRO desenvolvido. O Raspberry Pi é alojado em uma estrutura com um modelo físico antropomórfico, projetada para se aproximar dos moradores locais. Esse design visa facilitar a aceitação e a popularidade do sistema entre a comunidade, promovendo seu uso para fins educativos.



Fig. 4. SANDRO

Através do serviço open-source YouPHPTube, foi possível personalizar o streaming local offline. O serviço acessa o site desenvolvido em PHP, juntamente com o banco de dados local disponível no Raspberry Pi. Esse portal, destinado principalmente a conteúdos educativos, permite que os usuários acessem as informações utilizando seus smartphones via a rede Wi-Fi disponibilizada pelo sistema SANDRO, o primeiro protótipo do sistema foi instalado na Região das Ilhas em

Abaetetuba (PA). As Figuras 5 e 6 mostram o portal desenvolvido.

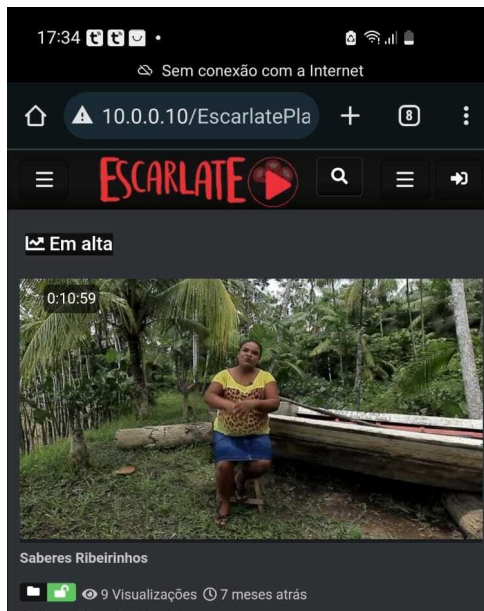


Fig. 5. Portal do SANDRO

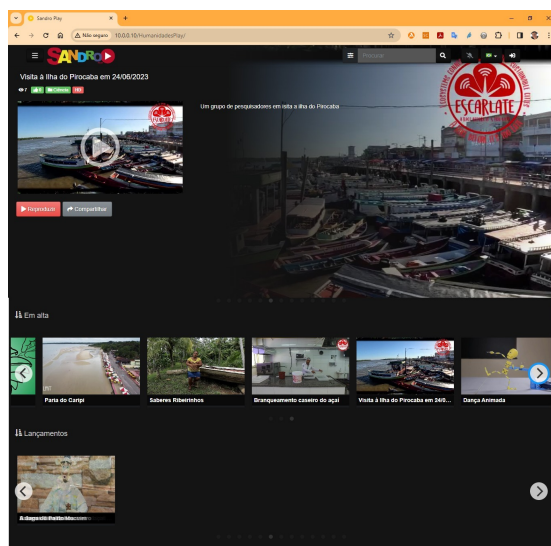


Fig. 6. Portal do SANDRO

O protótipo do circuito de automação de funcionamento de ligar e desligar do Sandro pode ser observado na Figura 7. Esse protótipo utiliza um Raspberry Pi Model B, um módulo RTC e um transistor PNP. O transistor é utilizado para manter o funcionamento da interface de comunicação I2C, pois tanto o pino do alarme quanto o pino SCL do módulo compartilham o mesmo pino SCL (GPIO5) no Raspberry Pi.

A partir do circuito, foi possível criar uma rotina para testar o sistema de liga e desliga automatizado. Quando o sistema é iniciado, um arquivo CSV é gerado para coletar dados de status do sistema, incluindo os horários de início do

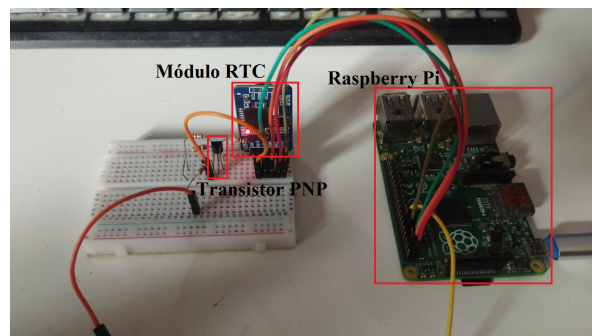


Fig. 7. Protótipo para automação do sistema SANDRO

sistema. O sistema foi programado para operar continuamente por trinta minutos. Antes de desligar, o alarme no RTC é configurado para reiniciar o sistema após um intervalo de dois minutos. Observa-se que o processo de inicialização do sistema leva aproximadamente quarenta segundos devido ao tempo necessário para o dispositivo carregar o sistema operacional. Essa rotina de ligar e desligar foi testada ao longo de aproximadamente três horas, os testes com o circuito foram feitos em laboratório sem a utilização de placa solar e bateria. A Figura 8 mostra os resultados coletados durante esse período.

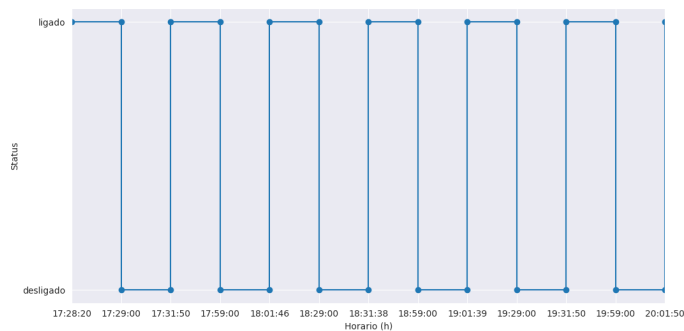


Fig. 8. Gráfico de status do sistema autônomo

Com um circuito de automação, é possível criar uma rotina de funcionamento para o sistema SANDRO, permitindo que os moradores possam facilmente acessar e utilizar os recursos oferecidos, mesmo sem conhecimento prévio de eletrônica.

V. CONCLUSÃO

Este estudo aborda o acesso limitado à tecnologia em comunidades ribeirinhas da Amazônia. Durante a pandemia de COVID-19, a falta de conectividade dificultou a educação nessas áreas. Propomos um sistema autônomo baseado em Raspberry Pi em conjunto com o módulo RTC DS3231, projetado para distribuir conteúdo educacional sem depender de redes convencionais ou energia estável. Testes demonstraram a viabilidade da solução, com distribuição de vídeos e auto liga/desliga funcionando com sucesso. Esta abordagem promove o acesso ao conhecimento e o desenvolvimento sustentável em comunidades isoladas.

Para trabalhos futuros, está prevista a implementação de um sistema de atualização dos conteúdos fornecidos pelo sistema SANDRO. Esse sistema aproveitará o trânsito de barcos feito pelos moradores. Cada barco terá um dispositivo Raspberry Pi operando em uma rede ad hoc. Quando um dispositivo reconhecer outro, ele transmitirá os conteúdos, e esse processo se repetirá sucessivamente para todos os outros dispositivos carregados pelos barcos, até que os conteúdos atualizados cheguem ao destino final do sistema SANDRO.

Também, a utilização de um conversor analógico-digital (ADC) para obter amostras da tensão da bateria. Isso possibilitará o monitoramento energético e o aprimoramento do processo de automação. Além disso, planeja-se implementar todo esse sistema automatizado no SANDRO, assim como realizar melhorias contínuas no processo de automação.

REFERENCES

- [1] AnalogDevice, “Ds3231,” Disponível em: <https://www.analog.com/media/en/technical-documentation/data-sheets/ds3231.pdf>, 2015.
- [2] M. Margalho, “Humanidades,” Disponível em: <https://www.margalho.pro.br/site-humanidades/index.html>, 2023.
- [3] S. J. Johnston and S. J. Cox, “The raspberry pi: A technology disrupter, and the enabler of dreams,” p. 51, 2017.
- [4] S. Kurkovsky and C. Williams, “Raspberry pi as a platform for the internet of things projects: Experiences and lessons,” in *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, 2017, pp. 64–69.
- [5] K. Venkatesh, P. Rajkumar, S. Hemaswathi, and B. Rajalingam, “IoT based home automation using raspberry pi,” *J. Adv. Res. Dyn. Control Syst.*, vol. 10, no. 7, pp. 1721–1728, 2018.
- [6] B. S. Srikar, S. Vijayaraghavan, D. Lakshman, and A. H. V. Kumar, “Smart class automation using raspberry pi.”
- [7] A. Zharkimbekova, A. Ospanova, K. Sagindykov, and M. Kokkoz, “Implementation and commercialization of the results of the “multidisciplinary mobile computer classroom based on raspberry pi” project,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 13, pp. 116–136, 2020.
- [8] AVideo, “Avideo,” Disponível em: <https://youphp.tube>, 2017.
- [9] B. Croston, “Rpi.gpio,” Disponível em: <https://pypi.org/project/RPi.GPIO/>, 2012.
- [10] K.-P. Lindegaard, “smbus2,” Disponível em: <https://pypi.org/project/smbus2/>, 2016.

Low-cost configurable data acquisition module design for real-time monitoring

Vitória J. Baratella, Anderson G. de Almeida, Caio T. B. Costa, Décio R. M. Faria, Danilo H. Spadoti

Abstract—This article consists of a low-cost data acquisition module for real-time monitoring with a configurable interface for selecting reading parameters. In this work, the complete development of the module hardware and tests in real applications are presented. The results demonstrated good module response even when tested in noisy environments, in addition to meeting the low cost requirement, costing approximately 26 dollars, 87% cheaper than commercial options. Therefore, the module makes up an accessible solution for low-cost applications for research and industry.

Data acquisition system (DAQ), instrumentation, embedded system, signal processing.

I. INTRODUCTION

Data acquisition is used in practically all everyday activities. In this context, it is clear that traditional analogue technologies are becoming obsolete, as users prefer innovative digital measurement techniques [1]. Thus, data acquisition systems (DAQ) are increasingly essential in the process of collecting real-world information, storing and analyzing data to use it as a basis for decision-making that results in process improvements [2]. To develop a data acquisition system, you need: a real-world phenomenon, sensors, signal conditioning, hardware for data acquisition and sensor control, dedicated software and communication interfaces [3].

According to research study The Data Acquisition Market Forecast [4], the data acquisition market was valued at 1.96 billion in 2019 and is expected to reach 2.73 billion by 2025 [4]. Therefore, the growing need to monitor parameters, adopt IoT solutions and perform industrial controls are current market trends that drive the demand for data acquisition systems [4]. These systems are also highly valued in academia, as teaching and research require DAQs to demonstrate practical applications of theoretical concepts [5]. However, the high cost of available systems makes it necessary to develop more affordable options.

II. PROPOSAL

The proposal consists of presenting the development of a versatile and configurable data acquisition module, which allows reading a wide range of sensors and data collection in noisy environments and other sources of interference. The module was designed to be low-cost and easy to use, making it an accessible solution for a wide variety of applications.

The authors thank CAPES, CNPQ -402093/2022-4, FAPEMIG, FAPESP 2021/06569-1 and the entire team at LabTel-LA IoT UNIFEI for their support to the research.

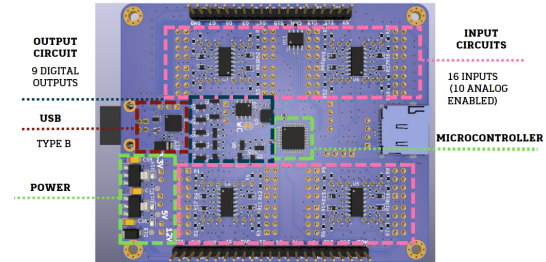


Fig. 1. Developed printed circuit board

The implementation of the data acquisition module was based on signal conditioning processes using a microcontroller, selected for its signal filtering capacity, gain, processing capability, and flexibility. Additionally, the approach was systematic and modular, allowing the addition of resources according to user needs.

The module is capable of reading different types of sensors, such as temperature, humidity, pressure, accelerometers, and gyroscopes, and can provide accurate and reliable data even in noisy environments and other sources of interference.

III. DEVELOPMENT

For the hardware development of the module, design requirements were defined that guided the choice of the microcontroller and other components. Thus, this section presents the operational circuit, power supply and system protection, input and accessory circuits. The printed circuit board layout is shown in Figure 1.

A. Power Supply

The power supply was designed for a nominal input voltage of 12V, with an operational range from 7V to 15V. At voltages lower than 12V, the signal will be attenuated but will continue to operate.

As the microcontroller and other components require 3.3V, the LM317 regulator was used with the appropriate resistor configuration. The LM317 was chosen for its high input-output voltage differential capability (up to 40V), providing significant protection against voltage pulses common in applications such as automotive environments [6]. Additionally, protection was ensured through the use of TVS diodes, reverse polarity diodes, and Zener diodes for both positive and negative over-voltages.

The circuit is presented in Figure 2.

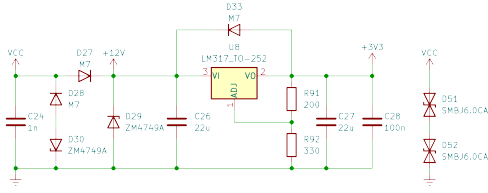


Fig. 2. Power Supply Circuit

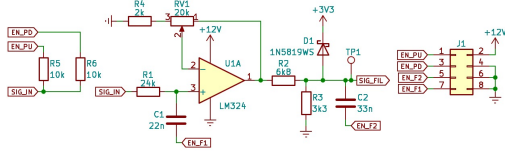


Fig. 3. Input Circuit

B. Input Circuit Design

1) *Requirements:* To read the signal, the following criteria were taken into consideration:

- High input impedance: The circuit must have low current from the signal to minimize line losses;
- Overvoltage protection: The circuit should not suffer damage to any of its parts in the event of an overvoltage, requiring protection that can withstand inputs up to 15V;
- High-frequency filtering: The circuit must filter out noise and remove signals with frequencies higher than specified by the Nyquist criterion. The cutoff frequencies used were 100Hz or 700Hz, allowing selection in the module.
- 5% attenuation tolerance in the passband;
- Acceptance of signals between 1-10V: The circuit must condition an input signal amplitude of 1 to 10V with the maximum resolution of the microcontroller;
- Maximum current of 1mA: The maximum current consumption of the input signal for high frequencies must be equal to or less than 1mA, due to the low-pass filter;
- Controllable pull-up and pull-down: It should be possible to add pull-up or pull-down at the signal input;

The input circuit is presented in Figure 3.

2) *Input Signals:* This conditioner was developed to accept both analog and digital signals, allowing for 3 levels of filtering, i.e., 3 filter configurations: 20kHz, 700Hz, and 100Hz, enabling noise reduction and greater precision. Each level is configured by a pin connector, as shown in Figure 3, so the filter is selected depending on which pins are connected. In total, 16 conditioning circuits were implemented, all accepting digital signal inputs, and 10 accepting analog signals.

The first level is selected when both EN_F1 and EN_F2 are not connected to GND, which disconnects capacitors C1 and C2 in Fig. 3, resulting in the behavior of only the operational amplifier. Analyzing the datasheet of the chosen operational amplifier, it has a maximum frequency of 25kHz due to its slow rate.

The second level is selected when EN_F2 is connected to GND and EN_F1 is disconnected. This way, a cutoff frequency of 700Hz is set. The third level is selected when both EN_F1 and EN_F2 are connected to GND. This way, a cutoff frequency of 100Hz is set. This circuit is able to read signals with amplitudes between 10V and 1V with the maximum resolution of the microcontroller's analog-to-digital converter.

3) *Operational Amplifier:* The operational amplifier is the central component of the input circuit, used to amplify the signal and provide high impedance at the signal inputs.

To select an appropriate operational amplifier for the project, it was considered necessary for it to support a 12V supply voltage and the established signal frequencies, while balancing the lowest possible cost.

Thus, an LM324DT quad-channel operational amplifier was chosen, which met all the specified requirements.

a) *High Impedance:* To ensure that the current from the signal is minimal, thereby minimizing losses in the line, it is necessary to have a high impedance at the signal input. For this purpose, the input of an operational amplifier was used.

Due to the structure of the circuit used, the input impedance also has a capacitance in parallel, which causes its impedance to change depending on the signal frequency. The desired maximum current will define the minimum value of the series resistance.

b) *Gain Control:* To have input signals between 10V and 1V, the operational amplifier configured to act as a non-inverting amplifier is used. This circuit uses a potentiometer to adjust its gain. The circuit should operate with the output of the operational amplifier at 10V when the input has its highest value, regulated by the potentiometer. In the presented circuit, a series resistor (R4) is placed with the potentiometer to prevent high gain when the potentiometer is set to an extreme.

To ensure that the gain operates within the specified range, it is necessary to determine the relationship between the resistance R4 and the potentiometer RV1. For this, the situation in which the potentiometer is at its extreme, meaning the resistance is entirely on the output side of the operational amplifier, resulting in its highest gain, is considered.

Assuming that there is no current at the input of the operational amplifier, and replacing the value of R4 with R and RV1 with $m \cdot R$, where m is the ratio between the potentiometer and R4, we have:

$$I = \frac{V_i}{R} = \frac{V_o - V_i}{mR} \quad (1)$$

$$V_i \times m = V_o - V_i \quad (2)$$

$$V_o = V_i \times (m + 1) \quad (3)$$

Thus, the maximum gain will be $(m+1)$. Knowing that the minimum input is 1V and the operational amplifier output must be 10V, the minimum value of m can be defined:

$$10 \leq 1 \times (m + 1) \quad (4)$$

$$m \geq 9 \quad (5)$$

Therefore, knowing this relationship, it is necessary to define its resistance. In the extreme case, the potentiometer will have a voltage difference of 10V, and the power consumption will be up to 5mW. The power consumption value is justified due to the need to minimize the current consumption of the operational amplifier feedback circuit. Thus, the minimum value of RV1 can be calculated.

$$P = \frac{V^2}{R} \quad (6)$$

$$5 \times 10^{-3} \leq \frac{10^2}{R_{V1}} \quad (7)$$

$$R_{V1} \geq 20k\Omega \quad (8)$$

A commercial potentiometer of 20k Ω was used. The resistance R4 can be found from the value of m.

$$m \geq 9 \quad (9)$$

$$mR \geq 20000 \quad (10)$$

$$R < 2.2k\Omega \quad (11)$$

Selecting a commercial value that matches the calculations, R4 was chosen to be 2k Ω and RV1 to be 20k Ω .

To ensure that the microcontroller can read the signal without being damaged, it is necessary to reduce this 10V value to 3.3V, the microcontroller's operating voltage, and for this purpose, a voltage divider composed of R2 and R3 is used, as shown in Figure 3.

c) *Overvoltage Protection:* To prevent errors in the potentiometer calibration or the signal voltage from exceeding the expected value, surpassing the maximum voltage supported by the microcontroller, a Schottky diode (D1) functioning as a clamp was used.

This way, the circuit limits the voltage at the microcontroller to 3.3V and takes into account the voltage drop across the diode, thus, a Schottky diode is ideal due to its low voltage drop.

According to the microcontroller datasheet, the maximum voltage allowed by the I/O ports is 4V, with a margin of 700mV. Thus, the chosen diode was the 1N5819WS, which has a drop of 300mV with a current of 100mA.

d) *Filtering Circuit:* To perform filtering, two stages of low-pass filters were used, as shown in Figure 3, and each stage must have a different cutoff frequency, as defined earlier.

Both stages are first-order systems that can be simplified to:

$$G(s) = K \frac{\omega_n}{s + \omega_n} \quad (12)$$

Given the attenuation limit, and ignoring the gain generated by K, the natural frequency at which a 5% tolerance is adopted can be found. Thus:

$$|G(i\omega_c)| = 0.95 \quad (13)$$

$$\omega_n = \omega_c \times \frac{19}{\sqrt{39}} \quad (14)$$

With this, it is possible to calculate the necessary values of resistors and capacitors to perform the filter. The first stage is made with components R1 and C1 from Figure 3, building a low-pass filter with a gain of 1. To calculate the transfer

function of this circuit, it is assumed that the current at the input of the operational amplifier is zero, and the calculation is done as a voltage divider, thus:

$$V_{oF1} = V_i \frac{\frac{1}{sC}}{R + \frac{1}{sC}} \times \frac{sC}{sC} = V_i \frac{1}{sRC + 1} \quad (15)$$

$$\frac{V_{oF1}}{V_i} = G_{F1}(s) = \frac{\frac{1}{sC}}{s + \frac{1}{RC}} = K \times \frac{\omega_n}{s + \omega_n} \quad (16)$$

$$\begin{cases} K = 1 \\ \omega_n = \frac{1}{RC} \end{cases} \quad (17)$$

The first stage will have a cutoff frequency of 100Hz, using the previous formula to calculate the natural frequency:

$$\omega_n = \omega_c \times \frac{19}{\sqrt{39}} = 2\pi \times 100 \times \frac{19}{\sqrt{39}} \quad (18)$$

$$RC \cong 5.231 \times 10^{-4} \quad (19)$$

Taking into account the requirement of a maximum signal current of 1 mA, it is necessary to calculate the equivalent impedance of the circuit:

$$Z_{eq}(\omega) = R + \frac{1}{i\omega C} \quad (20)$$

The highest current will occur with the lowest impedance. Given the impedance equation, it will have the lowest value when the frequency tends to infinity, and the equivalent impedance will be R. As the maximum signal voltage is 12V, the maximum current will be:

$$I_{max} = \frac{V_{max}}{Z_{min}} = \frac{12}{R} \leq 10^{-3} \quad (21)$$

$$R \geq 12000 \quad (22)$$

Analyzing the values of commercial resistors and capacitors, R1 was chosen to be 24k Ω and C1 to be 22nF, resulting in $RC = 5.28 \times 10^{-4}$. The second stage uses components R2, R3, and C2 for its operation. For the calculation of its transfer function, the currents at the nodes were used. Additionally, R2 is used as R, R3 as m*R, and C2 as C.

$$I = \frac{V_i - V_o}{R} \quad (23)$$

$$sV_o C + \frac{V_o}{mR} = I \quad (24)$$

$$\frac{V_o}{V_i} = G(s) = \frac{m}{smRC + m + 1} \quad (25)$$

$$= K \times \frac{\omega_n}{s + \omega_n} \quad (26)$$

$$K = \frac{m}{m + 1} \quad (27)$$

As specified earlier in Gain Control (3.b), the voltage divider must reduce the voltage from 10V to 3.3V, thus:

$$K = \frac{3.3}{10} = \frac{m}{m + 1} \quad (28)$$

$$m \cong 0.493 \quad (29)$$

The second stage should have a cutoff frequency of 700Hz.

$$RC \cong 2.265 \times 10^{-4} \quad (30)$$

As the operational amplifier used delivers a minimum output current of 10mA according to its datasheet, the minimum value

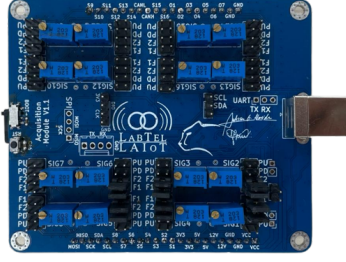


Fig. 4. Assembled Printed Circuit Board

of R can be defined.

$$I_{max} \leq 10 \times 10^{-3} = \frac{V_{max}}{R} \quad (31)$$

$$10 \times 10^{-3} \geq \frac{10}{R} \quad (32)$$

$$R \geq 10^3 \quad (33)$$

Thus, the following values were chosen: $R2 = 6.8k\Omega$, $R3 = 3.3k\Omega$, and $C2 = 33nF$.

C. Accessory Circuits

Additionally, several auxiliary circuits were implemented to enhance the module's functionality, including: output Ports, there were implemented 9 output circuit with 1A max current supply; USB port, used to calibrate and read ports and sensors; CAN bus, enabling expansion and communication with other boards; and accelerometer, for onboard applications.

IV. RESULTS

To analyze the behavior of the module and the circuits implemented on it, test procedures were performed in a laboratory setting using appropriate equipment. Below is Figure 8, showing the module manufactured.

A. Input Circuit Test

1) *Test Procedure:* For the correct reading of the signals, measurements were made regarding the filters implemented in the input circuit. To do this, signal sweeps were performed at the 3 filtering levels, i.e., at the 3 filter configurations: 20kHz, 700Hz, and 100Hz.

The signals inserted consisted of sine waves at various frequencies, starting with 1Hz and 5 points per decade. The frequency was increased respecting the limit of 500mV of the output voltage amplitude.

The DSO-X 2002A oscilloscope was used to generate sine waves and measure the peak-to-peak amplitude of the input and output voltages, as well as the frequency of the signals.

2) *Acquired Data:* For the first level, signals F1 and F2 were disconnected to achieve the widest passband, resulting in the operational amplifier behaving only as a filter. The resulting data can be observed in the Amplitude Bode Diagram in Figure 5:

Since an operational amplifier is used at its operating limit, the transformation distorts the signal frequencies.

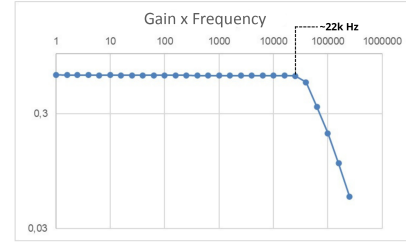


Fig. 5. Gain vs. Frequency

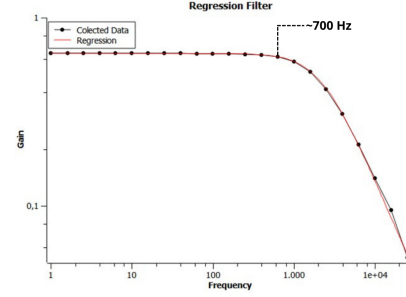


Fig. 6. Filter Regression (Gain vs. Frequency)

Therefore, to define the highest frequency without distorting the signal, the maximum point at which the amplifier can operate without distortion was analyzed, reaching the value of 22kHz.

In the second level, a single-stage filtering was performed with the filter in signal F2, generating the respective Amplitude Bode Diagram.

This signal does not undergo changes due to distortion caused by the operational amplifier since it has a low filtering frequency compared to the frequency analyzed by the amplifier.

To perform the regression, the transfer equation of a first-order filter was used, as follows:

$$|G(s)|_{s=i\omega} = \left| \frac{K\omega_n}{s + \omega_n} \right|_{s=i\omega} = \frac{K\omega_n}{\sqrt{\omega^2 + \omega_n^2}} \quad (34)$$

Applying the formula above to the regression yields the graph in Figure 6:

The following parameters were obtained:

$$f(x) = \left| \frac{a \times b}{\sqrt{x^2 + b^2}} \right| \quad (35)$$

$$\begin{cases} a = 0.645642713362156 \pm 0.12\% \\ b = 2153.85257996398 \pm 0.69\% \end{cases} \quad (36)$$

With this natural frequency, the cutoff frequency with 5% attenuation is also found, disregarding the fixed gain a (of 700Hz), achieving a result within the calculated value.

At the third and final level, a single-stage filtering was performed with the filter in signal F1 and F2, generating the Amplitude Bode Diagram.

To perform the regression, it is necessary to apply two stages, as the cutoff frequency of the previous stage is close:

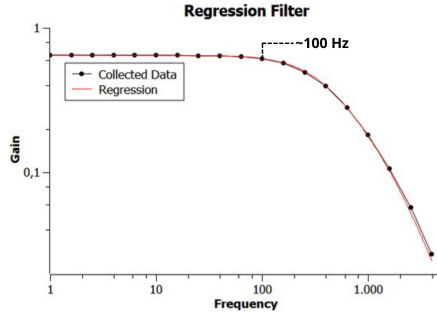


Fig. 7. Filter Regression (Gain vs. Frequency)

$$|G(s)|_{s=i\omega} = \left| \frac{K\omega_{n1}}{i\omega + i\omega_{n1}} \frac{\omega_{n2}}{i\omega + i\omega_{n2}} \right| = \frac{K\omega_{n1}}{\sqrt{\omega^2 + \omega_{n1}^2}} \frac{\omega_{n2}}{\sqrt{\omega^2 + \omega_{n2}^2}} \quad (37)$$

As the third level has the behavior of the second as one of the factors, we used $\omega_{n2} = 2153.853 \text{ rad/s}$. Applying this formula to the regression, we get the graph in Fig. 7:

The following parameters were obtained:

$$f(x) = \frac{a \times b}{\sqrt{x^2 + b^2}} \frac{c}{\sqrt{x^2 + c^2}}$$

$$\begin{cases} a = 0.645642713362156 \pm 0.09\% \\ b = 318.657604794926 \pm 0.74\% \\ c = 2153.85257996398 \pm 0.69\% \end{cases}$$

Again, with this natural frequency, the cutoff frequency with 5% attenuation is also found, disregarding the fixed gain a (of 100Hz), achieving a result within the calculated value.

Therefore, with the obtained data, it was possible to verify the correct functioning of the filters of the acquisition module and demonstrate its proper operation. Additionally, it was possible to experimentally determine the maximum frequency of a signal at 22kHz.

B. Application Example

For real application testing, the module was tested on a validation bench, which consists of an automotive brake system for the Formula SAE team from UNIFEI, named Cheetah Racing. Thus, it was possible to measure: brake disc temperature using an infrared sensor, brake line pressure sensor, applied system load using a load cell and disc rotation with an inductive sensor and tone wheel.

With all the acquired data, it was possible to determine the operating temperature ranges of the disc and pad. The graph of disc temperature as a function of time is presented in Fig. 8.

In the application, a quadratic increase in disc temperature as a function of time was expected. From the analysis of the graph, the correlation of the function can be observed, demonstrating the correct operation of the module.

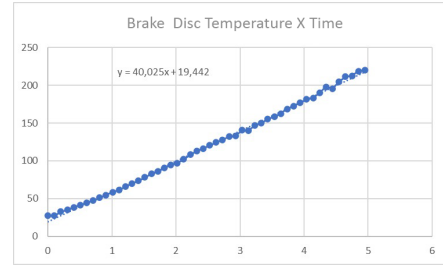


Fig. 8. Disc Temperature vs. Time

C. Module Pricing

As one of the objectives of the project is to develop a module with an affordable cost for the market, the total cost was \$26.05 per data acquisition module: connectors, passive components and power supply (USD 3.45); expansion peripherals and USB (USD 7.18); microcontroller (USD 4.34); input and output circuits (USD 6.79); printed circuit board (USD 4.29).

V. CONCLUSION

The development of the data acquisition module, after all the validation tests performed, showed satisfactory results, as it correctly fulfills all its functionalities.

Additionally, the module demonstrated good response even when tested in noisy environments, positively handling this important issue for the industry.

Commercial modules cost an average of 200 dollars, comparing costs, the developed prototype is approximately 87% cheaper than the commercial ones, meeting the economic requirement proposed in the article. Therefore, it becomes an affordable solution for various applications in research and industry.

Future works can include improvements to the module could include modifications such as changing the op amp to a rail to rail, expanding the input voltages. Implementing wireless communication to enable real time monitoring.

REFERENCES

- [1] M. Haizad, R. Ibrahim, A. Adnan, T. D. Chung, and S. M. Hassan, "Development of low-cost real-time data acquisition system for process automation and control," *2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)*, pp. 1–5, 2016.
- [2] H. Austerlitz, "Data acquisition techniques using pcs," *Academic Press*, vol. 2, no. 1, pp. 01–13, 2002.
- [3] S. Soloman, "Sensors handbook," *McGrawHill*, vol. 2, no. 1, 2009.
- [4] (2023) Data acquisition market size share analysis - growth trends forecasts (2023 - 2028). [Online]. Available: <https://www.mordorintelligence.com/industry-reports/global-data-acquisition-daq-market-industry>
- [5] S. Tehami, M. A. Khan, and O. Mazhar, "Hardware and software designing of usb based plug n play data acquisition device with c and labview compatibility," *2015 IEEE 21st International Symposium for Design and Technology in Electronic Packaging (SIITME)*, pp. 143–146, 2015.
- [6] B. Deutschmann, F. Magrini, and F. Klotz, "Internal ic protection structures in relation to new automotive transient requirements," in *2011 8th Workshop on Electromagnetic Compatibility of Integrated Circuits*, 2011, pp. 47–52.

Otimização Energética em RSMA 6G Utilizando Alocação de Potência via PSO

Reyso Teixeira

Universidade Federal do Pará - UFPA

Belém, Brasil

reyso.teixeira@itec.ufpa.br

Matheus Pantoja

Universidade Federal do Pará - UFPA

Belém, Brasil

matheus.pantoja@itec.ufpa.br

Albert dos Santos

Universidade Federal do Pará - UFPA

Belém, Brasil

albert.santos@itec.ufpa.br

Diego Cardoso

Universidade Federal do Pará - UFPA

Belém, Brasil

diego@ufpa.br

Abstract—À medida que avançamos para o 6G, a eficiência energética (EE) das redes móveis torna-se fundamental devido ao crescimento exponencial na demanda por dados. Este estudo investiga a alocação eficiente de potência em redes RSMA (Rate Splitting Multiple Access), propondo uma solução baseada na meta-heurística PSO (Particle Swarm Optimization). A abordagem visa otimizar a alocação de potência, resultando em melhorias significativas na EE e na redução de bloqueios na rede. Os resultados demonstram a eficácia da combinação RSMA e PSO para redes 6G, aprimorando a eficiência e sustentabilidade das comunicações móveis.

Index Terms—Rate Splitting Multiple Access, Alocação de Potência, Eficiência Energética, Otimização por Enxame de Partículas.

I. INTRODUÇÃO

Impulsionados pelo veloz avanço da tecnologia de comunicação móvel, os sistemas de comunicação sem fio trilham um caminho incessante rumo a taxas de dados cada vez mais elevadas. A rede 6G promete frequências de terahertz além de melhorar a cobertura e a taxa de dados [1]. No entanto, isto exigirá redes mais eficientes energeticamente, fundamentais para a sustentabilidade da infraestrutura de comunicação. É crucial projetar redes com estratégias de otimização de energia, tecnologias de transmissão eficientes e algoritmos inteligentes que maximizem a transmissão de informações com o menor consumo de energia.

A utilização de recursos em redes móveis de próxima geração exige a implementação de técnicas de acesso múltiplo eficientes, as quais facilitam o compartilhamento organizado entre os usuários. O Rate-Splitting Multiple Access (RSMA) se destaca como uma solução promissora para otimizar a eficiência espectral (SE) e energética em sistemas de comunicação sem fio [2]. O RSMA aborda os desafios de conectividade massiva e alta confiabilidade esperados no 6G, dividindo as mensagens dos usuários em partes privadas e comuns, resultando em maior taxa ergódica e menor número de SICs (Successive Interference Cancellation), reduzindo a complexidade [3].

O problema da alocação de potência em redes móveis representa um desafio complexo, pois envolve a determinação pre-

cisa da potência a ser distribuída entre os usuários para garantir uma comunicação eficiente. Esse problema é frequentemente classificado como um problema de otimização combinatória NP-difícil [15]. Em ambientes dinâmicos, encontrar a solução ideal pode ser computacionalmente inviável devido ao tempo necessário para processar a informação. Portanto, é fundamental o desenvolvimento de abordagens que proporcionem soluções aproximadas dentro de limites de tempo aceitáveis, de modo a equilibrar a eficácia na transmissão de dados e a eficiência computacional.

Assim, este trabalho consiste em otimizar a performance de redes RSMA 6G visando a melhor alocação de potência aos usuários conectados, considerando aspectos de Eficiência Energética (EE) de forma a atender a maior quantidade de usuários possível. Para tal abordagem, foi utilizada uma meta-heurística tradicional e bastante referenciada na literatura por sua menor complexidade e rápida convergência, conhecido como PSO (Particle Swarm Optimization). A aplicação de técnicas de otimização bioinspiradas, como o PSO, pode promover significativos ganhos em termos de EE e desempenho geral da rede, ajustando dinamicamente a alocação de recursos para atender às demandas variáveis do ambiente 6G.

II. TRABALHOS CORRELATOS

A literatura científica apresenta trabalhos que destacam cenários de superioridade do RSMA em relação ao Space-Division Multiple Access (SDMA) e Non-Orthogonal Multiple Access (NOMA) em termos de EE. Em [4] utilizou rádios definidos por software (SDR) para otimizar o *throughput* e a alocação de recursos onde foram implementadas as três técnicas de acesso múltiplo, enquanto [7] investigaram o equilíbrio entre SE e EE em sistemas *downlink multiple input single output* (MISO) usando Otimização Sucessiva por Aproximação Convexa (SCA).

Essas pesquisas destacam as vantagens do RSMA em termos de adaptabilidade a diferentes níveis de interferência, cargas de rede variáveis e EE, em comparação às técnicas tradicionais de acesso múltiplo. Os resultados de [4] mostram que o RSMA alcançou maior taxa de transferência total e

melhor equidade na alocação de recursos em comparação com SDMA e NOMA, além de apresentar vantagens em EE. [7] enfatizam o equilíbrio entre SE e EE proporcionado pelo RSMA, tornando-o uma técnica promissora para futuras gerações de sistemas de comunicação sem fio.

Além disso, [5] propuseram uma abordagem para a alocação dinâmica de potência em redes NOMA com Unmanned Aerial Vehicles (UAVs), utilizando o algoritmo PSO. Este método mostrou-se eficaz na otimização da alocação de potência, melhorando a taxa de sum-rate ergódica da rede e superando estratégias de alocação fixa e outros algoritmos meta-heurísticos.

Desta forma, percebe-se que a estratégia do RSMA destaca-se entre as técnicas de acesso múltiplo, bem como o uso de metaheurísticas é promissor no problema de alocação de potência, motivo pelo qual ambos foram objeto de pesquisa deste trabalho.

III. MATERIAIS E MÉTODOS

Nesta seção serão apresentados o cenário adotado bem como são delineadas as suposições e considerações formuladas durante a criação do modelo analítico.

A. Modelo do Sistema

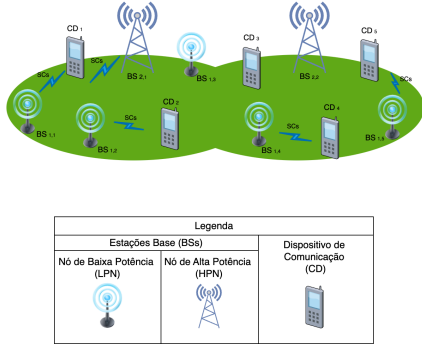


Fig. 1. Rede sem fio considerada com duas camadas de BSs.

O cenário apresentado na figura 1, consiste em uma rede RSMA 6G com M Estação Base (BS) de única antena e K camadas, com $M = [BS_{1,1}, BS_{1,2}, BS_{2,3}, BS_{2,4}, \dots, BS_{K,M}]$. As camadas podem ser do tipo LPN (Low-Power Nodes) ou HPN (High-Power Nodes) sendo $K = [1,2]$ o número de camadas presentes. Cada camada possui uma potência de transmissão máxima p_k^{Max} . A largura de banda W do sistema é dividida igualmente entre as S subportadoras (SC). Ou seja, a largura de banda B_s de cada SC corresponde a $B_s = \frac{W}{S}$. Além disso, é considerado que há U CDs (Communication Devices), com uma única antena, dispostos aleatoriamente no cenário.

De acordo com [13], no RSMA, as mensagens do transmissor são divididas em duas categorias: mensagens comuns e privadas. As mensagens comuns são decodificáveis por todos os usuários, mas ao decodificá-las, as mensagens privadas de outros usuários são consideradas ruído. Na parte privada, após aplicar a SIC para eliminar as mensagens comuns cada usuário

pode decodificar apenas suas próprias mensagens privadas, tratando as dos outros como ruído.

Para um par de usuários em um sistema RSMA, a Relação sinal-ruído (SINR) de cada usuário é calculada considerando a potência de transmissão, o ganho de canal, a potência do ruído e o cancelamento de interferência gerado pela SIC. A SINR da parte comum da mensagem do CD u (sobrescrito c) e da parte privada (sobrescrito p) é calculada de acordo com as Equações 1 e 2 [14]:

$$\gamma_{m,u,s}^{k,c} = \frac{p_{m,u,s}^{k,c} G_{m,u,s}^{k,c}}{\beta I_{INI}^{k,m,u,s} + I_{ICI}^{k,m,u,s} + \alpha^2} \quad (1)$$

$$\gamma_{m,u,s}^{k,p} = \frac{p_{m,u,s}^{k,p} G_{m,u,s}^{k,p}}{I_{INI}^{k,m,u,s} + I_{ICI}^{k,m,u,s} + \alpha^2} \quad (2)$$

sendo que $p_{m,u,s}^{k,c}$ representa a potência de transmissão da mensagem comum do CD u , alocado na BS m com camada k , suportadora s e ganho $G_{m,u,s}^{k,h}$. Além disso, $I_{INI}^{k,m,u,s}$ representa a interferência intra-cluster (INI), $I_{ICI}^{k,m,u,s}$ é a interferência inter-cluster, α^2 é a potência do ruído e β é o fator para simular a SIC e varia no intervalo de $[0,1]$ sendo que $\beta = 0$ é chamado de SIC perfeito e implica em cancelar toda a interferência INI.

A $I_{INI}^{k,m,u,s}$ é a interferência recebida de outros usuários dentro do mesmo cluster, calculada de acordo com a Equação 3:

$$I_{INI}^{k,m,u,s} = \sum_{u'=1, u' \neq u}^U p_{m,u',s}^k G_{m,u',s}^k \quad (3)$$

Essa soma envolve as potências de transmissão de todos os outros usuários no cluster ($p_{m,u'}^k$), exceto o usuário em análise (u), multiplicadas pelos seus respectivos ganhos de canal até o usuário em análise ($G_{m,u,s}^k$). Da mesma forma, a interferência inter-cluster ($I_{ICI}^{k,m,u,s}$) inclui a interferência de todos os usuários em outros clusters, e é dada pela Equação 4:

$$I_{ICI}^{k,m,u,s} = \sum_{k=1}^K \sum_{m'=1, m' \neq m}^M \sum_{u'=1, u' \neq u}^U p_{m',u'}^k G_{m',u}^k \quad (4)$$

Aqui, o cálculo envolve a soma das potências de transmissão ($p_{m',u'}^k$) de usuários em diferentes clusters (m') e diferentes subportadoras, mas ainda impactando o usuário atual (u), e multiplicadas pelos ganhos de canal desses usuários em relação ao usuário atual ($G_{m',u}^k$). Importante destacar que $I_{ICI}^{k,m,u,s}$ gerada por usuários diferentes do usuário u são oriundas de BS diferentes de m .

A capacidade do canal para um usuário, seja ele com maior ou menor ganho de canal dentro de um cluster, é determinada pela Equação 5 [17]:

$$R_k = B_s \log_2(1 + \gamma_{total}) \quad (5)$$

Onde R_k representa a taxa de transmissão de dados para o CD, refletindo a taxa máxima de transmissão de dados sob as condições atuais do canal, B_s é a largura de banda do subcanal, indicando a quantidade de espectro disponível

para a transmissão, e a γ_{total} é a relação sinal-ruído mais interferência calculada nas Equações 1 e 2 e que incorpora a qualidade do sinal diante do ruído e da interferência no canal.

B. Modelo Proposto

A Figura 2 apresenta o fluxograma do PSO adotado neste trabalho, o qual será descrito na subseção C. Abaixo, são explicadas as duas variáveis que a função objetivo busca otimizar, são elas EE e Probabilidade de bloqueio (OP).

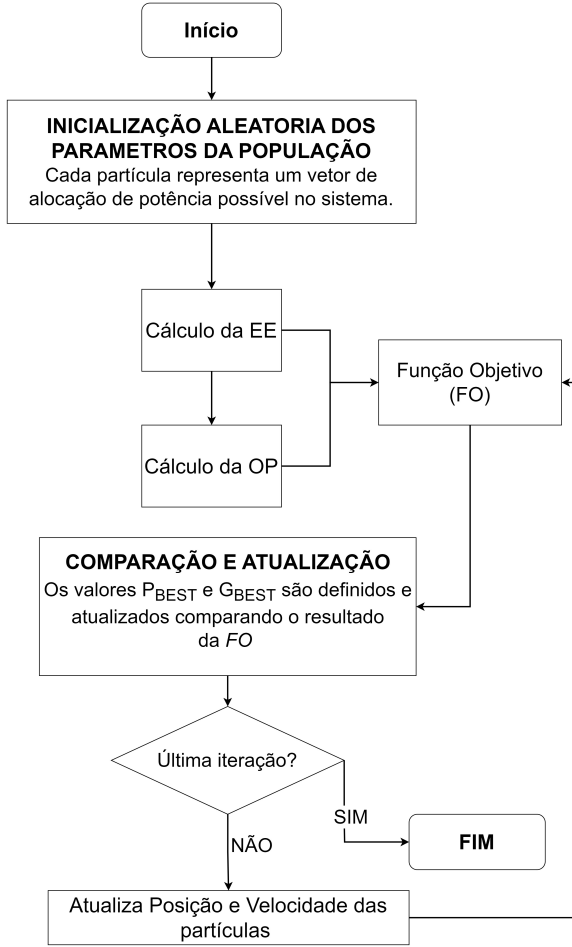


Fig. 2. Fluxograma geral do PSO

1) *Eficiência Energética Total do Sistema*: A EE é uma medida da quantidade de dados transmitidos por unidade de energia consumida. Portanto, a EE total do sistema é calculada considerando a taxa de dados transmitida e a potência consumida.

A EE total do sistema é dada pela equação:

$$EE = \frac{\sum_{k=1}^K R_k}{P_c + \sum_{k=1}^K p_k + p_{k,c}} \quad (6)$$

sendo que R_k , é a taxa de transmissão do usuário k , p_k é a potência atribuída à parte privada da mensagem do usuário k ,

$p_{k,c}$ é a potência atribuída à parte comum da mensagem do usuário k e P_c é o consumo fixo de energia para operar a BS.

2) *Probabilidade de Bloqueio*: A OP quantifica a proporção de CDs que não atingem a taxa de dados mínima θ_{min} . É considerado bloqueado se sua taxa de dados $R_{m,u,s}^k$ for menor que θ_{min} . A OP total é calculada pela fórmula:

$$OP_{total} = \frac{\sum_{u=1}^U \rho_u}{U} \quad (7)$$

onde ρ_u é 1 se o CD u está bloqueado e 0 caso contrário, e U é o número total de CDs.

3) *Problema de Otimização*: Com base nas equações anteriores apresentadas, é possível formular matematicamente o problema de maximização de EE total e mitigação da OP em uma rede RSMA 6G como uma otimização com dois objetivos:

$$\max_{p_{m,s,u}^k} EE \quad (8)$$

$$\min_{p_{m,s,u}^k} OP \quad (9)$$

Sujeito a :

$$\sum_{u=1}^U \sum_{s=1}^S p_{m,s,u}^k \leq p_k^{max}, \forall k, m \quad (10)$$

$$p_{m,u,s}^k > 0, \forall k, m, u, s \quad (11)$$

Sendo que (10) representa a restrição de potência total da BS e (11) indica que a potência atribuída aos usuários deve ser positiva.

C. Etapas de Otimização

O PSO é um algoritmo de otimização inspirado nos comportamentos colaborativos observados em grupos de animais, como bandos de pássaros e cardumes de peixes [6]. Este método é valorizado por sua simplicidade de implementação e menor necessidade de parâmetros de ajuste, quando comparado a outros algoritmos bioinspirados. O PSO utiliza operações aritméticas básicas para atualizar as posições das partículas, o que facilita a sua aplicação. Além disso, o algoritmo é conhecido por sua rápida convergência, resultado da cooperação entre partículas, o que permite uma exploração eficiente do espaço de busca.

O algoritmo inicia com uma população de soluções candidatas em posições aleatórias, cada partícula representando uma possível alocação de potência para os usuários da rede. A Função Objetivo (FO) do PSO considera o cálculo da EE e da OP. O objetivo é maximizar a EE enquanto minimiza a OP, considerando a alocação de potência e as taxas de transmissão entre os dispositivos. Esta função é então avaliada para cada partícula na população, utilizando a abordagem da soma ponderada, um método clássico que converte problemas multiobjetivos em um problema mono-objetivo ao atribuir um peso a cada FO normalizada e somando-os [7]. A Equação 12 representa a FO implementada, sendo que os pesos foram definidos para um cenário em que a OP tem maior prioridade do que a EE. As normalizações das variáveis $NormEE$ e

NormOP são necessárias para garantir que elas tenham o mesmo intervalo de valores, permitindo uma combinação na FO, como apresentado na equação 12

$$FO = 0.2 \times \text{NormEE} + 0.8 \times \text{NormOP} \quad (12)$$

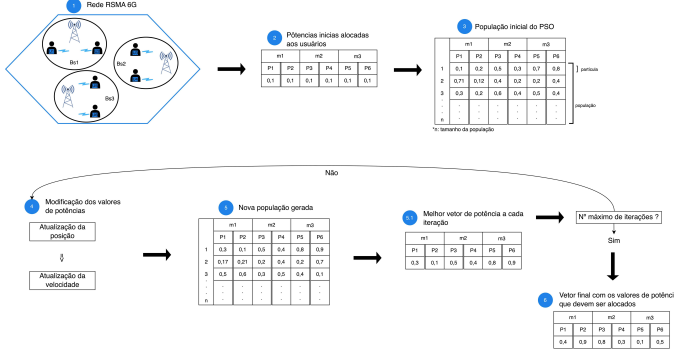


Fig. 3. Etapas de otimização do PSO utilizado.

A Figura 3 exemplifica o processo da otimização da alocação de potência dos usuários utilizando o PSO aplicado ao modelo de rede proposto. A etapa 1 ilustra a rede RSMA 6G que está sendo analisada com a disposição das BSs e dos usuários. Na etapa 2, são alocadas potências iniciais para os usuários conectados, nota-se que estes valores são iguais para todos os usuários.

Em seguida, na etapa 3 inicia-se a execução do PSO com a formação aleatória de uma população inicial. Cada partícula da população representa uma solução candidata contendo um vetor com as potências que devem ser alocadas para cada usuário. Cada CD recebe apenas uma SC. A população possui tamanho igual a n , sendo n o número máximo de partículas.

Na etapa 4 ocorrem os processos do PSO chamados de atualização da posição e velocidade das partículas, que consiste em alterar os valores de potência para gerar novas partículas. Com isso, é gerada uma nova população, como ilustrado na etapa 5. O PSO possui um número máximo de iterações sendo que em cada iteração uma nova população é gerada.

A cada iteração, é realizada uma avaliação para identificar a partícula que possui o vetor de potências resultando na maior EE e na menor OP, sendo esta considerada a melhor da iteração corrente (etapa 5.1). Para isso, o PSO realiza a comparação da iteração atual com os valores obtidos das demais iterações, P_{best} (melhor solução individual) e G_{best} (melhor solução global) encontrados até o momento. Posteriormente, ao considerar todas as iterações, a partícula que apresentou o vetor de potências que resultou na maior FO global é selecionada como a solução final. Esses valores de potência são então alocados aos respectivos usuários (etapa 6).

O PSO possui parâmetros que configuram a forma como a busca pelas soluções será realizada. Estes parâmetros do PSO estão dispostos na Tabela I

TABLE I
PARÂMETROS DE SIMULAÇÃO DO PSO.

População (n)	30
Coefficiente de inércia (w)	1
Constante de aceleração cognitiva (c1)	1.4962 [16]
Constante de aceleração social (c2)	1.4962 [16]
Coefficiente de amortecimento da inércia (wdamp)	0.99
Iterações	70

IV. RESULTADOS

Foi avaliada a eficiência de alocação de potência de transmissão do PSO em comparação com o ERPA (Equal RB Power Allocation) [11], método que aloca potências iguais para os CDs. Foram implementados cenários com diferentes quantidades de CDs dispostos aleatoriamente em um área de 1 km². As BSs, do tipo HPN e LPN, foram posicionadas de acordo com a Figura 4.

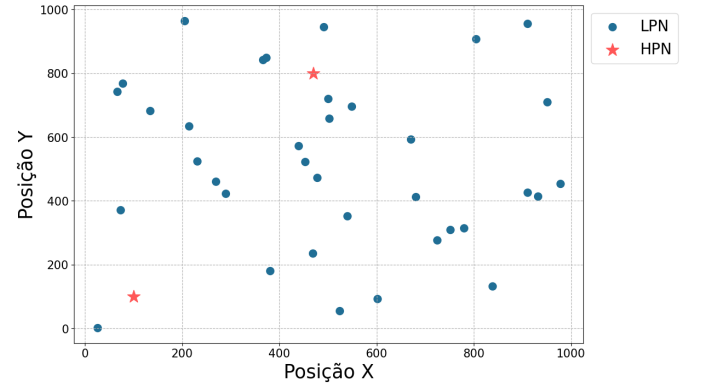


Fig. 4. Disposição das LPN e HPN.

Em relação à configuração da rede, o modelo de perda de caminho, que determina a atenuação do sinal à medida que se propaga pelo ambiente, foi considerado como $34.53 + 35\log_{10}(d)$ (d é a distância entre o CD e a BS a qual ele está conectado) [12]. A potência do ruído é $\sigma^2 = -20\text{dBm}$ [7]. A tabela II apresenta os demais parâmetros do cenário.

TABLE II
PARÂMETROS DO CENÁRIO

Parâmetros	Valor
Largura de banda do sistema (W) [9]	3 GHz
Número de HPNs	2
Número de LPNs	16
Potência de transmissão total LPN/HPN (P_{max})	1 W / 10 W
Número de CDs (U)	50-400
Taxa mínima dos CDs (θ_{min})	1.2 Mbps
Consumo de potência de circuito (P_c) [10]	0.001 W

A Figura 5 apresenta uma comparação de desempenho dos modelos PSO e ERPA à medida que o número de CDs aumenta. Especificamente, a Figura 5.A) mostra EE média do PSO, medida em bits por joules, enquanto a Figura 5.B) exibe a porcentagem de OP média. Para 50 CDs o PSO não obteve ganho em relação à EE, ou seja, ambos os modelos obtiveram

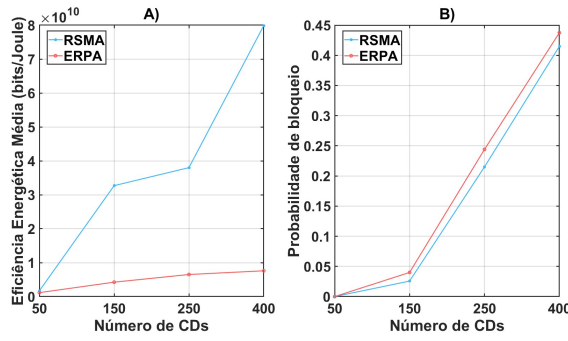


Fig. 5. Avaliação de desempenho dos modelos quando aumenta-se o número de CDs. (A) EE média, (B) OP média.

o mesmo desempenho devido ao número de CDs ser pequena para a simulação realizada.

Entretanto, podemos observar um maior ganho em EE cenário com 400 CDs, onde o modelo obteve um aumento expressivo em relação ao método ERPA. No entanto, é necessário salientar que o ERPA é uma baseline informal onde a potência é alocada igualmente a todos o usuário, sem nenhum processo de otimização. Em geral, o ganho obtido demonstra a superioridade de técnicas de otimização para alocações de potência em relação a uma abordagem tradicional. Isso proporciona uma redução do consumo de energia, sendo crucial para sustentabilidade das redes 6G. Além disso, a EE aprimorada contribui para a redução da pegada de carbono das operadoras de rede, alinhando-se com metas globais de sustentabilidade.

Na Figura 5.B), é observado que para um cenário com 50 CDs, ambos os modelos conseguem satisfazer a taxa mínima exigida pelos CDs. Esse resultado é atribuído à pequena quantidade de CDs em relação à disponibilidade de recursos na rede. Entretanto, ao aumentar o número de CDs para 150, o PSO demonstrou uma redução da OP, aproximadamente 56% menor em comparação ao ERPA. Essa tendência foi mantida para 250 e 400 CDs, onde o PSO resultou em reduções de OP de 13% e 6%, respectivamente, em relação ao ERPA.

A diminuição da OP resulta em menos falhas de conexão, proporcionando uma experiência de usuário mais estável. Ao utilizar o PSO para alocação de recursos, as operadoras podem reduzir custos operacionais com energia e gerenciar mais dispositivos sem comprometer a qualidade do serviço. Os usuários finais também desfrutaram de conexões mais estáveis e melhor qualidade de serviço, especialmente em áreas densamente povoadas.

V. CONCLUSÃO

Este estudo apresentou uma análise da aplicação do algoritmo PSO para a otimização da EE em redes RSMA 6G. Os resultados demonstram que o PSO supera significativamente o método ERPA em termos de EE e OP. Em termos de EE, a otimização proporcionou um aumento expressivo, evidenciando a capacidade do PSO de melhorar a alocação de potência e, consequentemente, o desempenho da rede. Os benefícios de uma maior EE são significativos. Para as

operadoras de rede, a implementação do PSO para alocação de recursos pode resultar em uma redução substancial nos custos operacionais devido ao menor consumo de energia. Além disso, a capacidade de gerenciar um número maior de dispositivos sem comprometer a qualidade do serviço é uma vantagem competitiva importante. Para os usuários finais, uma maior EE se traduz em conexões com menor risco de desconexão e melhor qualidade de serviço.

Em trabalhos futuros, pretende-se explorar um agrupamento de algoritmos de otimização para obter soluções ainda mais aprimoradas e compará-las com outros métodos mais complexos, tendo em vista que o ERPA é uma abordagem trivial de alocação de potência.

REFERENCES

- [1] Gkagkas, G., Vergados, D. J., Michalas, A., ' Dossis, M. (2024). The advantage of the 5G network for enhancing the Internet of Things and the evolution of the 6G network. *Sensors*, 24(8), 2455. [MDPI]
- [2] Dizdar, O., Mao, Y., Han, W., & Clerckx, B. (2020). Rate-Splitting Multiple Access: A New Frontier for the PHY Layer of 6G. In 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), 1-7.
- [3] Clerckx, B., Mao, Y., Schober, R., Jorswieck, E., Love, D. J., Yuan, J., Hanzo, L., Li, G. Y., & Larsson, E. G. (2021). Is NOMA Efficient in Multi-Antenna Networks? A Critical Look at Next Generation Multiple Access Techniques. *arXiv preprint arXiv:2101.04802*.
- [4] Lyu, Xinze, et al. "Rate-Splitting Multiple Access: The First Prototype and Experimental Validation of its Superiority over SDMA and NOMA." *IEEE Transactions on Wireless Communications* (2024).
- [5] Abdel-Razeq, Sharief, et al. "PSO-Based UAV Deployment and Dynamic Power Allocation for UAV-Enabled Uplink NOMA Network." *Wireless Communications and Mobile Computing* 2021.1 (2021): 2722887.
- [6] Wang, D., Tan, D., & Liu, L. (2018). Particle swarm optimization algorithm: an overview. *Soft Computing*, 22(2), 387-408.
- [7] Zhou, G., Mao, Y., & Clerckx, B. (2021). Rate-splitting multiple access for multi-antenna downlink communication systems: Spectral and energy efficiency tradeoff. *IEEE Transactions on Wireless Communications*, 21(7), 4816-4828.
- [8] Torres, F. de O., Santiago Júnior, V. A., da Costa, D. B., Cardoso, D. L., & Oliveira, R. C. L. (2023). Radio resource allocation in a 6G D-OMA network with imperfect SIC: A framework aided by a bi-objective hyper-heuristic. *Engineering Applications of Artificial Intelligence*, 119, 105830.
- [9] Akhtar, M. W., Hassan, S. A., Ghaffar, R., Jung, H., Garg, S., & Hossain, M. S. (2020). The shift to 6G communications: Vision and requirements. *Human-centric Computing and Information Sciences*, 10, 1-27.
- [10] Liu, Z., Hou, G., Yuan, Y., Chan, K. Y., Ma, K., & Guan, X. (2020). Robust resource allocation in two-tier NOMA heterogeneous networks toward 5G. *Computer Networks*, 176, 107299.
- [11] Al-Abbasi, Z. Q., & So, D. K. C. (2017). Resource allocation in non-orthogonal and hybrid multiple access system with proportional rate constraint. *IEEE Transactions on Wireless Communications*, 16(10), 6309-6320.
- [12] Hashemi, R., Beyranvand, H., Mili, M. R., Khalili, A., Tabassum, H., & Ng, D. W. K. (2021). Energy efficiency maximization in the uplink delta-OMA networks. *IEEE Transactions on Vehicular Technology*, 70(9), 9566-9571.
- [13] Sheng, X. (2024). Effective capacity of rate-splitting multiple access with channel estimation errors. *Physical Communication*, 64, 102331.
- [14] Mouni, N. S., Kumar, A., & Upadhyay, P. K. (2021). Adaptive user pairing for NOMA systems with imperfect SIC. *IEEE Wireless Communications Letters*, 10(7), 1547-1551.
- [15] Zhang, Y., Kang, C., Ma, T., Teng, Y., & Guo, D. (2018). Power allocation in multi-cell networks using deep reinforcement learning. In 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), 1-6.
- [16] Shaqarin, T., & Noack, B. R. (2023). A fast-converging particle swarm optimization through targeted, position-mutated, elitism (PSO-TPME). *International Journal of Computational Intelligence Systems*, 16(1), 6.
- [17] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423.

Avaliação de Desempenho de Filtros FIR com a Janela de Blackman-Harris para Redução de Emissões Fora da Banda de Sinais OFDM do Padrão IEEE 802.11a

João Vítor de Souza Lopes
Estudante de Graduação em Engenharia Elétrica
Universidade Federal do Rio Grande do Norte
Natal, Rio Grande do Norte. Brasil
e-mail: joao.vitor.lopes.103@ufrn.edu.br

Luiz Gonzaga de Queiroz Silveira Júnior
Departamento de Engenharia de Comunicações
Universidade Federal do Rio Grande do Norte
Natal, Rio Grande do Norte. Brasil
e-mail: luiz.silveira@ufrn.br

Resumo—Neste trabalho, é proposto o uso de um estágio adicional de filtragem para confinar o espectro de sinais do padrão IEEE 802.11a, reduzindo-se as emissões fora da banda, típicas do OFDM. Para se atingir esse objetivo, usa-se a janela de Blackman-Harris no projeto dos filtros FIR avaliados. A partir dos resultados obtidos, constatou-se um elevado nível de supressão, com atenuações maiores do que 100 dB, o que proporciona um desempenho satisfatório desses filtros na mitigação das emissões indesejadas. Portanto, a adição de um filtro FIR com janela de Blackman-Harris permite aumentar a eficiência no uso do espectro alocado para as redes Wi-Fi, ao mesmo tempo que provoca um aumento da PAPR bastante limitado.

Index Terms—IEEE 802.11a, OFDM, OOB, FIR Filters, Blackman-Harris Window.

I. INTRODUÇÃO

As redes Wi-Fi estão presentes no cotidiano das empresas e nos lares, tendo sido fundamentais para a popularização da internet, com bilhões de conexões sem fio ativas sendo responsáveis pelo escoamento de vários exabytes do tráfego de pacotes, gerado por um grande número de serviços.

Esse volume colossal de dados trafega por um canal de comunicações que é variante no tempo e tipicamente apresenta como característica o desvanecimento, isto é, variações lentas e rápidas na potência recebida pelo receptor, com redução da relação sinal-ruído, SNR (SNR, sigla inglesa para *Signal-to-Noise Ratio*) e aumento da interferência inter-simbólica, ISI (ISI, sigla inglesa para *Intersymbol Interference*) [1].

Para enfrentar os desafios impostos pela mobilidade dos terminais, com a presença de espalhadores e obstruções, ao mesmo tempo em que se provê baixa latência e altas taxas de transmissão, o padrão IEEE 802.11a empregou a transmissão multiportadora, isto é, o OFDM (OFDM, sigla inglesa para *Orthogonal Frequency Division Multiplexing*). Desde o seu advento, o OFDM tem sido empregado em um número crescente de sistemas de comunicações devido às suas características relevantes [1], [2]. A principal vantagem do OFDM explorada por esse padrão é a sua robustez diante da seletividade em frequência do ambiente de transmissão [1].

Apesar de suas vantagens, o OFDM apresenta características indesejáveis como, por exemplo, as emissões fora da banda, OOB (OOB, sigla inglesa para *Out-of-Band Emission*) e a razão entre a potência de pico e a potência média, PAPR (PAPR, sigla inglesa para *Peak-to-Average Power Ratio*). As emissões fora da banda são geradas pela sobreposição de um grande número de lóbulos laterais das subportadoras OFDM e podem causar interferência nas faixas de frequência licenciadas de canais adjacentes. Além disso, o sinal OFDM pode levar os amplificadores de potência a operarem fora do ponto de operação, causando distorções devido às não-linearidades dos dispositivos eletrônicos empregados [1], [3]–[5].

Uma solução para mitigar as emissões fora da banda é a adição de um intervalo de guarda, o que causa um desperdício de espectro. No padrão IEEE 802.11a, por exemplo, 83% da largura de banda de 20 MHz é utilizada efetivamente para a transmissão [1].

Esse trabalho propõe o uso de filtros digitais do tipo FIR (FIR, sigla inglesa para *Finite Impulse Response*) com janela de Blackman-Harris para a mitigação das emissões fora da banda, permitindo-se obter um aproveitamento da banda de frequência destinada ao Wi-Fi sem desperdício de espectro. Para isso, foi considerado como requisito de projeto uma atenuação na faixa de rejeição de pelo menos 100 dB, valor típico exigido para outras tecnologias também baseadas no OFDM. Essa atenuação garante uma supressão eficaz das emissões fora da banda [6].

O desenvolvimento desse trabalho foi realizado no ambiente *Simulink* do *software* Matlab®, licenciado e executando na nuvem, utilizando-se, para isso, de blocos das *toolboxes* “*Communications*” e “*Digital Signal Processing*”.

Este texto está organizado como segue: A Seção II identifica o problema das emissões fora da banda do sinal OFDM no padrão IEEE 802.11a e a Seção III discute a solução proposta neste artigo. Os resultados obtidos a partir das avaliações de desempenho realizadas são mostrados na Seção IV. Finalmente, a Seção V é dedicada a tecer as conclusões obtidas

com a realização desse estudo.

II. IDENTIFICAÇÃO DO PROBLEMA

As redes locais sem fio baseadas no padrão IEEE 802.11a apresentam diversas características, das quais citam-se o raio de cobertura em torno de 50 m, a disponibilidade de 8 canais de largura de banda de 20 MHz e a capacidade de transmissão de dados com taxas variáveis, de 6 – 54 Mbits/s [7], [8].

A principal característica do OFDM, explorada pelo padrão IEEE 802.11a, é o uso de portadoras ortogonais, que dividem a banda de transmissão em sub bandas de frequência, transmitindo os dados através de subportadoras, que estão espaçadas de 312,5 kHz. A Tabela I apresenta mais parâmetros do OFDM empregado no padrão IEEE 802.11a [8].

Tabela I: Parâmetros do sinal OFDM no padrão IEEE 802.11a

Parâmetro	Valor
Largura de Banda	20 MHz
Tamanho da FFT	64
Total de Subportadoras (N_{ST})	53
Número de Subportadoras para Dados (N_{SD})	48
Número de Subportadoras para Piloto (N_{SP})	4
Intervalo entre Subportadoras (F_{SP})	312,5 kHz (20MHz/64)
Período da IFFT/FFT (T_{FFT})	3,2 μ s ($1/F_{SP}$)
Intervalo de Guarda (T_{GI})	0,8 μ s ($T_{FFT}/4$)
Intervalo de símbolo (T_{SYM})	4 μ s ($T_{GI} + T_{FFT}$)
Número de Subportadoras na Banda de Guarda	11 (FFT Length - N_{ST})
Comprimento do Prefixo Cíclico	16

Como um símbolo nesse padrão é o resultado de 53 subportadoras (53 = 4 piloto + 48 para dados + 1 nula), tem-se que a ocupação efetiva é de 16.562,5 KHz (53 $\times$ 312,5 KHz). Dos 20 MHz disponíveis, cerca de 3.437,5 KHz (11 subportadoras para banda de guarda) não é efetivamente usado.

Portanto, a ocupação líquida do espectro não é eficiente, pois serve para “acomodar” (em uma faixa de frequência bastante desejada por usuários não licenciados) as emissões fora da banda geradas pelo OFDM. Essa desvantagem será explicada com mais detalhes na sequência.

A. Emissões fora da banda

Na transmissão serial de única portadora, toda a banda disponível, W , é ocupada pela mesma e única portadora. Contudo, os sistemas multiportadoras, como, por exemplo, OFDM, utilizam várias sub portadoras. Nesse caso, a banda total, W , poderia ser dividida em N sub bandas, cada uma das quais apresentando uma largura de banda aproximadamente igual a W/N . Quando comparado com o FDM tradicional, o OFDM permite uma maior aproveitamento do espectro, de cerca de 50 % [1]. A Figura 1 ilustra a ocupação no espectro com uma portadora única em contraste com subportadoras OFDM [3].

No OFDM, a sequência de bits a ser transmitida é dizimada, de modo a gerar N subsequências de bits, que serão mapeados por símbolos complexos de uma constelação de sinais. A Figura 2 ilustra esse procedimento no caso de uma modulação BPSK.



Figura 1: Banda ocupada com portadora única *versus* ocupada por subportadoras. Adaptado de [3].

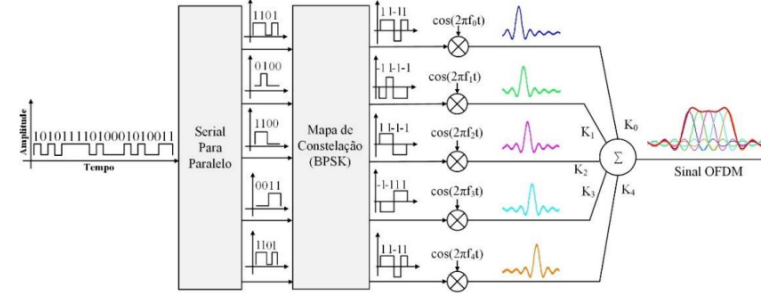


Figura 2: Exemplo da geração de um sinal OFDM com modulação BPSK. Adaptado de [4].

Posteriormente, cada símbolo de duração T é modulado por subportadoras $\cos(2\pi f_k t)$. Desde que o espaçamento entre subportadoras seja $F_{SP} = \frac{1}{T}$, a amplitude máxima de uma subportadora ocorrerá no instante em que as demais subportadoras cruzarão o zero. Dessa forma, a ortogonalidade entre subportadoras é garantida, mesmo que haja sobreposição dos lóbulos laterais. Na sequência, essas subportadoras são somadas de forma a se obter a transmissão multiportadora, isto é, o sinal OFDM. A Figura 3 ilustra o espectro típico de um sinal OFDM.

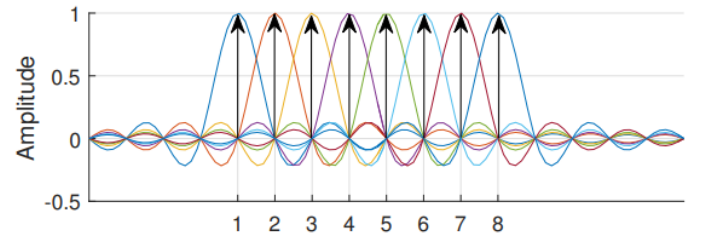


Figura 3: Espectro de um sinal OFDM. Adaptado de [5].

A soma dos lóbulos laterais não pode ser desprezada e, de fato, será responsável por gerar as emissões fora da banda, OOBs, típicas do OFDM. Assim, de forma a mitigar possíveis interferências nas faixas de frequência licenciadas de canais adjacentes, o padrão IEEE 802.11a utiliza cerca de 83% dos 20 MHz de cada canal e destina o restante da banda para a acomodação destas emissões indesejadas.

B. Vantagens do OFDM para o padrão IEEE 802.11a

Atualmente, o OFDM está presente em diversos sistemas de comunicações sem fio. Duas características do OFDM são destacadas para se justificar o emprego dessa transmissão multiportadora no padrão IEEE 802.11a.

Inicialmente, destaca-se a robustez dessa transmissão sobre um canal sujeito ao desvanecimento seletivo em frequência.

Com a divisão da sequência de bits em múltiplas subseqüências, o número de subseqüências precisa ser definido de maneira que o tempo de cada sub símbolo seja maior do que o tempo de coerência do subcanal por onde haverá a transmissão de cada subseqüência. Dessa forma, mesmo que o canal de comunicações de largura de banda integral, W , esteja sujeito ao desvanecimento seletivo em frequência, o desvanecimento será plano em cada subcanal. Isso acarreta uma menor probabilidade para a ocorrência de ISI.

Ainda, adiciona-se um prefixo cíclico após a formação do símbolo OFDM. Esta adição oferece como vantagem a possibilidade de uso de equalizadores mais simples do lado do receptor, o que elimina a ISI e exclui o possível efeito de interferência entre subportadoras, ICI (ICI, sigla inglesa para *Inter Carrier Interference*).

Finalmente, destaca-se outra vantagem do OFDM, que trata-se da forma como essa técnica é implementada de forma totalmente digital, isto é, mediante a implementação de um algoritmo rápido de cálculo da Transformada Discreta Inversa de Fourier, o que favorece sua implementação nos processadores de sinais digitais, DSP (sigla inglesa para *Digital Signal Processor*) comerciais, evitando-se, assim, a complexidade de *hardware* dedicado com o emprego de vários osciladores analógicos, necessários para gerar as subportadoras. De fato, uma implementação analógica do OFDM é impraticável.

III. SOLUÇÃO PROPOSTA

O objetivo desse trabalho é propor e avaliar a implementação de filtros digitais do tipo FIR para a mitigação das emissões fora da banda de sinais OFDM do padrão IEEE 802.11a. Assim, este trabalho propõe uma solução ao problema da emissão fora de banda, que é um dos principais efeitos indesejados do OFDM, que surge com a soma dos lóbulos laterais das funções *sinc*(ω) empregadas.

A adição de um estágio de filtragem digital a ser utilizado no sinal OFDM resulta no sinal OFDM filtrado, f-OFDM (f-OFDM, sigla inglesa para "*filtered-OFDM*"), e é realizada de maneira a combater o desperdício de espectro necessário para acomodar as emissões fora da banda, OOB. Com a utilização de filtros, há o confinamento do sinal na banda alocada, reduzindo-se, consideravelmente, o desperdício de espectro. Dessa forma, tal estágio é quase que indispensável em um sistema de transmissão baseado em OFDM que promova um melhor aproveitamento do espectro [2], [6].

Para isso, o projeto do filtro FIR será realizado com a técnica de janelas. Devido às suas características, esse trabalho propõe o uso da janela de Blackman-Harris. No projeto desses filtros, serão avaliadas as atenuações na faixa de rejeição para diferentes ordens de filtros. Aqui, o requisito de projeto é alcançar uma atenuação na faixa de rejeição de pelo menos 100 dB de forma a suprimir eficazmente as emissões fora da banda [6].

A. Projeto de filtros FIR com a Janela de Blackman-Harris

Os filtros digitais tipo FIR são sempre estáveis e de fácil implementação computacional, em contraste aos filtros IIR

(IIR, sigla inglesa para *Infinite Impulse Response*), que exigem um protótipo em tempo contínuo e nem sempre permanecem estáveis após a quantização dos coeficientes do filtro.

A resposta ao impulso de um filtro FIR pode ser matematicamente descrita pela equação (1) [9]

$$y[n] = c_1x[n] + c_2x[n-1] + c_3x[n-2] + \dots + c_kx[n-m], \quad (1)$$

sendo m a ordem do filtro, $x[n]$ o sinal de entrada, $y[n]$ o sinal de saída e c_k os k coeficientes do filtro. Logo, um filtro FIR não apresenta recursividade e não possui polos.

Um dos métodos clássicos para o projeto de filtros FIR é o método da função janela. Neste método, a resposta de um filtro ideal deslocado, $h_d[n]$, que é sempre não causal e apresenta uma duração infinita no tempo discreto, é truncada por uma função janela, $w[n]$, obtendo-se a resposta ao filtro real, $h[n]$, que é causal e de duração finita [9]. Matematicamente, tem-se

$$h[n] = h_d[n] \cdot w[n]. \quad (2)$$

Na literatura existe uma variedade de funções janela clássicas como, por exemplo, Kaiser, Blackman, Hanning, Hamming, Retangular, Nuttall, dentre várias outras, que são usadas para propósitos variados.

Devido às suas características de atenuação na banda de rejeição e por ser uma janela ainda pouco explorada, a janela de Blackman-Harris foi escolhida neste estudo. Ela é descrita matematicamente pelas Equações (3) e (4) [9]:

$$w(n) = 0,35875 - 0,48829 \cos\left(\frac{2\pi n}{N-1}\right) + 0,14128 \cos\left(\frac{4\pi n}{N-1}\right) - 0,01168 \cos\left(\frac{6\pi n}{N-1}\right), \quad (3)$$

para $0 < |n| < N-1$.

No domínio da frequência, têm-se

$$W(\omega) = 0,7175 \frac{\text{sen}[\omega(N-1)]}{\omega} - 0,48829 \left[\frac{\text{sen}\left[\left(\omega + \frac{2\pi}{(N-1)}\right)(N-1)\right]}{\omega + \frac{2\pi}{(N-1)}} + \frac{\text{sen}\left[\left(\omega - \frac{2\pi}{(N-1)}\right)(N-1)\right]}{\omega - \frac{2\pi}{(N-1)}} \right] + 0,14128 \left[\frac{\text{sen}\left[\left(\omega + \frac{4\pi}{(N-1)}\right)(N-1)\right]}{\omega + \frac{4\pi}{(N-1)}} + \frac{\text{sen}\left[\left(\omega - \frac{4\pi}{(N-1)}\right)(N-1)\right]}{\omega - \frac{4\pi}{(N-1)}} \right] - 0,01168 \left[\frac{\text{sen}\left[\left(\omega + \frac{6\pi}{(N-1)}\right)(N-1)\right]}{\omega + \frac{6\pi}{(N-1)}} + \frac{\text{sen}\left[\left(\omega - \frac{6\pi}{(N-1)}\right)(N-1)\right]}{\omega - \frac{6\pi}{(N-1)}} \right]. \quad (4)$$

As Figuras 4 e 5 apresentam os comportamentos da janela de Blackman-Harris para diferentes ordens, nos domínios do tempo e da frequência, respectivamente.

Observa-se que o aumento da ordem do filtro promove a redução da largura do lóbulo principal, assim como um

aumento na amplitude do lóbulo lateral de pico. Desse modo, a faixa de transição torna-se menor e a atenuação na faixa de rejeição eleva-se [10], [11]. As Figuras 6 e 7 traçam um paralelo entre a janela de Blackman-Harris e as janelas tradicionais.

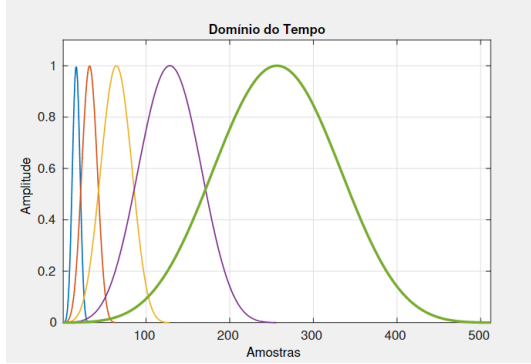


Figura 4: Janela de Blackman-Harris, para diferentes ordens, no domínio do tempo. Fonte: Autoria Própria.

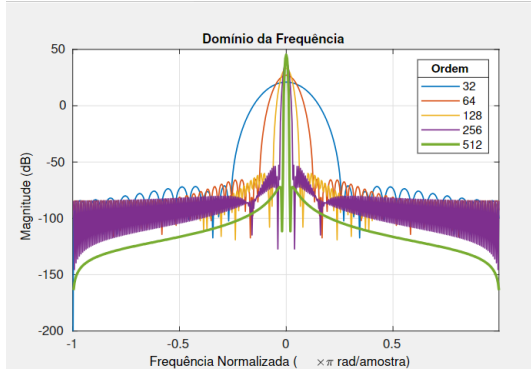


Figura 5: Janela de Blackman-Harris, para diferentes ordens, no domínio da frequência. Fonte: Autoria Própria.

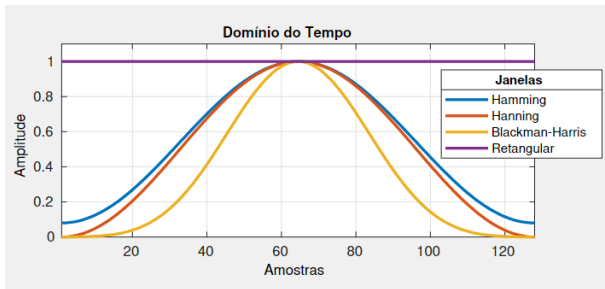


Figura 6: Comparação entre janelas de Blackman-Harris em contraste com outras janelas, com ordem igual a 128, no domínio do tempo. Fonte: Autoria Própria.

Através dos resultados apresentados nas Figuras 6 e 7 é possível constatar que a janela de Blackman-Harris possui uma largura de lóbulo principal maior, com uma amplitude de lóbulo secundário reduzida, em comparação com as demais janelas avaliadas nesse trabalho. Esta redução na amplitude dos lóbulos secundários proporciona uma maior atenuação na

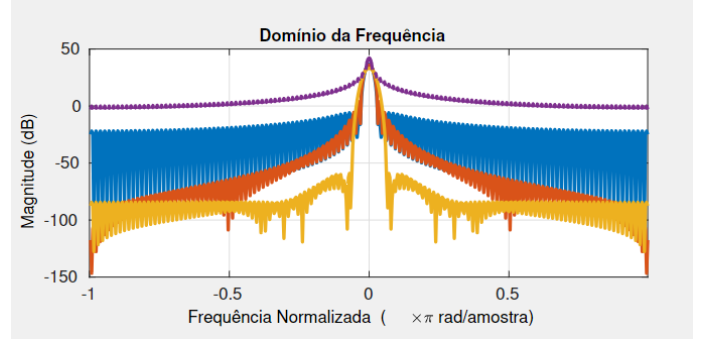


Figura 7: Comparação entre janelas de Blackman-Harris em contraste com outras janelas, com ordem igual a 128, no domínio da frequência. Fonte: Autoria Própria.

banda de rejeição, característica desejada para um filtro que mitigue os vazamentos espectrais do OFDM.

O desempenho do filtro projetado será avaliado também com relação ao aumento do parâmetro PAPR, pois embora o uso de filtros FIR mitiguem o problema do OOB, seu emprego pode levar ao aumento da PAPR.

Matematicamente, define-se a PAPR [12]

$$PAPR = \frac{|P_{pico}|^2}{|P_{rms}|^2}, \quad (5)$$

sendo P_{pico} a potência de pico do sinal e P_{rms} a potência eficaz. Assim, percebe-se que a diminuição da emissão fora de banda por um filtro digital reduz sua potência eficaz, o que eleva a PAPR.

É desejado que um sistema de telecomunicações apresente baixa PAPR, uma vez que um elevado valor poderia causar distorções não lineares, dentre outros problemas. Portanto, o projeto do filtro digital é crítico no contexto do padrão IEEE 802.11a.

IV. RESULTADOS

Os filtros FIR propostos nesse artigo foram implementados no ambiente *Simulink*, na nuvem do Matlab®, versão R2023, licenciada. Foram projetados seis filtros FIR passa-baixas de diferentes ordens, pelo método da janela, todos com a janela de Blackman-Harris, implementados através do bloco “*Digital Filter Design*” (*fdatool*) [13]–[15].

Também nas simulações desenvolvidas, utilizamos os parâmetros da Tabela I, que definem as características do OFDM utilizado no padrão IEEE 802.11a. Para se gerar a forma de onda, foi usado o aplicativo “*Wireless Waveform Generator*” (*wirelessWaveformGenerator*), o qual permite gerar as formas de onda OFDM, com modulação 64-QAM, que foi a mesma adotada para todos os filtros avaliados [14] e que é a de maior ordem suportada pelo padrão IEEE 802.11a. Os resultados obtidos estão mostrados nas Figuras 8 e 9.

Na Figura 8, é possível observar 6 curvas, correspondentes aos espectros dos sinais f-OFDM obtidos a partir do emprego de 5 diferentes filtros, de ordens distintas: 256, 512, 1024, 2048 e 4096, em contraste com o sinal OFDM.

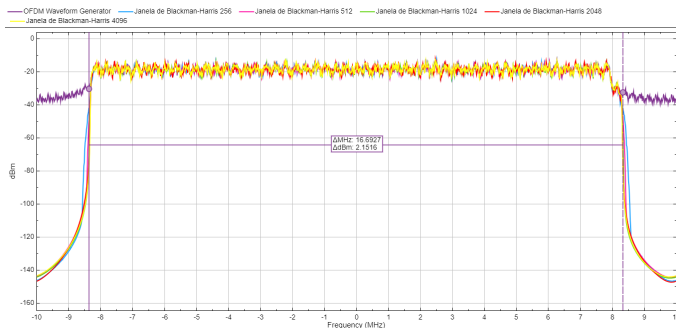


Figura 8: Espectro dos sinais OFDM e f-OFDM.

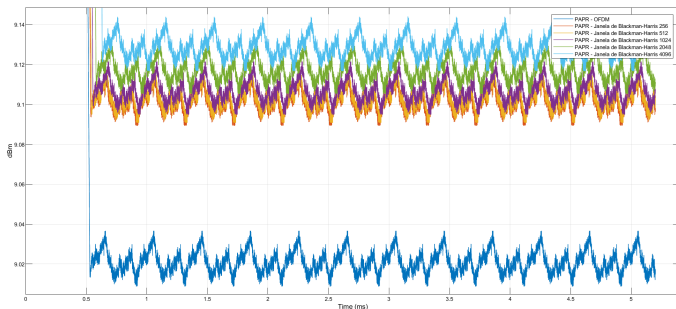


Figura 9: PAPR dos sinais OFDM e f-OFDM.

Mesmo os filtros de ordens mais baixas, de 256 e 512, conseguem prover uma redução eficaz das emissões fora de banda de um sinal OFDM. Entretanto, à medida que a ordem do filtro aumenta, há uma redução mais acentuada da largura de banda da faixa de transição. Assim, para uma frequência de corte de 8,3 MHz, que é a máxima frequência ocupada pelas 53 subportadoras, é possível constatar uma atenuação bastante elevada na faixa de rejeição do sinal OFDM, que chega a atingir valores próximos a 140 dBm, portanto, maior do que 100 dBm, o que é desejado para suprimir as emissões fora da banda nas aplicações com o OFDM presente no padrão IEEE 802.11a. Dessa forma, um estágio adicional de filtragem digital com filtro FIR com janela de Blackman-Harris permite reduzir as emissões fora da banda de forma bastante eficaz, eliminando a necessidade de uma banda de guarda de cerca de 3,5 MHz.

A Figura 9 apresenta os valores da PAPR para os sinais OFDM e f-OFDM. Fica evidente que o uso de filtros digitais provoca aumento da PAPR nos sistemas f-OFDM. Isto se deve ao fato de que o processo de filtragem reduz a potência média do sinal, tornando menor o denominador da Equação (5).

Contudo, é perceptível que o aumento da ordem do filtro produz um aumento bastante limitado da PAPR, considerando-se o emprego de filtros FIR com janela de Blackman-Harris.

V. CONCLUSÃO

Este trabalho propôs a redução das emissões fora da banda de sinais OFDM típicos do padrão IEEE 802.11a através do uso de filtros FIR com a janela de Blackman-Harris. Por meio

dos resultados obtidos, observou-se uma redução satisfatória das emissões fora de banda mesmo para ordens menores, como 256 ou 512. Além disso, a PAPR aumentou de forma limitada com o aumento da ordem, não sendo um critério decisivo para a escolha do filtro. Desse modo, ficou evidente que o filtro de 256 coeficientes é adequado para atenuar o efeito da OOB, provendo a menor complexidade de projeto, a menor PAPR dentre todos os filtros simulados e atenuação na banda de rejeição satisfatória. O estágio adicional de filtragem tornou desnecessária a banda de guarda de cerca de 3,5 MHz, aumentando a eficiência no uso do espectro para essa aplicação.

REFERÊNCIAS

- [1] PINTO, E. L. e ALBUQUERQUE, C. P., "A Técnica de Transmissão OFDM", *Revista Científica Periódica - Telecomunicações - Volume 05 - Número 01*, ISSN 1516-2338., 2002.
- [2] H. F. Arrano e C. A. Azurdia-Meza., "OFDM: today and in the future of next generation wireless communications," *2016 IEEE Central America and Panama Student Conference (CONESCAPAN)*, Guatemala City, Guatemala, pp. 1-6, doi: 10.1109/CONESCAPAN.2016.8075209., 2016.
- [3] XIE, J., "A Constant Envelope Variation of OFDM Waveform", *10.13140/RG.2.2.26686.79688*, 2017.
- [4] SILVA, L.C.A.M., "Estudo da utilização de OFDM para mitigar o efeito de mistura de quatro ondas em sistemas DWDM", *Trabalho de Conclusão de Curso (TCC em Engenharia Eletrônica e Telecomunicações)* - UFU, 2022.
- [5] DOGAN, S. and TUSHA, A. and ARSLAN, H., "OFDM with Index Modulation for Asynchronous mMTC Networks", *Sensors* 2018, 18, 1280; doi:10.3390/s18041280, 2018.
- [6] P. Martin-Martin, R. Bregovic, A. Martin-Marcos, F. Cruz-Roldan and T. Saramaki., "A Generalized Window Approach for Designing Transmultiplexers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 9, pp. 2696-2706, Oct. 2008, doi: 10.1109/TCSI.2008.920095., 2008.
- [7] Pathak, P., "A Comparison Between WLAN (IEEE 802.11a, B, G, N and Ac) Standards", *GRIN Verlag*, 2017.
- [8] Gast, M., "802.11 Wireless Networks. The Definitive Guide", *O'Reilly Media*, 2005.
- [9] PRABHU, K.M.M., "Window Functions and Their Applications in Signal Processing", *CRC Press*, 2014.
- [10] LATHI, B.P., "Sinais e Sistemas Lineares", 2.ed, Bookman, 2007.
- [11] LATHI, B.P., "Sistemas de Comunicações Analógicos e Digitais Modernos", 4.ed, LTC, 2012.
- [12] BIRADAR, R. L. and NARAYANAMMA, G., "PAPR Reduction: Techniques, Analysis and Applications", *Archers & Elevators Publishing House*, 2023.
- [13] MathWorks, *blackmanharris*. Disponível em: <https://www.mathworks.com/help/signal/ref/blackmanharris.html>.
- [14] MathWorks, *Wireless Waveform Generator*. Disponível em: <https://www.mathworks.com/help/comm/ref/wirelesswaveformgenerator-app.html>.
- [15] MathWorks, *Digital Filter Design*. Disponível em: <https://www.mathworks.com/help/dsp/ref/digitalfilterdesign.html>.

Avaliação de Desempenho de Filtros FIR com a Janela de Nuttall para Redução de Emissões Fora da Banda de Sinais OFDM do Padrão IEEE 802.11a

Ayrton Pinheiro de Araújo Silva
Estudante de Graduação em Engenharia Elétrica
Universidade Federal do Rio Grande do Norte
Natal, Rio Grande do Norte. Brasil
e-mail: ayrton.pinheiro.701@ufrn.edu.br

Luiz Gonzaga de Queiroz Silveira Júnior
Departamento de Engenharia de Comunicações.
Universidade Federal do Rio Grande do Norte
Natal, Rio Grande do Norte. Brasil
e-mail: luiz.silveira@ufrn.br

Resumo—Este trabalho propõe o uso de um filtro digital do tipo FIR baseado na Janela de Nuttall para reduzir as emissões fora da banda típicas de transmissões OFDM, sendo a técnica OFDM revisitada nesse trabalho no âmbito do padrão IEEE 802.11a. Nesse contexto, a escolha do filtro FIR é particularmente crítica para sistemas OFDM filtrados devido ao aumento da PAPR decorrente da aplicação de um filtro digital. A partir dos resultados obtidos, constata-se que a filtragem digital constitui uma alternativa promissora para combater o desperdício de espectro de redes wifi, ao mesmo tempo que minimiza o aumento da PAPR.

Index Terms—IEEE 802.11a, OFDM, OOB, FIR Filters, Nuttall Window.

I. INTRODUÇÃO

Desde o seu advento, a técnica de multiplexação por divisão de frequências ortogonais, OFDM (OFDM, sigla inglesa para *Orthogonal Frequency-Division Multiplexing*) foi extensivamente aplicada nas redes de comunicações sem fio (e também com fio), desde que surgiu originalmente no padrão IEEE 802.11a. Dos sistemas sem fio que a utilizam, destacam-se os padrões amplamente usados nas redes wifi, IEEE 802.11a e 802.11g, os padrões de redes celulares, 3G, 4G e 5G NR da 3GPP, o rádio digital, DAB (DAB, sigla inglesa para *Digital Audio Broadcasting*) e a tv digital, DVB-T (DVB-T, sigla inglesa para *Digital Video Broadcasting – Terrestrial*). Já para sistemas com fio, o OFDM foi inicialmente empregado em ADSL (ADSL, sigla inglesa para *Asymmetric Digital Subscriber Line*), que é um padrão que utiliza a linha telefônica comum como meio de transmissão [1].

No que diz respeito as vantagens, dentre várias de suas características, três resumem os motivos principais pelos quais a técnica OFDM é empregada massivamente: (i) a facilidade de implementação, (ii) possibilidade do uso de equalizadores mais simples no receptor e, (iii) maior robustez em canais seletivos em frequência. Contudo, algumas das suas desvantagens são bem conhecidas: (i) emissão fora da banda, OOB (OOB, sigla inglesa para *Out-of-Band Emission*), (ii) o problema da PAPR (PAPR, sigla inglesa para *Peak-to-Average Power Ratio*) e (iii) *Frequency offset* e ruído de fase.

Devido ao OOB, no padrão IEEE 802.11a, da largura de banda de 20 MHz, apenas 16,6 MHz são destinados para a transmissão efetiva, ou seja, 17% da largura de banda total é desperdiçada para mitigar a emissão fora da banda [2].

Dessa forma, para combater o desperdício de espectro nas redes wifi baseadas nesse padrão, este trabalho propõe o uso de filtros digitais do tipo FIR com a janela de Nuttall para filtrar sinais OFDM. Devido ao aumento da PAPR observado após esse estágio adicional de filtragem, a escolha dos parâmetros desse filtro digital é crítica. Portanto, tal escolha será guiada de maneira a se obter um bom *trade-off* entre a quantidade de coeficientes do filtro e o aumento observado na PAPR.

A partir dos resultados obtidos, constata-se a vantagem do uso do filtro FIR com janela de Nuttall e ordem igual a 256, o qual proporciona um aumento da eficácia do uso do espectro de redes wifi baseadas no padrão IEEE 802.11a, com um aumento limitado da PAPR.

Esse artigo está organizado como segue: Na Seção II, a técnica OFDM será revisitada de maneira a destacar suas vantagens e desvantagens no contexto do padrão IEEE 802.11a. Esta Seção ajuda não só na compreensão da técnica OFDM, mas também favorece o uso desse trabalho como material educacional para a área de processamento digital de sinais para sistemas de telecomunicações, sendo, portanto, uma das contribuições desse trabalho. A Seção III apresenta os detalhes do projeto de filtros FIR com a janela de Nuttall para sinais OFDM do padrão IEEE 802.11a. Na Seção IV, são apresentados e discutidos os resultados obtidos. Finalmente, a Seção V é dedicada a tecer as conclusões obtidas com a realização desse trabalho.

II. A TÉCNICA OFDM

De forma simples, pode-se entender a técnica OFDM como uma técnica de transmissão de dados que combina modulação e multiplexação para enviar dados com uma alta taxa de transmissão, sem que sejam afetados por um canal sujeito ao desvanecimento seletivo em frequência. A ideia básica explora o princípio “dividir para conquistar”, no qual uma portadora de alta taxa de dados é “dividida” em várias subportadoras de

baixa taxa, sendo os dados enviados por muitas subportadoras, através de muitos subcanais. A largura de banda de cada subcanal é aproximadamente igual a divisão da largura de banda total do canal pela quantidade de subportadoras, sendo definida de forma que seja menor do que a largura de banda de coerência do subcanal através do qual a parcela de dados será transmitida. Como resultado, um canal com desvanecimento seletivo em frequência é “convertido” em muitos subcanais com desvanecimento plano, o que favorece o desenvolvimento de esquemas de recepção mais simples [2].

Trata-se, portanto, de um aperfeiçoamento da técnica FDM (FDM, sigla inglesa para *Frequency Division Multiplexing*), que exige uma banda de guarda entre as subportadoras no domínio da frequência para evitar a interferência entre subportadoras de canais adjacentes. Assim, o OFDM, por empregar portadoras ortogonais, é robusto à sobreposição espectral de subportadoras, já que tal interferência mútua entre as subportadoras é desprezível. Isso ocorre devido às sinalizações de banda estreita e ao controle cuidadoso da estrutura de tempo/frequência do sinal, além da sincronização de relógio e da utilização do prefixo cíclico. Essa melhor eficiência espectral é devido à ortogonalidade inerente da técnica OFDM [1].

A. Geração de sinais OFDM via FFT^{-1}

A correta implementação dessa técnica, respeitando o que alguns autores chamam de “*wish list*” (“lista de desejos”), considera um número ideal de subportadoras para determinado canal, com o objetivo de evitar os problemas causados pela seletividade em frequência como, por exemplo, a interferência entre símbolos (ISI) e o patamar de erro de bit irredutível. A *Wish List* é definida pela Equação (1)

$$f_0 > \frac{1}{T_s} > \frac{1}{T_0}, \quad (1)$$

sendo f_0 a largura de banda de coerência, $1/T_s$ a taxa de símbolo e T_0 o tempo de coerência [2].

Com OFDM, desprezando-se a sinalização, a taxa de símbolo é dada pela banda total dividida pela quantidade de subportadoras ($1/T_s = W_{\text{signal}}/N_s$). Assim, após pequenas manipulações, tem-se o intervalo que define o número ideal de subportadoras, N_s , apresentado pela Equação (2) [2]

$$\frac{W_{\text{signal}}}{f_0} < N_s < W_{\text{signal}} \cdot T_0. \quad (2)$$

É bem conhecido que todo esse processo de transformar um canal seletivo em frequência em subcanais com desvanecimento plano implica em grandes vantagens aos sistemas OFDM, como, por exemplo, o uso de equalizadores mais simples do lado do receptor [2].

Além disso, para garantir a ortogonalidade e, consequentemente, o uso de equalizadores lineares, é frequente a utilização do prefixo cíclico, que torna o sistema ainda mais robusto na transmissão de sinais por canais com condições severas de multipercurso. Embora, comumente, já seja utilizado um intervalo de guarda entre símbolos OFDM (para assegurar que não

ocorra a interferência entre símbolos), a utilização do prefixo cíclico evita, também, a interferência entre subportadoras, ICI (ICI, sigla inglesa para *Inter Carrier Interference*) [2].

Esta técnica consiste em adicionar uma cópia de uma certa quantidade de amostras da parte final do símbolo no espaço que estaria vazio do intervalo de guarda. Esse processo acaba fazendo a conversão do que normalmente seria uma convolução linear, entre o sinal transmitido e a resposta ao impulso do canal, em uma convolução circular, garantindo que o sinal visto pelo receptor tenha o mesmo tamanho do sinal transmitido - o que simplifica o processo de equalização e preserva a ortogonalidade. Como consequência, o tempo de símbolo aumenta, passando a ser a soma do tempo útil de símbolo OFDM com o tempo do intervalo de guarda [2].

A geração de um sinal OFDM se inicia com a conversão série-paralelo de símbolos de uma certa quantidade de bits (X_k , para $k = 0, 1, 2, \dots, N_s - 1$) oriundos do estágio de modulação, que pode ser QAM, por exemplo [3]. Em seguida, esse bloco de símbolos em paralelo é modulado por um conjunto de senoides janeladas com frequências múltiplas de $1/T_s$ ($f_k = k/T_s$), como apresentado na Figura 1.

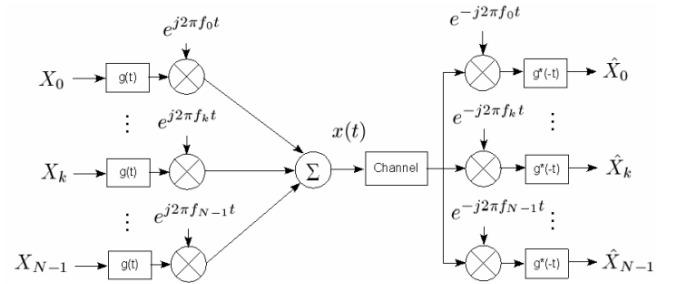


Figura 1. Diagrama de blocos da modulação do sinal OFDM. Adaptado de [3].

De forma analógica, conceitualmente, emprega-se osciladores coerentes: cada símbolo da saída da modulação é multiplicado por uma das senoides geradas pelos osciladores que, por fim, são adicionadas, gerando-se o sinal OFDM em banda base, $x(t)$ [1]. No entanto, para canais que exigem um número elevado de subportadoras (N_s muito alto), essa tarefa se torna inviável, pois o número de osciladores necessários é equivalente à quantidade de subportadoras, tanto na transmissão, quanto na recepção do sinal.

Contudo, verifica-se que ao amostrar no domínio do tempo o sinal OFDM em banda base, $x(t)$, formado nesse processo ilustrado na Figura 1, o resultado se assemelha à definição de uma Transformada Discreta de Fourier Inversa (IDFT), $x[n]$. Isso significa que, ao aplicar-se a IDFT ao conjunto de símbolos complexos da constelação de sinais, X_k , tem-se, praticamente, um sinal OFDM de tempo discreto, $x[n]$.

Como a implementação da geração de sinais OFDM passa a ser realizada utilizando-se de algoritmos de Transformada Inversa Rápida de Fourier, FFT^{-1} (FFT, sigla inglesa para *Fast Fourier Transform*), sobre arquiteturas otimizadas para DSP (DSP, sigla inglesa para *Digital Signal Processor*), é inegável que a solução digital facilita o emprego do OFDM em muitos

sistemas de comunicações digitais, pois evita-se o emprego de osciladores analógicos; o que seria impraticável, dada a necessidade de um grande número de subportadoras. Por sua vez, na recepção, utiliza-se do algoritmo da Transformada Direta, FFT, para se recuperar os símbolos da constelação, X_k . [3].

B. Emissão fora da banda

Quando a transformada de Fourier é aplicada a uma senoide infinita, no domínio da frequência tem-se um impulso centralizado na frequência da senoide. Mas, no caso de uma senoide janelada, a transformada de Fourier gera, no domínio da frequência, uma função $\text{sinc}(\cdot)$ infinita centralizada na frequência da senoide, ou seja, o lóbulo principal fica centralizado na frequência da senoide, mas tem-se infinitos lobos laterais de pequenas amplitudes interferindo em outras frequências. Isso pode ser visto nas Figuras 2a e 2b [4].

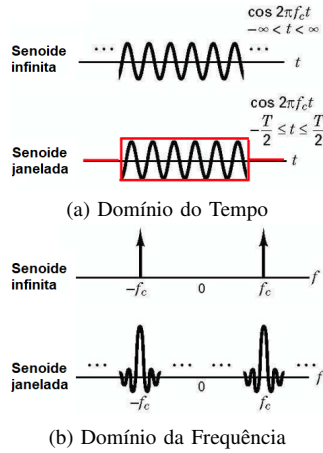


Figura 2. Transformada de Fourier da senoide infinita e da senoide janelada. Adaptado de [4].

Após aplicação da FFT^{-1} , como o sinal OFDM é gerado por uma soma de senoides janeladas, então, no domínio da frequência, as várias funções $\text{sinc}(\cdot)$ espaçadas por $1/T_s$ são somadas, produzindo como resultado uma soma de lobos laterais que resulta numa emissão fora da banda disponível, como é possível de ser observado na Figura 3.

Diante disso, alguns padrões já deixam de utilizar uma porcentagem das partes laterais da sua largura de banda total, a fim de diminuir alguns problemas causados pela emissão fora da banda. Por exemplo, no padrão IEEE 802.11a é disponibilizado uma largura de banda de 20 MHz, dos quais somente 16,6 MHz são destinados a dados úteis, ou seja, 17% da largura de banda total é desperdiçada para mitigar a emissão fora da banda [1], [2].

A Tabela I apresenta alguns dos parâmetros do padrão IEEE 802.11a, que foram considerados no desenvolvimento desse trabalho.

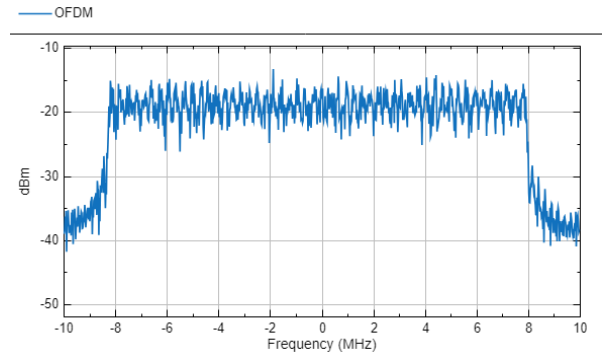


Figura 3. Emissão fora de banda do IEEE 802.11a. De autoria própria.

Tabela I
PARÂMETROS DO SINAL OFDM NO PADRÃO IEEE 802.11A [5]

Parâmetro	Valor
Largura de Banda	20 MHz
Comprimento da FFT (L_{FFT})	64
Subportadoras Totais (N_{ST})	52
Subportadoras de Dados (N_{SD})	48
Subportadoras Piloto (N_{SP})	4
Frequência de Espaçamento entre Subportadoras (F_{SP})	312.5 kHz (20 MHz/64)
Banda de Guarda das Subportadoras	12 ($L_{FFT} - N_{ST}$)
Comprimento do Prefixo Cíclico	16

C. PAPR

No domínio do tempo, o sinal transmitido é formado por uma soma de senoides janeladas com frequências múltiplas de $1/T_s$. Essa soma ocasiona uma grande variação de amplitude no sinal e esta é diretamente proporcional à quantidade de subportadoras. Isso faz com que haja valores de pico bem maiores que a média, que origina uma elevada PAPR (PAPR, sigla inglesa para *Peak-to-Average Power Ratio*; cuja tradução literal é a razão entre potência de pico e potência média). Nesse caso, pode haver degradações quando o sinal passar por um amplificador (não-linear) de alta potência [6].

III. PROJETO DE FILTROS FIR COM JANELA

Portanto, para combater o desperdício de espectro nos sistemas baseados no padrão IEEE 802.11a, esse trabalho propõe, de forma inovadora, com base no melhor conhecimento que se tem para o problema, o emprego de filtros digitais do tipo FIR com janela de Nuttall para filtrar as emissões fora da banda de sinais OFDM baseados no Padrão IEEE 802.11a, levando-se à obtenção de um sinal OFDM “filtrado”, f-OFDM (sigla inglesa para *filtered-OFDM*).

No presente estudo, a escolha do filtro digital do tipo FIR deu-se em função da sua característica de fase linear e da sua estabilidade. A partir disso, decidiu-se utilizar o método da janela para projetá-lo devido à facilidade de implementação e por haver um grande número de aplicações desse tipo de filtro, sendo que essa técnica já é bastante consolidada [7].

O filtro digital é do tipo passa-baixas, visto que o sinal está em banda base. Como o filtro passa-baixas ideal leva à

uma resposta ao impulso de duração infinita e não-causal, sua implementação se torna impraticável. Sendo assim, utiliza-se o método da janela para tornar possível a implementação de um filtro passa-baixas que não é mais ideal, mas que supre as necessidades do projeto. Detalhando melhor: tem-se que a resposta em frequência de um filtro passa-baixas ideal é dada pela Equação (3), onde ω_c é a frequência de corte e α é um número inteiro definido pela metade do valor da ordem do filtro ($\alpha = N/2$, pois isso garante a simetria da resposta ao impulso, que é a condição necessária para se ter fase linear) [8].

$$H_d(e^{j\omega}) = \begin{cases} e^{-j\omega\alpha}, & \text{se } |\omega| \leq \omega_c \\ 0, & \text{c.c.} \end{cases} \quad (3)$$

A resposta ao impulso correspondente do filtro ideal é dada pela Equação (4)

$$h_d[n] = \frac{\sin[\omega_c(n - \alpha)]}{\pi(n - \alpha)}, \quad -\infty < n < \infty. \quad (4)$$

A partir disso, é possível perceber que essa resposta ao impulso tem duração infinita, sendo esse tipo de sistema não-causal [8].

Portanto, utilizando-se o método da janela, esses dois problemas são resolvidos. Esse método consiste em multiplicar a resposta ao impulso ideal $h_d[n]$ por uma janela, $w[n]$, de modo que mantenha-se as propriedades de um filtro passa-baixas desejado, i.e.

$$h[n] = h_d[n] \cdot w[n]. \quad (5)$$

A maioria das funções usadas como janelas têm uma forma semelhante no domínio da frequência: um lóbulo principal e vários lóbulos laterais. A Figura 4 apresenta um comportamento típico de uma janela no domínio da frequência.

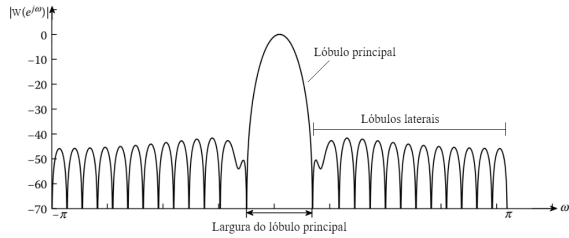


Figura 4. Características das janelas de um filtro FIR. Adaptado de [9].

Como resultado, o filtro desejado apresenta um módulo da resposta em frequência na banda passante que não é mais constante, enquanto que na banda de rejeição tal ganho não é mais igual a zero. Além disso, entre a banda de passagem e a banda de rejeição há uma faixa de transição.

Para o projeto de um filtro digital do tipo FIR com janela, as características mais importantes que devemos nos ater para a escolha de “boas” janelas são a largura do lóbulo principal, que contribui para a faixa de transição do filtro, $\Delta\omega$, e a amplitude máxima dos lóbulos laterais, que contribui para o nível de atenuação na banda de rejeição. O parâmetro $\Delta\omega$ é obtido através da Equação (6), em que K é uma constante determinada pela escolha da janela utilizada, e N é a ordem do filtro.

$$\Delta\omega = \frac{K}{N} \quad (6)$$

A. janela de Nuttall

A janela escolhida para esse estudo foi a de Nuttall, que apresenta uma elevada atenuação na banda de rejeição, de $-98,17$ dB [9]. Esse nível de atenuação é uma característica relevante para mitigar as emissões fora da banda dos sistemas OFDM baseados no padrão IEEE 802.11a.

A função janela de Nuttall é definida pela Equação (7) [9]

$$w[n] = 0,3635819 - 0,4891775 \cdot \cos\left(2\pi \frac{n}{N-1}\right) + 0,1365995 \cdot \cos\left(4\pi \frac{n}{N-1}\right) - 0,0106411 \cdot \cos\left(6\pi \frac{n}{N-1}\right). \quad (7)$$

Na comparação com janelas clássicas, janelas retangular e blackman, é possível observar na Figura 5 o comportamento no domínio da frequência apresentado para essas janelas, cujos resultados aqui exibidos foram obtidos através da ferramenta Matlab®. Logo, a janela de Nuttall tem uma menor amplitude máxima dos lóbulos laterais quando comparada com as janelas clássicas de Blackman e Retangular, isso influencia numa maior atenuação na banda de rejeição. Contudo, a largura do lóbulo principal é um pouco maior que a de Blackman, gerando um filtro com uma faixa de transição um pouco maior.

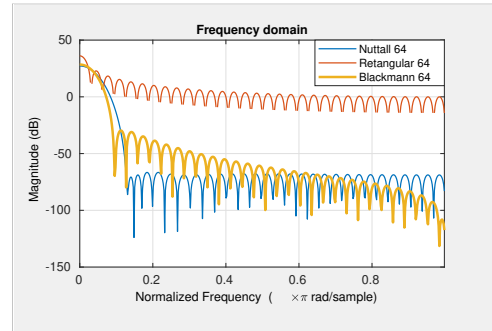


Figura 5. Comparação entre janelas usando o Matlab®. De autoria própria.

Embora a avaliação da PAPR não seja o foco deste trabalho, este parâmetro é um ponto importante no projeto do filtro digital para sistemas f-OFDM, visto que qualquer que seja o filtro digital empregado, esse efetivamente elevará a PAPR. Isso ocorre porque o filtro diminui a potência média, enquanto a potência de pico não é alterada [6].

Sendo assim, esse trabalho tem o objetivo de utilizar o método da janela no projeto de filtros digitais de resposta ao impulso finita (FIR) que leve a uma boa relação de custo-benefício entre a redução da emissão fora de banda e o aumento indesejado da PAPR causado pelo emprego do filtro digital.

Para avaliar o desempenho do uso de filtros, foram utilizados cinco filtros FIR com a janela de Nuttall. Cada um com uma das seguintes ordens: 256, 512, 1024, 2048 e 4096.

IV. RESULTADOS

Para a elaboração do projeto foi utilizada a ferramenta de modelagem Simulink®. Dentro desta ferramenta, inicialmente, foi adicionado um bloco gerador de sinais OFDM, criado através de uma ferramenta do Matlab® chamada de “Wireless Waveform Generator”, utilizando-se dos parâmetros apresentados na Tabela I, além do número de símbolos OFDM igual a 100 e da modulação QAM.

Para os filtros, foi utilizado o bloco “Digital Filter Design” com frequência de corte de 8,3 MHz, definida pela largura de banda destinada aos dados úteis do padrão IEEE 802.11a. Em seguida, para visualizar o espectro da frequência de cada sinal, foi utilizado o bloco “Spectrum Analyzer”. Por último, para medir e apresentar a PAPR foram utilizados, respectivamente, os blocos “Power Meter” e “Scope”, que encontram-se nas *toolboxes* “DSP System”, “Signal Processing” e “Communications” [10].

As Figuras 6 e 7 apresentam os resultados obtidos.

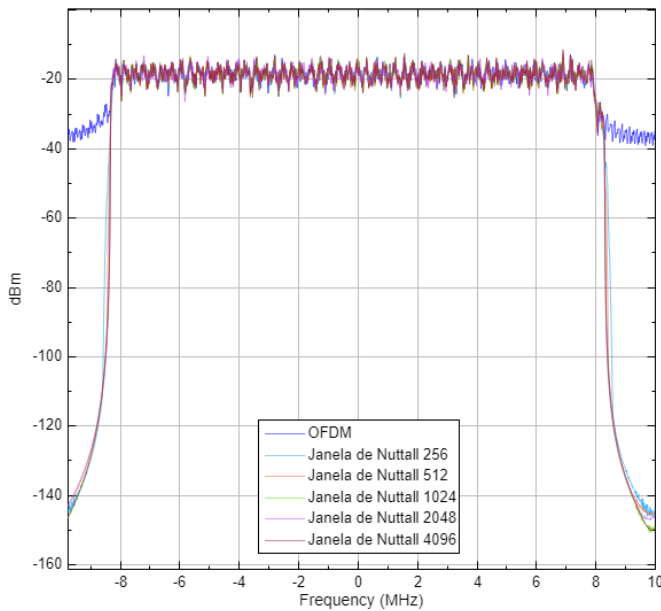


Figura 6. Espectro da Frequência do sinal OFDM original e dos sinais f-OFDM. De autoria própria.

Por meio dos resultados de desempenho obtidos, que estão ilustrados na Figura 6, é possível constatar que os filtros FIR com a janela de Nuttall avaliados nesse estudo levaram à obtenção de níveis de mitigação elevados para as emissões fora da banda, chegando a valores entre -140 e -150 dBm.

Admitindo-se que uma boa atenuação para supressão de interferências nos sistemas modernos de comunicações sem fio situa-se em torno de -100 dBm [11], então qualquer escolha de um dos filtros FIR com janela de Nuttall avaliados (independente da ordem) leva ao mesmo resultado, da redução notável das emissões fora da banda. Contudo, visando diminuir os requisitos de memória e latência, pode-se escolher o filtro FIR de ordem menor, 256.

A Figura 7 apresenta os resultados das avaliações da PAPR para cada um dos filtros FIR com janela de Nuttall propo-

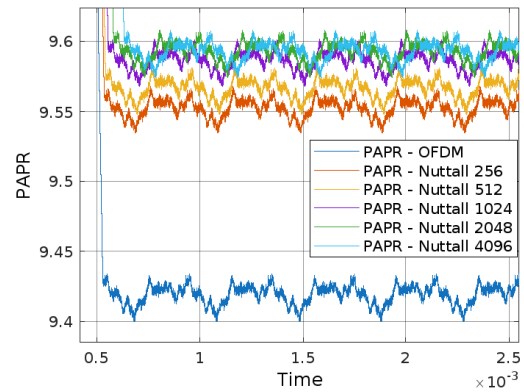


Figura 7. PAPR do sinal OFDM original e dos sinais f-OFDM. De autoria própria.

tos. Inicialmente, comprova-se o que foi visto na teoria: a utilização de filtro aumenta a PAPR. Portanto, o filtro FIR com ordem igual a 256 é o filtro digital que causa o menor aumento da PAPR.

Desse modo, conclui-se que, dos filtros avaliados, o de ordem 256 é o que leva à melhor relação custo-benefício entre OOB e PAPR para o padrão IEEE 802.11a.

V. CONCLUSÃO

Este trabalho propõe o uso de filtros FIR com janela de Nuttall para supressão das emissões fora da banda de sinais OFDM do padrão IEEE 802.11a. A partir dos resultados obtidos, foi possível perceber que o uso de um filtro digital do tipo FIR com janela de Nuttall e ordem igual a 256 é capaz de mitigar as emissões fora de banda, ao mesmo tempo que produz um aumento menor na PAPR.

REFERÊNCIAS

- [1] PINTO, Ernesto L. e ALBUQUERQUE, Claudio P., *A Técnica de Transmissão OFDM*. Revista Científica Periódica - Telecomunicações - Volume 05 - Número 01, ISSN 1516-2338. Junho, 2002.
- [2] HARRIS, Fredric J. e SKLAR, Bernard, *Digital Communications: Fundamentals and Applications*, 3ª ed. Pearson. Dezembro, 2020.
- [3] Silicon DSP Corporation, *Modulation for Multipath: Orthogonal Frequency Division Multiplex (OFDM)*. 2007.
- [4] SKLAR, Bernard, *The ABCs of OFDM - Part 2*. IEEE Presentation. Março, 2021.
- [5] Gast, M., *802.11 Wireless Networks. The Definitive Guide*, O'Reilly Media, 2005.
- [6] PAREDES, Martha C. P. e GARCÍA, M. Julia Fernandez-Getino, *The Problem of Peak-to-Average Power Ratio in OFDM Systems*. Mar, 2015.
- [7] OPPENHEIM, A. V.; SCHAFER, R. W. *Processamento em tempo discreto de sinais*. 3ª ed. São Paulo: Pearson, 2013.
- [8] HIGUTI, Ricardo T., *Filtros Digitais Tipo FIR*, 1ª ed. Notas de Aula da disciplina de Processamento Digital de Sinais, FEIS - Unesp. 2022. Disponível em: https://www.feis.unesp.br/Home/departamentos/engenhariaeletrica/ele1095_8b_filtrosfir_v2.pdf. Acesso em: 25 abril 2024.
- [9] PRABHU K. M. M., *Windows Functions and Their Applications in Signal Processing*, 1ª ed. CRC Press, 2014.
- [10] The MathWorks, Inc., *Online*. Disponível em: <https://www.mathworks.com/>. Acessado em Maio de 2024.
- [11] MARTÍN-MARTÍN, P.; BREGOVIC, R.; MARTÍN-MARCOS, A.; CRUZ-ROLDÁN, F. e SARMAKI, T., *A Generalized Window Approach for Designing Transmultiplexers*, em IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 55, no. 9, pp. 2696-2706, Outubro de 2008, doi: 10.1109/TCSI.2008.920095.

Gerenciamento de Largura de Banda em Sistemas 3GPP: Uma Análise de Desempenho

Eriberto Silva, Iago Rego, Vicente Sousa, Dario Vieira e Augusto Venâncio.

Resumo—A colisão entre transmissões de diferentes células é o principal efeito colateral do reuso de frequência, gerando a interferência co-canal. A disposição dos usuários também tem efeito negativo sobre o desempenho do sistema. Um grande número de usuários concentrados espacialmente (gerando um *hotspot*) pode causar uma maior possibilidade de transmitirem na mesma frequência (maior interferência co-canal) e de receberem uma menor potência útil, pois a potência útil pode ser dividida por mais usuários. Como a capacidade do canal de comunicação depende tanto da largura de banda quanto da qualidade do canal (que é inversamente proporcional a interferência), o gerenciamento de largura de banda é crítico quando existe interferência variante espacialmente na célula, como é o caso do cenário com *hotspots*. Este trabalho aborda técnicas de coordenação de interferência entre células (ICIC) em um cenário dinâmico com a presença de *hotspots*. O impacto do *hotspot* é discutido à luz da distribuição de banda entre borda e centro da célula. As investigações foram efetuadas utilizando o simulador de redes *ns-3*. Pelos resultados obtidos é possível verificar a necessidade de, além do uso de ICIC, alocar de forma dinâmica a distribuição de banda.

Index Terms—ICIC, largura de banda, LTE, *hotspots*

I. INTRODUÇÃO

A maioria dos sistemas modernos de comunicação utilizam fator de reuso de frequência igual a 1, ou seja, as células reutilizam todo o espectro disponível. No entanto, essa maneira de reutilização de frequência não traz somente benefícios e deve ser utilizada com cuidado, especialmente em redes com muitas estações bases próximas. Nessa situação, a interferência co-canal, ou *Inter-Cell Interference* (ICI), pode aumentar de maneira descontrolada, degradando a *SINR* (*Signal-to-Interference-plus-Noise Ratio*) e, conseqüentemente, a qualidade de serviço (QoS) em termos de taxa útil e taxa de erro. Por outro lado, um bom planejamento do fator de reuso reduz a ICI e aumenta a SINR, levando a melhor QoS, mesmo considerando a diminuição de largura de banda útil da célula, devido a um reuso de frequência menor que 1 [1].

Os sistemas 3GPP têm flexibilidade no gerenciamento de banda e de interferência utilizando, por exemplo, estratégias como MIMO *beamforming*, controle de potência, *bandwidth parts*, uso de *small cells* e o Reuso Fracionário de Frequência (FFR). Especificamente relacionadas a gerência de interferência, técnicas de coordenação de ICI foram desenvolvidas como forma de mitigar os efeitos da interferência, atuando na alocação de banda, na potência de transmissão e na distribuição de usuários por regiões da

célula. Essas técnicas são chamadas de *Inter-Cell Interference Coordination* (ICIC).

Esse gerenciamento de interferência torna-se ainda mais complexo em cenários dinâmicos, nos quais, frequentemente, surgem picos de tráfego inesperados devido as concentrações de usuários decorrentes de eventos aleatórios, por exemplo, eventos artísticos, feiras, jogos, dentre outros. Esses aglomerados de usuários são denominadas como *hotspot* e, geralmente, aparecem de forma aleatória, disputando os recursos disponibilizados aos usuários da região [2].

A aleatoriedade do *hotspot* ocorre tanto em posição quanto em quantidade de dispositivos, sendo o seu surgimento capaz de alterar a capacidade projetada do sistema e aumentar a disputa dos usuários pelos recursos de rádio. Além disso, é possível experimentar um aumento de ICI, principalmente para os usuários próximo à borda da célula, que também experimentam maior degradação de sinal [3].

Neste artigo, propõe-se um estudo do comportamento de um sistema de comunicação móvel da família do 3GPP ao se aplicar a técnica de ICIC em um cenário dinâmico com o surgimento de *hotspots*. A técnica de ICIC alvo do estudo faz uso do FFR, permitindo o gerenciamento da alocação de banda e da potência em diferentes áreas da célula, visando maior eficiência espectral.

II. TRABALHOS RELACIONADOS

A. Análise Quantitativa

A análise quantitativa é materializada pela exposição da quantidade de artigos, fruto da busca por palavras-chaves nas bases de dados Scopus e Web of Science (WoS), visando quantificar a relevância do tema abordado neste artigo. Como mostra a Tabela I, a estratégia de busca foi definida em três grupos de palavras-chaves e escopo.

Tabela I: Número de artigos por grupo de palavras-chaves.

Grupo	Escopo da Busca/Palavras-Chaves	Scopus	WoS
#1	Title("ICIC") OR	490	303
	Title("Inter-Cell Interference Coordination") OR		
	Title("fractional frequency reuse")		
#2	Title("ICIC") OR	267	218
	Title("Inter-Cell Interference Coordination") OR		
	Title("fractional frequency reuse") AND Abstract("Performance")		
#3	Title("ICIC") OR	4	51
	Title("Inter-Cell Interference Coordination") OR		
	Title("fractional frequency reuse") AND Abstract("Performance") AND All fields("ns-3")		

Dario é do Efrei Research Lab, Paris, France. E-mail: dario.vieira@efrei.fr. Os demais autores são do Leading Advanced Technologies Center of Excellence (LANCE), UFRN, Brasil. E-mails: {eriberto.souto.095, iago.diogenes.072}@ufrn.edu.br, vicente.sousa@ufrn.br, augusto@dimap.ufrn.br. Esse estudo foi financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A busca do Grupo #1 retornou o maior número de artigos, tendo em vista que as palavras-chaves não restringem a busca a um aspecto específico sobre ICIC.

Refinando a pesquisa ao incluir o termo *Performance* no campo *Abstract* dos artigos (Grupo #2), o número de publicações foi menor, pois vários artigos do tipo *survey* foram excluídos, retornando somente os artigos com algum estudo de desempenho sendo explorado.

O objetivo do Grupo #3 é restringir ainda mais a busca ao incluir o termo *ns-3*, simulador utilizado neste trabalho. A quantidade de artigos reduziu consideravelmente, especialmente na Scopus, com 4 trabalhos, em que somente um deles versa sobre *hotspot*.

Enquanto o *ICIC* explora a dimensão da frequência ao gerenciar a largura de banda, o *e-ICIC* (*enhanced Inter-Cell Interference Coordination*) silencia *slots* em *downlink* de algumas células ou setores para evitar interferência co-canal, enquanto o *Fe-ICIC* (*Further enhanced Inter-Cell Interference Coordination*) faz o mesmo para *uplink*. A Fig. 1 mostra a quantidade de artigos na base WoS para cada tipo de algoritmo de ICIC. É possível observar um número de artigos proporcional ao tempo que cada técnica foi proposta pelo 3GPP: *ICIC* foi proposto no *Release 8/9*, o *e-ICIC* no *Release 10* e o *Fe-ICIC* no *Release 11*.

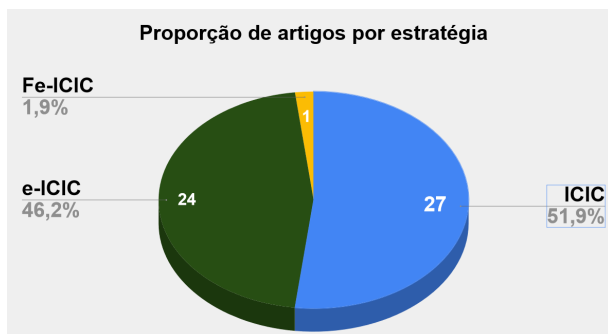


Figura 1: Quantidade de artigos por estratégia ICIC.

Este artigo foca no *ICIC* devido a sua simplicidade de implementação em sistemas 4G e 5G, bem como na ausência de trabalhos que explorem a distribuição de banda do *ICIC* em cenários com *hotspot*.

B. Análise Qualitativa

Para essa seção foram selecionados os artigos diretamente relacionados a *ICIC* utilizando reuso de frequência com alocação de banda.

O artigo [4] propõe um método de *ICIC* que gerencia dinamicamente a largura de banda da borda da célula. O método se baseia no número de usuários conectados e o uso da interface X2 para troca de informações entre células.

O trabalho [5] propõe um novo esquema de reuso e alocação de frequência chamado *First Tier Interference Free (FTIF)*. Ele setoriza as células sem reduzir a potência de transmissão, eliminando a região de interferência entre a célula adjacente e a borda da célula vizinha.

Em [1], os autores apresentam um *survey* acompanhado de uma análise de desempenho de técnicas de *ICIC* para diferentes condições de rádio e cargas da rede. O artigo apresenta uma análise comparativa entre o FFR e o *Strict Frequency Reuse*

(SFR), discutindo a natureza estática de cada algoritmo e seus desempenhos perante aos reúsos 1 e 3.

Por fim, os autores de [3] abordam a interferência entre células em cenários dinâmicos com e sem a presença de *hotspot*, avaliando os danos que ele causa no desempenho do sistema LTE. Uma solução utilizando aprendizado de máquina é dedicada ao ajuste da distribuição de usuários entre as diferentes zonas de reuso da célula. A proposta é baseada em *Q-Learning* e ajusta os parâmetros de distribuição de usuários entre o centro e a borda da célula de forma dinâmica, conforme o surgimento dos *hotspots*. O trabalho [3] não explora a gerência de largura de banda do *ICIC*.

Apesar das publicações identificadas, a literatura é carente de trabalhos sobre técnicas de *ICIC* que façam alocação de banda dinâmica entre centro e borda da célula em cenários com presença de *hotspot*. Contribuir para o preenchimento dessa lacuna é a principal contribuição deste artigo.

Esse trabalho aborda o uso do algoritmo de SFR, alocando diferentes distribuições de banda entre centro e borda das células. Essa avaliação contribui para a investigação do comportamento da gerência espacial de banda com objetivo de mitigar os efeitos da ICI, agravados pela presença dos *hotspots*.

III. SOLUÇÃO DE FFR E METODOLOGIA DE PROTOTIPAGEM

O esquema SFR consiste em dividir a banda total do sistema em uma sub-banda central compartilhada e em uma sub-banda privada de borda, com largura de banda e potências distintas. Os usuários são alocados para cada sub-banda dependendo de sua RSRQ, obedecendo um limite preestabelecido.

O sistema multicelular alvo da análise de desempenho deste trabalho foi prototipado no Network Simulator (*ns-3*). Ele é um simulador de uso gratuito e de código aberto, bastante utilizado no meio acadêmico e em prototipagem na indústria de telecomunicações. Além disso, possui uma comunidade ativa e colaborativa na solução de problemas em experimentos utilizando o simulador.

O *ns-3* suporta algoritmos de *ICIC* por intermédio da classe abstrata *LteFfrAlgorithm*. Além do SFR, existem mais seis algoritmos de FFR implementados, descritos brevemente a seguir [6]:

- *ns3::LteFrNoOpAlgorithm*: algoritmo padrão com reuso 1;
- *ns3::LteFrHardAlgorithm*: algoritmo com reuso 3, com três estações base com mesma potência;
- *ns3::LteFrStrictAlgorithm*: a banda total é dividida em sub-banda central compartilhada e em sub-banda privada de borda, com larguras de banda e potências distintas;
- *ns3::LteFrSoftAlgorithm*: divide a banda em duas regiões, mas a região de borda também é compartilhada;
- *ns3::LteFfrSoftFractionalAlgorithm*: a célula é subdivida em três regiões distintas: central, média (reuso 1) e externa (reuso 3). A região central reutiliza frequências das bordas das células adjacentes;
- *ns3::LteFfrEnhancedAlgorithm*: cada célula adjacente recebe uma parte do espectro com frequências ortogonais (o segmento primário, que tem partes com reuso 1 e 3), e os demais canais são os segmentos secundários (compartilhados);
- *ns3::LteFfrDistributedAlgorithm*: Seleciona *Physical Resource Blocks* (RBs) para os usuários de borda baseado

nas informações de células adjacentes (especialmente a distribuição de *RSRP*).

É possível selecionar o algoritmo de ICIC e configurar os atributos necessários para seu funcionamento. Os parâmetros disponíveis dependem de cada algoritmo, mas geralmente, exceto para a opção *ns3::LteFrNoOpAlgorithm*, é possível definir a largura de banda e a potência de cada região de reuso da célula (centro e borda). Outro parâmetro importante é o limiar *RsrqThreshold*. Ele determina se o usuário será alocado para a região do centro ou da borda da célula, dependendo de seu valor de *Reference Signal Received Quality* (RSRQ) [6].

O algoritmo de *ICIC* é selecionado no objeto *LteHelper*, utilizando o método *SetFfrAlgorithmType*. A largura de banda total de *uplink* e de *downlink* são configuradas alterando as variáveis *DLBandwidth* e *ULBandwidth* no *LteEnbNetDevice*. Para este trabalho, optou-se pelo algoritmo *ns3::LteFrStrictAlgorithm* (SFR), pois apresenta equilíbrio de desempenho em diferentes cenários e são eficientes na mitigação dos efeitos de ICI [7]. A Fig. 2 ilustra como configurar o algoritmo no ns-3, exemplificando a distribuição de banda 64-12, ou seja, 64 PRBs na sub-banda compartilhada (região do centro da célula) e 12 PRBs em cada sub-banda privada (região da borda de três células vizinhas).

O método *SetBandwidth* (Fig. 2b) foi construído para facilitar essa configuração. Note que três células são configuradas de acordo com o cenário da Fig. 3. As sub-bandas também são servidas com níveis de potência diferentes, com a potência de transmissão para a região do centro configurada para -6 dB e, para as bordas, 3 dB, relativos à potência do sinal de referência [9]. O critério de RSRQ (*RsrqThreshold*), que determina para qual sub-banda um usuário é alocado, é um parâmetro de entrada e é configurado com o método *SetFfrAlgorithmAttribute*.

IV. CENÁRIOS DE AVALIAÇÃO

O trabalho analisa o comportamento do sistema quando os *hotspots* surgem próximos e distantes das estações base. Foram adotados dois cenários com diferentes posicionamentos dos *hotspots*, mas mantendo a dimensão de célula, bem como a distância entre elas, como ilustrado na Fig. 3. O cenário celular é composto por 3 estações bases distanciadas entre si por 500 metros. Os usuários são distribuídos em dois grupos: (i) distribuição uniforme, com 30 usuários uniformemente distribuídos por toda região da célula; e (ii) *hotspots*, que delimitam usuários distribuídos em círculos com raio de 50 m. Existem 3 *hotspots* com 20 usuários cada. A Fig. 3 ilustra os dois cenários, com os *hotspots* na cor vermelha, posicionados a 50 m da estação base, representando o Cenário 1. O Cenário 2 é representado pelos *hotspots* na cor preta, posicionados a 450 m da estação base.

A largura de banda do sistema é de 100 PRBs (20 MHz) e é distribuída na seguinte proporção: 1 fração direcionada ao centro e 3 frações da banda iguais para cada borda das três células vizinhas (reuso 3). As proporções de banda utilizadas nos experimentos são: 16-28, 28-24, 40-20, 52-16 e 64-12.

Os valores 30, 31 e 32 foram utilizados como limiar de *RsrqThreshold*. Segundo [10], utilizar o limiar 30 significa que os usuários com valores de *RSRQ* menores que -4,5 dB são alocados na borda. Os usuários com valores de *RSRQ* maiores usam os recursos de rádio da região central. A mesma regra é

```
// ***** FFR ALGORITHM CONFIGURATION *****
lteHelper->SetFfrAlgorithmType ("ns3::LteFrStrictAlgorithm");
NS_LOG_DEBUG ("FfrAlgorithmType: " << ffrAlgorithmType);

if (frAlgorithmType == "LteFrHardAlgorithm")
{
    //Nothing to configure here in automatic mode
}
else if (frAlgorithmType == "LteFrStrictAlgorithm")
{
    lteHelper->SetFfrAlgorithmAttribute("RsrqThreshold", UIntegerValue(RsrqThreshold));

    if (CenterPowerOffset == 0)
    {
        lteHelper->SetFfrAlgorithmAttribute("CenterPowerOffset",
                                           UIntegerValue(LteRrcSap::PdschConfigDedicated::dB_6));
    }
    else if (CenterPowerOffset == 1)
    {
        lteHelper->SetFfrAlgorithmAttribute("CenterPowerOffset",
                                           UIntegerValue(LteRrcSap::PdschConfigDedicated::dB3));
    }

    lteHelper->SetFfrAlgorithmAttribute("EdgePowerOffset",
                                       UIntegerValue(LteRrcSap::PdschConfigDedicated::dB3));
    lteHelper->SetFfrAlgorithmAttribute("CenterAreaTpc", UIntegerValue(0));
    lteHelper->SetFfrAlgorithmAttribute("EdgeAreaTpc", UIntegerValue(3));

    if (BandwidthDistribution == "64-12")
    {
        SetBandwidth(lteHelper, enbDevs, enbNodes, 64, 12, bandwidth);
    }
    else if (BandwidthDistribution == "52-16")
    {
        SetBandwidth(lteHelper, enbDevs, enbNodes, 52, 16, bandwidth);
    }
}
```

(a) Configuração de sub-bandas do SFR com *SetBandwidth*.

```
void
SetBandwidth(Ptr<LteHelper> lteHelper,
             NetDeviceContainer& enbDevs,
             NodeContainer enbNodes,
             int centerBW,
             int edgeBW,
             int bw)
{
    lteHelper->SetFfrAlgorithmAttribute("FfrCellTypeId", UIntegerValue(0));

    lteHelper->SetEnbDeviceAttribute("DLBandwidth", UIntegerValue(bw));
    lteHelper->SetEnbDeviceAttribute("ULBandwidth", UIntegerValue(bw));

    //configuring eNB 1
    lteHelper->SetFfrAlgorithmAttribute("DLCommonSubBandwidth", UIntegerValue(centerBW));
    lteHelper->SetFfrAlgorithmAttribute("DLEdgeSubBandOffset", UIntegerValue(0));
    lteHelper->SetFfrAlgorithmAttribute("DLEdgeSubBandwidth", UIntegerValue(edgeBW));

    lteHelper->SetFfrAlgorithmAttribute("ULCommonSubBandwidth", UIntegerValue(centerBW));
    lteHelper->SetFfrAlgorithmAttribute("ULEdgeSubBandOffset", UIntegerValue(edgeBW));
    lteHelper->SetFfrAlgorithmAttribute("ULEdgeSubBandwidth", UIntegerValue(edgeBW));
    enbDevs.Add(lteHelper->InstallEnbDevice(enbNodes.Get(0)));

    //configuring eNB 2
    lteHelper->SetFfrAlgorithmAttribute("DLCommonSubBandwidth", UIntegerValue(centerBW));
    lteHelper->SetFfrAlgorithmAttribute("DLEdgeSubBandOffset", UIntegerValue(edgeBW));
    lteHelper->SetFfrAlgorithmAttribute("DLEdgeSubBandwidth", UIntegerValue(edgeBW));

    lteHelper->SetFfrAlgorithmAttribute("ULCommonSubBandwidth", UIntegerValue(centerBW));
    lteHelper->SetFfrAlgorithmAttribute("ULEdgeSubBandOffset", UIntegerValue(edgeBW));
    lteHelper->SetFfrAlgorithmAttribute("ULEdgeSubBandwidth", UIntegerValue(edgeBW));
    enbDevs.Add(lteHelper->InstallEnbDevice(enbNodes.Get(1)));

    //configuring eNB 3
    lteHelper->SetFfrAlgorithmAttribute("DLCommonSubBandwidth", UIntegerValue(centerBW));
    lteHelper->SetFfrAlgorithmAttribute("DLEdgeSubBandOffset", UIntegerValue(2*edgeBW));
    lteHelper->SetFfrAlgorithmAttribute("DLEdgeSubBandwidth", UIntegerValue(edgeBW));

    lteHelper->SetFfrAlgorithmAttribute("ULCommonSubBandwidth", UIntegerValue(centerBW));
    lteHelper->SetFfrAlgorithmAttribute("ULEdgeSubBandOffset", UIntegerValue(2*edgeBW));
    lteHelper->SetFfrAlgorithmAttribute("ULEdgeSubBandwidth", UIntegerValue(edgeBW));
    enbDevs.Add(lteHelper->InstallEnbDevice(enbNodes.Get(2)));
}
```

(b) Método *SetBandwidth*.

Figura 2: Configuração do algoritmo SFR (*ns3::LteFrStrictAlgorithm*) no ns-3 [8].

utilizada ao definir os limiares como 31 e 32 que são -4 dB e -3,5 dB, respectivamente.

A Tabela II apresenta as demais configurações utilizadas nos experimentos.

Tabela II: Parâmetros do Experimento.

Parâmetro	Valor
Usuários uniformes	30
Usuários por <i>hotspot</i>	20
Algoritmo de escalonamento	<i>Proportional Fair</i>
Modo de transmissão (MIMO)	4x4
Taxas oferecidas (Mbps)	2, 1 e 0,5

V. RESULTADOS

Antes de analisar os resultados de *throughput*, a Fig. 4 apresenta a ECDF (*Empirical Cumulative Distribution*

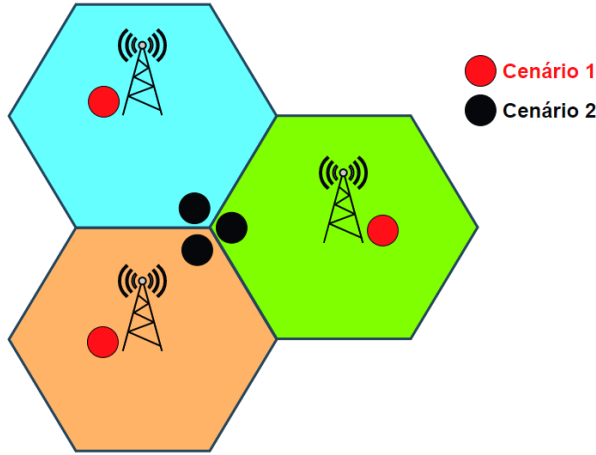


Figura 3: Posicionamento dos *hotspots* nos cenários avaliados.

Function) da *RSRQ* para *RsrqThreshold* igual a 30. Especialmente para esse gráfico, o cenário sem *hotspot* (somente com usuários distribuídos uniformemente) é mostrado. Como esperado, é possível observar que o Cenário 2 (*hotspots* longe) tem os menores valores de *RSRQ*, e o Cenário 1 (*hotspots* perto) tem os maiores valores. A distribuição de *RSRQ* tem comportamento similar para todas as cargas e limiares de *RSRQ*. Assim, ao variar o limiar entre 30 e 32, mais usuários são alocados para a borda da célula. Esse comportamento é mais evidente para o Cenário 2.

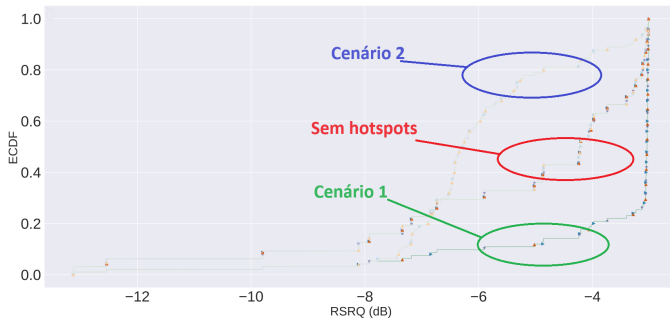


Figura 4: ECDF da *RSRQ* - *RSRQThreshold* 30.

Os resultados de *throughput* são apresentados nas Figs. 5 e 6, para *RsrqThreshold* igual a 30; Figs. 7 e 8, para *RsrqThreshold* igual a 31; e, finalmente, Figs. 9 e 10, para *RsrqThreshold* igual a 32. Os pares de figuras apresentam resultados para os Cenários 1 e 2. As barras representam o *throughput* de célula. Cada conjunto de três barras representa uma distribuição de banda, sendo cada barra desse conjunto o *throughput* para os valores das taxas oferecidas 2 Mbps, 1 Mbps e 500 kbps da esquerda para à direita, respectivamente.

Os experimentos com *RSRQThreshold* igual a 30 revelam que ao comparar o *throughput* dos Cenários 1 e 2, o melhor desempenho é alcançado com diferentes configurações de distribuição de banda, como identificado nas Figs. 5 e 6. Se os *hotspots* estão próximos à estação base (Cenário 1), alocar mais banda para o centro da célula (distribuição 64-12) resulta em maior *throughput*. Contudo, quando os *hotspots* estão afastados, o maior valor de *throughput* é alcançado com a distribuição de banda 28-24, uma alocação equilibrada entre borda e centro da célula.

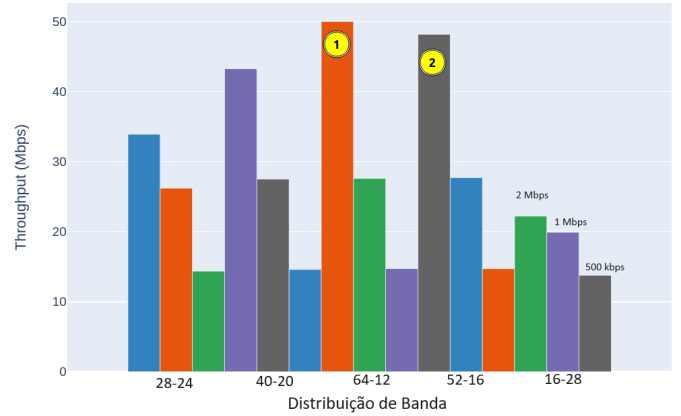


Figura 5: Cenário 1: *Throughput* de célula - *RSRQThreshold* 30.

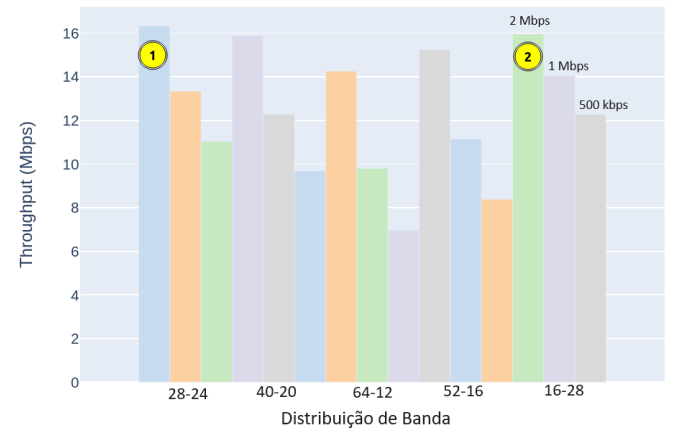


Figura 6: Cenário 2: *Throughput* de célula - *RSRQThreshold* 30.

Os resultados do Cenário 1 para *RSRQThreshold* igual a 31 (Figs. 7 e 8) são semelhantes ao anterior, com maior *throughput* para distribuição de banda 64-12 (mais banda para o centro da célula). Contudo, diferente do primeiro conjunto de resultados do Cenário 2, a melhor distribuição de banda é a 16-28, uma configuração com mais banda para borda da célula.

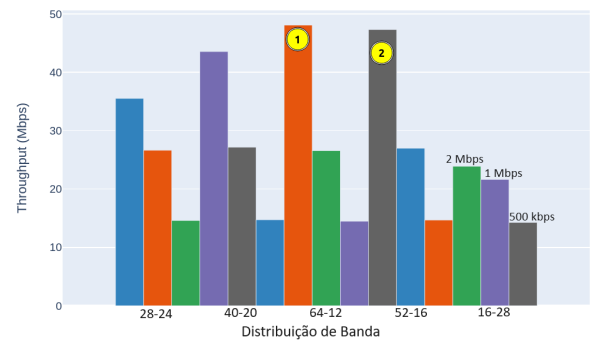


Figura 7: Cenário 1: *Throughput* de célula - *RSRQThreshold* 31.

Por fim, os resultados dos experimentos para *RSRQThreshold* igual a 32 mostram que as mesmas distribuições de banda para *RSRQThreshold* igual a 31 resultam nos melhores *throughputs*. Esse conjunto de

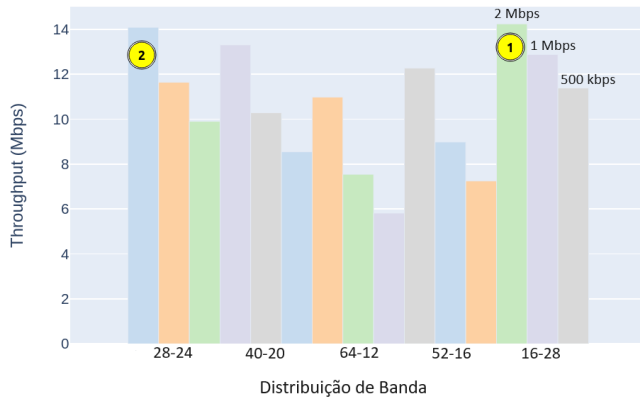


Figura 8: Cenário 2: *Throughput* de célula - *RSRQThreshold* 31.

resultados mostra que não existe um único valor ótimo de limiar *RSRQThreshold* e de distribuição de banda para todos os cenários de distribuição de *hotspots*.

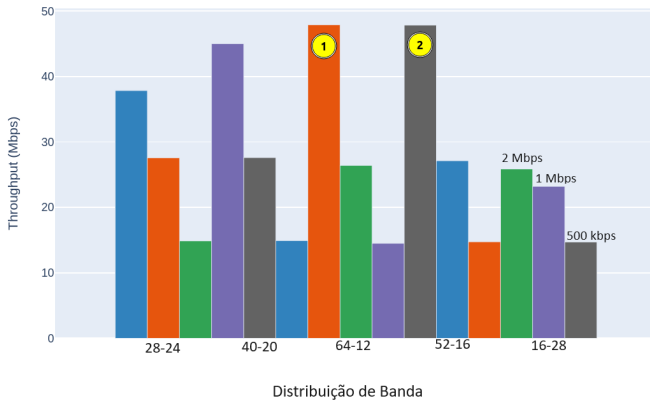


Figura 9: Cenário 1: *Throughput* de célula - *RSRQThreshold* 32.

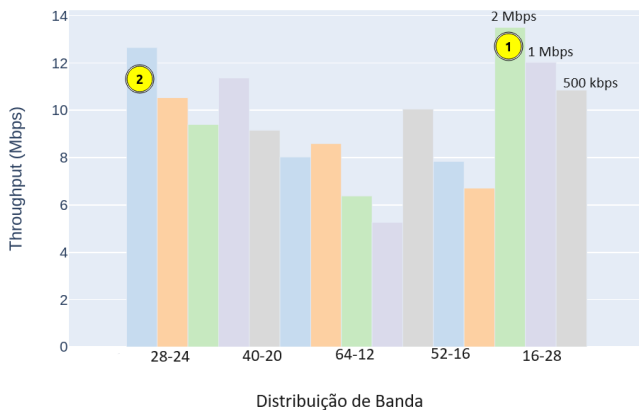


Figura 10: Cenário 2: *Throughput* de célula - *RSRQThreshold* 32.

VI. CONCLUSÕES

Este trabalho apresenta uma análise sistêmica de técnicas de ICIC em cenários com *hotspots*. Os resultados indicam que a variação da posição dos *hotspots* afeta significativamente o

throughput do sistema, e que se agrava quando se aumenta a taxa oferecida, chegando ao ponto de reduzir mais de 50% de desempenho ao comparar os dois cenários testados. Além disso, o trabalho apresenta o quanto o *throughput* do sistema é afetando quando não se adota uma alocação dinâmica de banda para os cenários dinâmicos. A Tabela III resume os resultados do artigo ao mostrar a *porcentagem* de redução do *throughput* do Cenário 1 (*hotspots* perto) em relação ao Cenário 2 (*hotspots* longe), i.e., quanto se perde ao mover o *hotspot* de perto para longe da estação base. A coluna *Melhor Escolha* mostra a redução de *throughput* ao se escolher a melhor distribuição de banda para ambos cenários. A coluna *Escolha Fixa* mostra a redução de *throughput* ao se escolher a melhor distribuição de banda do Cenário 1 e aplicá-la ao Cenário 2. É possível concluir que manter a mesma distribuição de banda nos dois cenários pode aumentar as perdas de *throughput* em mais de 17% (para o caso de *RSRQThreshold* 32 e 500 kbps).

Tabela III: Redução do *Throughput* entre os Cenários 1 e 2.

<i>RSRQThreshold</i> 30		
Taxa Oferecida	Melhor Escolha	Escolha Fixa
2 Mbps	67,34%	71,48%
1 Mbps	48,55%	59,77%
500 kbps	16,57%	52,78%
<i>RSRQThreshold</i> 31		
2 Mbps	70,37%	77,15%
1 Mbps	52,55%	62,13%
500 kbps	22,69%	42,06%
<i>RSRQThreshold</i> 32		
2 Mbps	71,80%	82,05%
1 Mbps	56,44%	66,84%
500 kbps	27,27%	44,76%

O presente trabalho abre possibilidades para futuras investigações voltadas a concepção de algoritmos de aprendizado de máquina aplicados a ICIC em cenários em que a posição do *hotspot* varia ao longo do tempo, algo que é comum em cenários reais. Trabalhos futuros explorarão o ajuste dinâmico da distribuição de banda utilizando técnicas de aprendizado por reforço.

REFERÊNCIAS

- [1] M. Yassin, M. A. AboulHassan *et al.*, "Survey of ICIC Techniques in LTE Networks under Various Mobile Environment Parameters," *Wireless Networks*, Springer Verlag, vol. 23, no. 2, pp. 403–418, 2017.
- [2] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*. Second edition, 2nd ed. Wiley, 2011.
- [3] I. D. do Rego and J. Vicente A. de Sousa, "Solution for Interference in Hotspot Scenarios Applying Q-Learning on FFR-Based ICIC Techniques," *Sensors*, 2021.
- [4] D. Kimura, Y. Harada, and H. Seki, "De-Centralized Dynamic ICIC using X2 Interfaces for Downlink LTE Systems," *IEEE, 73rd Vehicular Technology Conference (VTC Spring)*, 2011.
- [5] H. Wu, H. Gao, and X. Xiong, "ICIC Based on Adjacent Cells Interference Elimination for Cell Edge," *IEEE, 17th International Conference on Communication Technology (ICCT)*, 2017.
- [6] The University of Washington NS-3 Consortium, "Network simulator(ns-3)," ns-3, Tech. Rep., 2024. [Online]. Available: <https://www.nsnam.org/docs/models/html/lte-user.html>
- [7] I. D. do Rego, L. F. de O. Nobrega, and V. A. de Sousa Jr., "Análise de desempenho de técnicas ICIC usando o ns-3," *ECOP*, 2018.
- [8] The University of Washington NS-3 Consortium, "Alocação de recursos do model LTE-EPC," ns-3, Tech. Rep., 2024. [Online]. Available: <https://www.nsnam.org/docs/models/html/lte-design.html#resource-allocation-model>
- [9] 3GPP, "TS 36.133: Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); requirements for support of radio resource management (Release 8)," 3GPP, Technical Report, 2010.
- [10] The University of Washington NS-3 Consortium. (2024) 4g/lte - phy measurement. [Online]. Available: https://sharetechnote.com/html/Handbook_LTE_RSRQ.html

Licenciamento de Espectro para uso Científico: Legislação e Estudo de Caso de Sistema 5G Aberto

Antonio Campos, Vicente Sousa, João Oliveira, Matheus Dória, Carlos Lima, Nelson Oliveira, Paulo Filho, Paulo Eduardo, Ricardo Silva, Daniel R. de Luna, Marcelo Fernandes, Augusto Neto e Halysson Mendonça

Resumo—Com a disponibilidade de código fonte aberto, é possível criar protótipos de sistemas 5G utilizando plataformas de rádio definido por software (SDR) capazes de realizar transmissões OTA (*over the air*). Contudo, existe o desafio da concessão do espectro eletromagnético, que dependendo da faixa, precisa envolver diferentes esforços do requerente. Este trabalho tem o objetivo de descrever o processo de licenciamento de espectro para uso científico, compartilhando aspectos práticos e de legislação, tendo como referência um caso de uso real da Universidade Federal do Rio Grande do Norte. São descritos os passos para solicitação de autorização e todo processo junto ao órgão regulador ANATEL, bem como algumas medições necessárias. O estudo de caso trata da solicitação para uso em rede 5G experimental na UFRN, utilizando código aberto do OpenAirInterface e plataformas USRPs. Os aparatos de softwares e hardwares do *setup* são descritos, assim como os resultados das medições da radiação não ionizante (RNI) de sete estações base 5G são apresentados. Este artigo contribui para a melhor compreensão e a facilitação do processo licenciamento de espectro para uso científico, bem como na disseminação de protótipos de sistemas 5G utilizando código aberto, algo que se tornará cada vez mais comum em iniciativas de ensino, pesquisa e extensão em universidade e setor produtivo.

Index Terms—Licenciamento de espectro, Sistemas 5G, ANATEL, CONTEL, OpenAirInterface, USRP.

I. INTRODUÇÃO

Os aperfeiçoamentos das redes 5G e sua evolução para as redes 6G passam pela concepção e prototipagem de redes de acesso de rádio (RANs) abertas. Alinhado com os sistemas 3GPP, as RAN abertas tem dois objetivos principais [1]: (i) fornecer interoperabilidade de vários fornecedores entre diferentes componentes de hardware e software em um sistema de telecomunicações; e (ii) criar um ecossistema aberto para desenvolver, implantar e operar aplicativos de terceiros (chamados xApps e rApps) para oferecer suporte ao gerenciamento de recursos de rádio (RRM), otimização de procedimentos de camadas superiores, otimização de políticas em RAN e fornecer orientação, parâmetros, políticas e modelos de Inteligência Artificial/Machine Learning (AI/ML) para cortar custos, melhorar a Qualidade de Serviço (QoS) e gerar novos fluxos de receita [2]–[4]. Com esforços liderados pela O-RAN Alliance [1], é possível ter acesso gratuito a código funcional de uma rede 5G aberta, o que está aumentando a demanda no Brasil por pedidos de acesso ao espectro para uso científico e o acesso a investimentos em redes abertas [5].

Halysson Mendonça é da ANATEL, Natal, RN, E-mail: halysson@anatel.gov.br. Todos os demais autores são do Leading Advanced Technologies Center of Excellence (LANCE), Universidade Federal do Rio Grande do Norte, E-mails: {antonio.campos, vicente.sousa}@ufrn.br, {joao.guilherme.016, matheus.fagundes.067, carlos.lima.106, paulo.filho.071, paulo.eduardo.093, ricardo.queiroz.105, daniel.luna.088}@ufrn.edu.br, nelson@imd.ufrn.br, mfernandes@dca.ufrn.br e augusto@dimap.ufrn.br. Esse estudo foi financiado em parte pela Lenovo, com recursos de P&D de Lei de Informática, e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Por se tratar de um recurso limitado, é de extrema importância a existência de regulamentação que torne o uso do espectro coordenado e eficaz. No Brasil, a radiocomunicação foi regulamentada em maio de 1931 por meio do decreto nº 20.047 [6]. Esse decreto passou por atualizações e foi tornado sem efeito em 1962, ano em que foi aprovada a lei nº 4.117/62, que ficou conhecida como Código Brasileiro das Telecomunicações (CBT). Por fim, é necessário destacar que, a versão atual da regulamentação das telecomunicações no Brasil é a Lei Geral das Telecomunicações (LGT), Lei nº 9.472/97, que tem um papel crucial na regulação do espectro de RF [7].

Outra faceta relacionada ao espectro de radiofrequências (RF) é a fabricação de dispositivos capazes de operar em múltiplas bandas, considerando que cada país é livre para determinar a faixa ocupada por diversas tecnologias de comunicação sem fio, desde comunicação marítima e aeronáutica até telefonia celular e redes sem fio [8]. Otimizar o número de faixas que os dispositivos devem ser fabricados resulta em redução de custos e manutenção de operação em escala mundial, por exemplo, como o que acontece com sistemas Wi-Fi, pois opera em um conjunto de faixas de frequências bem definidas.

Finalmente, os sistemas 5G nasceram e estão evoluindo com a prerrogativa de virtualização. O núcleo de rede (5GC) é baseado em funções virtuais de rede (NFVs) que podem ser executados em hardwares padrão e não dedicados. Esse movimento também envolveu a RAN, mesmo necessitando de processamento muito sensível ao atraso, foi virtualizada. Dessa forma, parte da RAN 5G continua sendo executada em hardware dedicado (e.g., em FPGAs, DSPs, placas aceleradoras e SDRs), mas a implementação de camadas superiores a PHY são feitas inteiramente em software, rodando em servidores ou computadores tradicionais. Essa perspectiva, aliada à disponibilização de código livre de sistemas 5G (e.g., OAI [9] e srsRAN [10]) traz a tendência de uso exaustivo de sistemas 5G reais em iniciativas de ensino, pesquisa e prototipagem, seja por universidades ou por fabricantes de dispositivos. Assim, somado à metodologia de prototipagem via simuladores e emuladores, com o código aberto de sistemas 5G, existe a facilidade de uso de dispositivos SDR (e.g., USRPs [11] e Cyan Per Vices [12]) para gerar protótipos que incluem a geração do sinal eletromagnético. Esses aspectos, somados aos demais apresentados anteriormente, contribuem ainda mais para o aumento dos pedidos de licença de uso do espectro para fins científicos.

Este trabalho visa descrever o processo de requisição da autorização do uso do espectro eletromagnético para fins científicos e experimentais. Após descrição de todo processo junto ao órgão regulador ANATEL, é apresentado um estudo de caso da rede 5G aberta da UFRN, no qual é descrito todo o *setup*, incluindo as especificações dos softwares e hardwares

utilizados, além dos resultados de radiação não ionizante (RNI) medidos quando o sistema está em operação.

O artigo está organizado como a seguir: Na Seção I-A, são apresentados trabalhos relacionados à pesquisa, destacando as suas principais contribuições. Na Seção II é descrito o procedimento de requisição do uso do espectro para fins científicos e experimentais, além de mostrar os limites de exposição à RNI da legislação. Na Seção III é apresentado um estudo de caso envolvendo a rede 5G aberta da UFRN. Na Seção IV são apresentados os resultados provenientes das medições de RNI realizadas no *setup* descrito na Seção III. Por fim, na Seção V, são discutidas as principais conclusões sobre o que foi apresentado em todo o trabalho, assim como são descritas as expectativas de trabalhos futuros.

A. Trabalhos Relacionados e Contribuições

É possível destacar a amplitude de abordagens ao falar sobre trabalhos relacionados à regulamentação do espectro eletromagnético encontrados na literatura.

Em [13], o foco do autor é destacar a importância do órgão regulador na gestão eficiente do uso do espectro eletromagnético no Brasil. Ele destaca a atuação da ANATEL em prol da competição saudável entre os diversos *players* do mercado de telecomunicações no Brasil, além de desempenhar um papel crucial na defesa do direito dos consumidores.

Outra abordagem encontrada em artigos científicos diz respeito à visão do uso do espectro eletromagnético como bem público [14], [15]. No trabalho desenvolvido em [14], é feita a análise de como se dá o processo de regulação do espectro no Brasil, além de comparar com processos mais antigos e bem estabelecidos, como, por exemplo, da União Europeia e dos Estados Unidos. O autor destaca a necessidade de maior transparência e flexibilidade no uso do espectro, de forma a beneficiar a sociedade, pois isso traria novas oportunidades de mercado e desenvolvimento tecnológico. Em [15], foi realizada uma análise crítica sobre as possibilidades jurídicas de flexibilização da regulação do espectro, assim como, foram descritos diversos pontos de melhoria no modelo atual. Está em destaque novamente a discussão sobre a flexibilização da regulamentação, possibilitando o aumento da utilidade e da função social do espectro.

O presente trabalho é fruto do uso prático da legislação brasileira no processo de solicitação de uso do espectro entre 3,7 e 3,8 GHz para uso científico em uma rede 5G de código aberto na UFRN. O trabalho contribui nos seguintes aspectos:

- Apresentar detalhadamente todo o procedimento de requisição, pontuando todas as etapas necessárias, incluindo pormenores relacionados a documentação, atores, responsabilidades e procedimento de assinaturas;
- Apresentar detalhes operacionais, frutos de trabalho consultivo com a equipe da ANATEL;
- Apresentar detalhamento de hardware e software para instalação e implantação de um sistema 5G virtualizado, operando com celulares comerciais;
- Demonstrar o procedimento de medição dos níveis de RNI em um *setup* de sete estações base 5G (gNBs) operando em 100 MHz de largura de banda.

Na literatura, esse é o primeiro trabalho que apresenta todos esses aspectos em conjunto.

II. USO DO ESPECTRO PARA FINS CIENTÍFICOS

Por definição, o serviço especial para fins científicos e experimentais é uma modalidade de serviço fixo ou móvel,

explorado em regime privado, de interesse restrito, para uso próprio, utilizada para efetuar experiências que possam contribuir para o progresso da ciência e da técnica em geral [16]. A norma que estabelece as condições para expedição de autorizações para execução desse serviço é a Norma Técnica NTC N° 22, aprovada pela Resolução n° 24, de 22 de setembro de 1966, do Conselho Nacional de Telecomunicações (CONTEL) e publicada no Diário Oficial da União – D.O.U. em 18 de outubro de 1966 [16].

Essa seção descreve os principais pontos acerca do serviço especial para fins científicos e experimentais, em especial o procedimento para requisitar esse tipo de serviço e a recomendação brasileira sobre o limite de radiação não-ionizante, referente a essa solicitação.

A. Legislação Brasileira

A legislação brasileira estabelece que a autorização se dá a título oneroso, devendo ser recolhido o Preço Público pelo Direito de Exploração de Serviços de Telecomunicações e pelo Direito de Exploração de Satélite (PPDESS) no valor de R\$ 20,00. Essa autorização é expedida por prazo determinado, mas uma vez findos os experimentos que motivaram sua expedição, a mesma torna-se sem efeito [16].

Se necessário o uso de radiofrequências, deve, também, ser solicitada Outorga de Autorização para Uso de Radiofrequências, associada à Autorização do referido Serviço. Cabe esclarecer que esta autorização é outorgada por prazo determinado, sendo que sua vigência deve coincidir com o prazo das experiências autorizadas [16].

A outorga obtida não pode ser usada para executar outros serviços de telecomunicações, que não com fins exclusivamente científicos ou experimentais. Para o início da operação das estações de telecomunicações, a autorizada deve obter Licença para Funcionamento de Estação junto à Agência Nacional de Telecomunicações (ANATEL), observada a regulamentação [16].

B. Procedimento de Requisição

No Brasil, o serviço especial para fins científicos ou experimentais deve ser implementado na faixa de frequências de 3,7 a 3,8 GHz. Essa faixa é compartilhada por todos os órgãos autorizados a operarem esse tipo de serviço. Assim, uma das coisas que deve constar na lista de documentos é um plano de coordenação de experimentos, para não ocorrer interferências entre as entidades autorizadas.

A permissão para a execução de serviço especial para fins científicos ou experimentais é outorgada pelo CONTEL, mediante requerimento instruído. As entidades que podem solicitar esse serviço são indústrias de telecomunicações, universidades, entidades brasileiras com fins científicos ou experimentais e concessionárias ou permissionárias de serviço de telecomunicações.

Antes de iniciar o processo de solicitação, o engenheiro que ficará como responsável técnico pelas instalações de rádio deverá cadastrar no sítio de serviços do CREA uma anotação de responsabilidade técnica (ART) com essa finalidade. O contratante e o proprietário devem ser a mesma entidade, mesmo que as instalações fiquem em um e o pagamento seja feito por outro. Por exemplo, mesmo que a rede fique em uma universidade e o pagamento do responsável técnico seja feito por uma fundação, o contratante e o proprietário na ART devem ser os mesmos.

Após isso, o responsável técnico deve receber procuração da entidade, indicando que o mesmo é o responsável técnico junto à ANATEL. Essa procuração deve ser cadastrada no Sistema Eletrônico de Informações (SEI). A entidade também deve ser cadastrada no SEI, bem como o responsável técnico, antes de qualquer procedimento. O procedimento para cadastro é apresentado em [17].

Para se cadastrar como usuário externo no SEI, deve-se preencher o formulário online de cadastro de usuário externo, bem como preencher o Termo de Concordância e Veracidade [18] e enviá-lo para o e-mail *sei@anatel.gov.br*. Após isso, o solicitante recebe um email informando que o cadastro como usuário externo no SEI está liberado. Após fazer o login na página de acesso externo do SEI, é possível assinar contratos, convênios, termos, acordos e outros instrumentos congêneres; fazer Petições Eletrônicas em processos novos ou já existentes; receber e responder Intimações Eletrônicas; emitir e gerir Procurações Eletrônicas; e visualizar processos com restrição de acesso aos interessados.

Caso a entidade não tenha a outorga de serviços de interesse restrito, deve-se solicitar o Serviços de Interesse Restrito - 002 (ou Coletivo - 001), notificando a ANATEL, por meio do sistema Mosaico [19]. Após a Outorga, deve-se protocolar a documentação complementar para dar prosseguimento ao cadastro do contrato, canalização e liberação de acesso [16]. Feito o cadastro do contrato, deve-se prosseguir com o cadastro das estações a serem licenciadas no sistema STEL [20].

Após isso, os interessados devem instruir o requerimento de serviços de telecomunicações com as seguintes informações:

- 1) Nome e endereço;
- 2) Programa das experiências a serem realizadas, incluindo objetivos em vista;
- 3) Tipo de operação; Prazo necessário para realização do serviço; e Classe de estação e natureza de serviço;
- 4) Locais da operação pretendida; Equipamentos a serem utilizados, incluindo nome do fabricante, modelo e número de unidades a serem utilizadas;
- 5) Frequências desejadas; e Características da antena;
- 6) Responsável pelas experiências;
- 7) Nome do Engenheiro ou Técnico de Telecomunicações, registrado no CONTEL, que se responsabilizará pelo funcionamento das instalações.

Após o preenchimento e submissão do requerimento de serviços de telecomunicações, os seguintes documentos devem ser anexados no SEI [16]:

- 1) Formulário de Solicitação de Serviço de Telecomunicações; e Formulário de Solicitação de Autocadastramento de Estações;
- 2) Descrição do Sistema com todas as informações;
- 3) Programa das experiências a serem realizadas, incluindo os objetivos pretendidos;
- 4) O período necessário para a realização dos experimentos (prazo máximo de 24 meses);
- 5) A área de abrangência da execução dos experimentos (município/UF);
- 6) Relação e descrição dos equipamentos a serem utilizados;
- 7) Dados técnicos necessários referentes à canalização da(s) frequência(s): TX/RX; Potência; Classe de Estação; Designação de Emissão; e Largura de Faixa;
- 8) Formulário Simplificado para Licenciamento (TRI, Laudo Conclusivo e Declaração de RNI);

- 9) Anotação de Responsabilidade Técnica (ART) relativa ao projeto;
- 10) Cópia autenticada do Ato Constitutivo da entidade requerente;
- 11) Documento de investidura do poder de assinar do requerente;
- 12) Declaração, assinada pelo representante legal da empresa de que está ciente da necessidade de efetuar coordenação com os demais usuários que fazem uso das mesmas radiofrequências.

C. Determinações sobre Limite de Exposição à Radiação Não Ionizante

Por meio da Lei Nº 11.934/2009 [21], da Resolução Nº 700/2018 [22] e do Ato Nº 458/2019 [23] da ANATEL são estabelecidos os limites e a metodologia de avaliação de exposição humana a campos elétricos, magnéticos e eletromagnéticos, advindos de estações transmissoras de radiocomunicação, de terminais de usuário e de sistemas de energia elétrica em frequências até 300 GHz, os quais se destacam como fontes artificiais de Radiação Não Ionizante (RNI). Os limites adotados nestes documentos foram propostos pela Comissão Internacional de Proteção Contra Radiação Não Ionizante (ICNIRP) em 1998 [24].

Em [23] são detalhados os limites de exposição ocupacional e da população em geral a tais campos, sendo estes últimos mais restritivos por se tratar de uma população que não sabe que está sendo exposta a RNI. A TABELA I lista os limites de exposição para a população em geral.

Tabela I: Limites para exposição da população em geral (valores eficazes não perturbados). Adaptado de [23].

Faixa de Frequência	Intensidade de Campo E (V/m)	Densidade de Potência da Onda Plana Equivalente S_{eq} (W/m ²)
1 MHz a 10 MHz	$87/f^{1/2}$	-
10 MHz to 400 MHz	28	2
400 MHz to 2000 MHz	$1.375f^{1/2}$	$f/200$
2 GHz to 300 GHz	61	10

O Ato 458 [23] também especifica diretrizes a serem seguidas durante as medições, como: 1) para frequência entre 100 kHz e 10 GHz, deve-se garantir que a média temporal atenda ao limite em qualquer período de seis minutos; 2) para realização da média temporal em um único ponto, a sonda deve estar localizada, preferencialmente na altura da máxima intensidade de campo, ou a 1,70 m de altura; 3) manter um distanciamento mínimo de 2 metros entre o equipamento e pessoas e objetos durante a medição; e 4) não utilizar dispositivos portáteis de telecomunicações inclusive telefones celulares próximos ao equipamento de medição durante a coleta dos dados.

Todas essas recomendações foram atendidas no estudo de caso apresentado a seguir.

III. ESTUDO DE CASO: REDE 5G ABERTA DA UFRN

A. Descrição do sistema e disposição das estações

As estações estão dispostas no Laboratório do Centro de Excelência Lenovo em Redes Avançadas (LANCE), situado em duas salas (403 e 404) do quarto andar de um prédio de quatro andares na Universidade Federal do Rio Grande do Norte (UFRN). A Fig. 1 mostra a disposição aproximada das

gNBs e do medidor de campo elétrico nas salas. É importante destacar que a posição das gNBs foi definida de tal modo a tentar distribuir a potência uniformemente por toda a sala, para que a cobertura do sistema proporcione experimentos de *handover* e multi-conectividade.

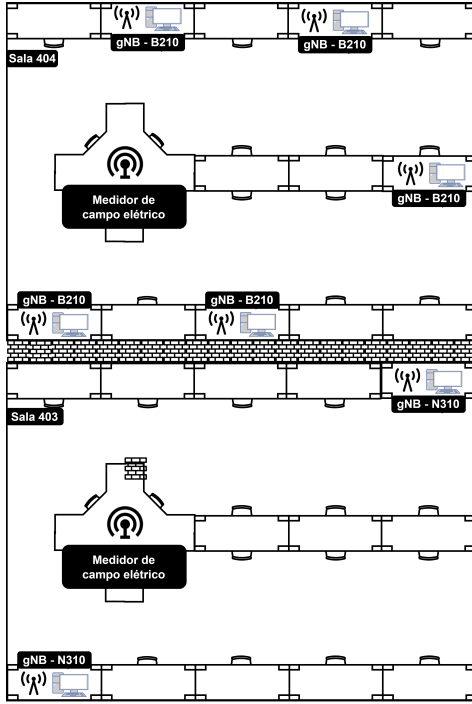


Figura 1: Disposição das gNBs e dos pontos de medição.

B. Configuração de hardware e software

As gNBs são compostas por dois equipamentos inter-conectados: o primeiro abriga a gNB (versão V2.0.0 da 5G RAN do *OpenAirInterface*¹) e o segundo é a USRP e suas antenas. As gNBs são instaladas e implantadas em um *desktop* ou em um servidor remoto, enquanto as USRPs disponíveis são cinco B210² e duas N310³ da Ettus [11]. O 5GC (versão V1.5.0 do *OpenAirInterface*) foi instalado em um dos desktops, se comunicando com as gNBs via rede *Ethernet* 10 Gbps. Dois *setups* foram montados para o experimento:

- 1) *Setup* com as cinco B210: desktop Lenovo, modelo ThinkCentre N50s Gen 3, 12th Gen Intel(R) Core(TM) i7-12700, 2.10 GHz, 64 GB RAM;
- 2) *Setup* com as duas N310: servidor Lenovo, modelo ThinkSystem SR630, Intel Xeon Gold 5218R, com 20 Cores 2.10GHz, 128 GB RAM. Versões do Docker e do Docker-compose são 24.0.7 e 2.21.0, respectivamente.

IV. PROCEDIMENTO E SETUP DE MEDIÇÃO

A. Medição de ocupação de espectro

Para medir a ocupação espectral e realizar os experimentos de medição de RNI, foi utilizado o *software* *IPERF 3* para realizar a transmissão de pacotes UDP entre um UE conectado a uma gNB a uma taxa de 1 Gbps durante 180 s, forçando as gNBs a atuarem em potência máxima. As Figs. 2 e 3 mostram

telas de monitoramento da ocupação do espectro do analisador de espectro portátil (HSA) N9342C da Keysight, para os experimentos com as USRPs B210 e N310, respectivamente. Utilizando a função *spectro monitor*, foi possível observar a ocupação de 40 MHz de banda (Fig. 2) e de 60 MHz de banda (Fig. 3), pois 106 PRBs foram configurados para as gNBs com as USRPs B210 e 162 PRBs para USRPs N310. O traço do espectro mostra que a frequência central é de 3,75 GHz, com um espalhamento de 100 MHz, ou seja, a faixa medida vai de 3,7 GHz a 3,8 GHz.

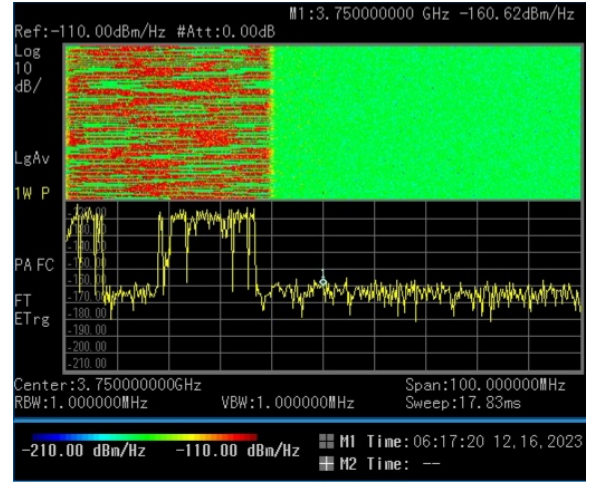


Figura 2: Ocupação espectral USRP B210: 106 PRBs (40 MHz).

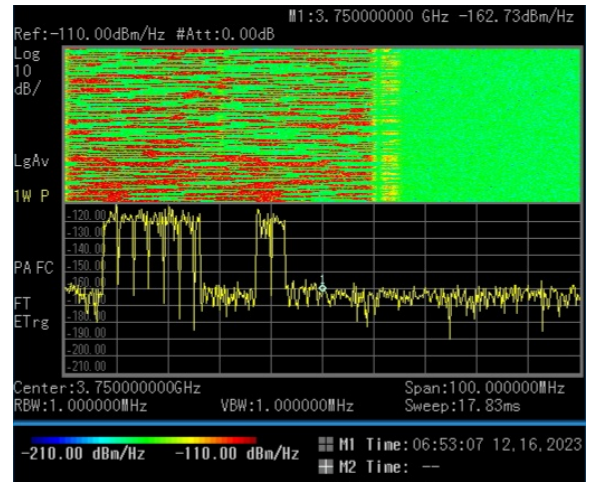


Figura 3: Ocupação espectral USRP N310: 162 PRBs (60 MHz).

A ocupação do espectro no momento das medições é ilustrado na Fig. 4.

B. Medição de RNI

Seguindo as determinações de [23], o equipamento de medição foi posicionado aproximadamente no centro do ambiente, a fim de respeitar os 2 m de distância de superfícies refletoras e de pessoas, bem como a região de campo distante das antenas. As medições consideram a média temporal de 6 minutos e o limite adotado para a faixa de frequência alvo é de 61 V/m, conforme mostrado na TABELA I.

Para a avaliação da RNI foi utilizado o monitor de campo eletromagnético contínuo MonitEM conectado a sonda de

¹<https://gitlab.eurecom.fr/oai/openairinterface5g/-/tree/v2.0.0>

²<https://www.ettus.com/all-products/ub210-kit/>

³<https://www.ettus.com/all-products/usrp-n310/>

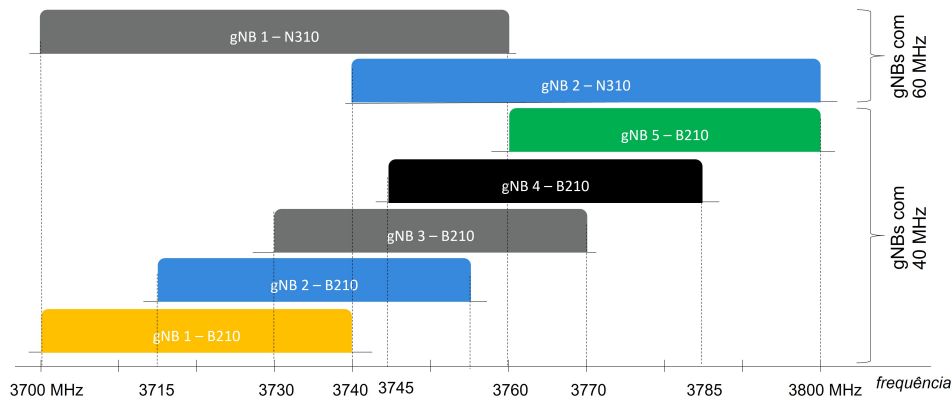


Figura 4: Ocupação espectral: 5 gNBs (40 MHz) e 2 gNBs (60 MHz).

campo elétrico WPF8, ambos da empresa Wavecontrol. Este arranjo é o mesmo usado pela ANATEL para medir campo elétrico (V/m) em banda larga, na faixa de frequência entre 100 kHz e 8 GHz. Por se tratar de um equipamento de medição em banda larga, durante as medições foram desligados todos os equipamentos da rede Wi-Fi do andar onde está localizado o laboratório.

C. Apresentação e Discussão de Resultados

A TABELA II mostra os valores de pico e de média dos campos elétricos e as taxas de exposição, definidas como a porcentagem obtida da relação entre o valor de pico medido e o limite de exposição que consta na legislação.

Tabela II: Valores de pico e de média dos campos elétricos e as taxas de exposição.

Ambiente	Pico do Campo Elétrico (V/m)	Média do Campo Elétrico (V/m)	Taxa de Exposição (%) ($f = 3,7 \text{ GHz}$)
Sala 403	0,45	0,059	0,737
Sala 404	0,24	0,1	0,39

Uma vez que os valores foram muito abaixo de 50% do limite legislado, não foi executada a média espacial em diferentes alturas, conforme exigiria [23].

Por se tratarem de medições em banda larga, os valores obtidos não se restringem apenas ao sistema 5G instalado na UFRN, abrangem toda a faixa de frequência do equipamento de medição, implicando que a contribuição de emissão eletromagnética das gNBs é baixa, considerando os valores limites estabelecidos pela legislação brasileira.

V. CONCLUSÕES

Este trabalho apresentou uma descrição detalhada sobre o processo de licenciamento de espectro para fins científicos e experimentais, mostrando o caso de uso real na UFRN. Foram descritos os passos para solicitação de autorização e todo processo junto à ANATEL, tratando da solicitação para uso em rede 5G aberta na própria UFRN. O *setup* foi descrito, incluindo as especificações necessárias tanto de softwares quanto de hardwares utilizados. Medições de RNI foram efetuadas, mostrando que os índices medidos estão dentro dos limites estabelecidos pela ANATEL.

REFERÊNCIAS

- [1] O-RAN Alliance, "About o-ran alliance," 2023. [Online]. Available: <https://www.o-ran.org/about>
- [2] R. Fontes, A. Martins *et al.*, "Open-source emulation-based test environment to settle O-RAN-compliant trials," *ITU Journal on Future and Evolving Technologies*, 2023.
- [3] I. Rego, L. Medeiros *et al.*, "Prototyping near-real time RIC O-RAN xApps for Flexible ML-based Spectrum Sensing," in *2022 IEEE NFV-SDN*, 2022, pp. 137–142.
- [4] —, "Frequency Scanning xApps: O-RAN RIC and GNU Radio with RTL-SDR Use Case," in *XL SBRT*, 2022.
- [5] RNP, "OpenRAN@Brasil," RNP, Feb 2024. [Online]. Available: <https://www.rnp.br/projetos/openranbrasil>
- [6] E. Rebouças and M. Martins, "Evolução da regulamentação da mídia eletrônica no brasil," *EDITORES*, 2007.
- [7] T. R. Coimbra, "Regulação do espectro de radiofrequências: uma análise técnica do modelo brasileiro," 2006.
- [8] J. A. Machado, "O modelo de expansão da banda larga no brasil: uma análise de políticas de infraestrutura e de gestão do espectro para a universalização do acesso à internet," *Revista Gestão & Políticas Públicas*, vol. 7, no. 2, pp. 199–219, 2017.
- [9] OAI, "OpenAirInterface Software Alliance," OSA, Feb 2024. [Online]. Available: <https://openairinterface.org/>
- [10] srsRAN, "srsRAN - Project Open Source RAN," srsRAN, Feb 2024. [Online]. Available: <https://www.srsrte.com/>
- [11] Ettus, (2025) Ettus website. [Online]. Available: <https://www.ettus.com/>
- [12] Per Vices, "Cyan Per Vices," Per Vices, Feb 2024. [Online]. Available: <https://www.pervices.com/cyan/>
- [13] G. d. S. Cardoso, "O papel da agência reguladora na gestão eficiente do espectro de radiofrequência no brasil," 2023.
- [14] R. d. P. Falcão, "Bens públicos e a regulamentação do espectro de radiofrequência: uma análise sobre o setor de telecomunicação," 2021.
- [15] G. P. Pinheiro, "A regulação do espectro de radiofrequências no brasil: uma visão crítica sob a perspectiva dos bens públicos," 2015.
- [16] ANATEL, "Serviço especial para fins científicos e experimentais," Abril 2021. [Online]. Available: <https://www.gov.br/anatel/pt-br/regulador/outorga/servico-especial-para-fins-cientificos-e-experimentais>
- [17] —, "Manual do usuário externo do sei." [Online]. Available: <https://docs.google.com/document/d/1tBRrH1E4s25Q2ZBe6sW0qp75HsNlVkwFqmeoVU8MQLo/pub>
- [18] —, "Termo de declaração de concordância e veracidade." [Online]. Available: <https://www.anatel.gov.br/Portal/verificaDocumentos/documento.asp?numeroPublicacao=336902>
- [19] —, "Serviços de interesse restrito," Junho 2022. [Online]. Available: <https://www.gov.br/anatel/pt-br/regulador/outorga/servico-de-interesse-restrito>
- [20] —, "STEL - SISTEMA DE SERVIÇOS DE TELECOMUNICAÇÕES." [Online]. Available: <https://sistemas.anatel.gov.br/stel>
- [21] Brasil, "Lei nº 11.934, de 05 de maio de 2009," *Diário Oficial [da] República Federativa do Brasil*, 2009. [Online]. Available: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/lei/11934.htm
- [22] ANATEL, "Resolução nº 700," 2018.
- [23] —, "Ato nº 458," 2019.
- [24] A. Ahlbom, U. Bergqvist *et al.*, "Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz)," *Health physics*, vol. 74, no. 4, pp. 494–521, 1998.

Otimização de Redes Sem Fio de Ondas Milimétricas com Algoritmo Morcego para Posição Dinâmica dos Usuários e Prioridade de Aplicações

Marcos Hiarley
FACOMP
UFPA

Castanhal, Brazil
marcoshiarley.silva@gmail.com

Robert Gabriel
FACOMP
Estácio

Bragança, Brazil
robertdsgabriel@gmail.com

Thalita Ayass
PPGEE
UFPA

Belém, Brazil
thalita_ayass@gmail.com

Evelin Cardoso
FACOMP
UFPA

Castanhal, Brazil
ehs.cardoso@gmail.com

Jasmine Araujo
PPGEE
UFPA

Belém, Brazil
jasmine.araujo@gmail.com

Renato Francês
PPGEE
UFPA

Belém, Brazil
rfrances@ufpa.br

José Jailton
FACOMP
UFPA
Castanhal, Brazil
jjj@ufpa.br

Abstract— With the advancement in the development of wireless network transmission technologies, there has been a significant investment in research in the field of mobile networks, especially in 5G and 6G technologies. This investment is driven by the exponential increase in the number of mobile network users who utilize a wide variety of services with this technology. However, one challenge in planning wireless networks is optimizing the allocation of user positions to maximize network efficiency. This paper will conduct wireless network simulations with millimeter waves, using the bat algorithm to determine new user positions based on their current positions and the type of application used. The goal is to improve coverage and efficiency of next-generation networks, especially in areas with difficult access.

Keywords— 5G Networks, 6G Networks, millimeter waves, user's position.

Resumo - Com o avanço no desenvolvimento de tecnologias de transmissão em redes sem fio, houve um investimento significativo em pesquisa na área de redes móveis, especialmente nas tecnologias 5G e 6G. Esse investimento é impulsionado pelo aumento exponencial do número de usuários de redes móveis que utilizam uma grande variedade de serviços com essa tecnologia. No entanto, um desafio no planejamento de redes sem fio é otimizar a alocação de posições dos usuários para maximizar a eficiência da rede. Este artigo realizará simulações de redes sem fio com ondas milimétricas, utilizando o algoritmo Morcego para determinar novas posições de usuários com base em suas posições atuais e no tipo de aplicação utilizada. O objetivo é melhorar a cobertura e a eficiência das redes de próxima geração, especialmente em áreas de difícil acesso.

I. INTRODUÇÃO

As redes 5G representam um avanço significativo em relação às gerações anteriores de tecnologia móvel, oferecendo velocidades de download e upload muito mais altas, latência reduzida e maior capacidade de conectar dispositivos

simultaneamente. Esse avanço tecnológico permitirá o desenvolvimento de aplicações como realidade aumentada e virtual, telemedicina e veículos autônomos. Além disso, as redes 5G são fundamentais para a expansão da Internet das Coisas (IoT), permitindo a conexão de bilhões de dispositivos, desde eletrodomésticos inteligentes até sensores industriais, promovendo uma transformação digital em várias indústrias. O desenvolvimento das redes 6G promete levar essas capacidades ainda mais longe. Espera-se que as redes 6G sejam implementadas a partir da próxima década, visando fornecer velocidades de transmissão de dados ainda mais altas, latências quase zero e maior eficiência energética. Além disso, as redes 6G estão projetadas para integrar tecnologias emergentes, como inteligência artificial, computação em nuvem distribuída e comunicações holográficas. Esses avanços possibilitarão novas possibilidades para serviços avançados, como comunicação em tempo real entre máquinas e a criação de ambientes virtuais hiper-realistas, proporcionando experiências completamente novas para os usuários e facilitando avanços tecnológicos em áreas como saúde, educação e entretenimento [1].

As ondas milimétricas desempenham um papel crucial na concretização dos benefícios das redes 5G e 6G. Essas ondas operam em frequências muito mais altas do que as redes móveis tradicionais, permitindo a transmissão de grandes quantidades de dados de forma rápida e eficiente. No entanto, as ondas milimétricas têm um alcance mais curto e são mais suscetíveis a interferências físicas, exigindo uma infraestrutura de rede mais densa com um número maior de antenas. Apesar desses desafios, o uso de ondas milimétricas é essencial para atender à crescente demanda por largura de banda e para suportar aplicações que exigem taxas de dados extremamente altas e baixa latência [1].

As redes 5G e 6G, juntamente com o uso de ondas milimétricas, têm o potencial de transformar regiões onde a conectividade é frequentemente limitada devido às grandes distâncias e desafios de infraestrutura. A implementação dessas tecnologias pode melhorar significativamente a comunicação em áreas remotas, facilitando o acesso à educação e à saúde, por exemplo, através da telemedicina, permitindo consultas médicas e diagnósticos remotos. Além disso, a conectividade aprimorada pode apoiar iniciativas de monitoramento ambiental, que são essenciais para a conservação florestal, fornecendo dados em tempo real sobre desmatamento e mudanças climáticas [2].

Este artigo visa simular redes sem fio utilizando ondas milimétricas com o objetivo de realocar usuários para alcançar um melhor desempenho da rede. A realocação dos usuários é feita utilizando o algoritmo Morcego, levando em conta as posições dos usuários e a prioridade do tipo de aplicação. Os benefícios da proposta foram analisados avaliando a qualidade do sinal e o atraso na transmissão, comparando as posições iniciais com as posições otimizadas dos usuários.

As seções deste artigo estão organizadas da seguinte forma: A Seção 2 discute os benefícios das redes 5G. A Seção 3 descreve a melhoria de assistência ao usuário. A Seção 4 aborda a metodologia utilizada no artigo. A Seção 5 apresenta os resultados obtidos das simulações. A Seção 6 conclui o artigo.

II. REDES 5G

Nos últimos anos, o desenvolvimento tecnológico promoveu uma verdadeira revolução na maneira como nos conectamos e compartilhamos informações. A quinta geração de redes móveis, conhecida como 5G, surge como protagonista dessa transformação, prometendo uma conectividade mais rápida, confiável e eficiente do que nunca. Uma das características mais notáveis do 5G é o uso de ondas milimétricas em altas frequências, abrindo caminho para um novo paradigma de comunicação sem fio.

O avanço da automação e da tecnologia no setor industrial também é notável. A adoção do 5G na indústria visa proporcionar melhorias substanciais na versatilidade, flexibilidade, eficiência e usabilidade das futuras fábricas inteligentes. Assim, a Indústria 4.0 incorpora tecnologias 5G, promovendo integração contínua em toda a cadeia de valor e em todas as camadas de automação.

As características do 5G têm gerado grandes expectativas, especialmente em comparação com o 4G, devido aos avanços significativos que essa tecnologia promete trazer. Melhorias notáveis incluem a velocidade da banda larga móvel e a baixa latência, essenciais para aplicações críticas como realidade virtual, inteligência artificial, veículos autônomos e robótica. Além disso, tecnologias como *network slicing* oferecem níveis de serviço personalizados, otimizando a infraestrutura para diferentes casos de uso e proporcionando uma experiência de usuário aprimorada [4].

Essa evolução tecnológica também pode desempenhar um papel crucial na preservação ambiental, especialmente em áreas como a Amazônia, onde o desmatamento é uma preocupação global. A implementação de redes 5G pode permitir um

monitoramento remoto mais eficaz das florestas, ajudando a detectar atividades de desmatamento ilegal e incêndios florestais com maior precisão e rapidez. Além disso, o uso de tecnologias como a Internet das Coisas (IoT) e a análise de dados em tempo real pode contribuir para a gestão sustentável dos recursos naturais, reduzindo o impacto humano nesses ecossistemas vitais.

Essas tecnologias e os avanços associados ao 5G representam uma evolução significativa no campo das comunicações sem fio, oferecendo oportunidades de transformação em vários setores, desde a indústria até a preservação ambiental, contribuindo para a conectividade, eficiência e inovação em escala global [3].

III. MELHORIA DE ASSISTÊNCIA AO USUÁRIO

Nas redes 5G, melhorar a conectividade para oferecer uma experiência superior ao usuário é essencial. O Aprimoramento Assistido pelo Usuário se destaca ao implementar tecnologias avançadas que não apenas melhoram a conexão, mas também a tornam mais robusta e adaptável às necessidades dinâmicas dos usuários.

Em um ambiente onde a demanda por dados e conectividade é cada vez mais crítica, tecnologias como Inteligência Artificial (IA) e Técnicas de Otimização desempenham um papel crucial. Essas tecnologias permitem a personalização avançada da conectividade, ajustando-se dinamicamente às necessidades individuais dos usuários e prevendo demandas de largura de banda para garantir conexões estáveis e de alta qualidade em todos os momentos [5].

Além disso, as redes 5G se beneficiam de avanços em antenas adaptativas e técnicas de formação de feixe, otimizando a eficiência espectral e melhorando significativamente a cobertura e a capacidade da rede. Isso resulta em uma conexão mais estável e de alta velocidade, especialmente em áreas urbanas densas e locais com alta demanda de tráfego.

A capacidade de compartilhamento dinâmico do espectro também é uma característica crucial das redes 5G, garantindo a alocação inteligente da largura de banda de acordo com as demandas variáveis, melhorando ainda mais a qualidade da conexão para todos os usuários.

Em resumo, o Aprimoramento Assistido pelo Usuário nas redes 5G não apenas proporciona maiores velocidades, mas também garante uma conectividade confiável que se adapta às necessidades específicas de cada usuário. Com essas tecnologias avançadas e práticas de gerenciamento eficientes, as redes 5G estão preparadas para transformar a maneira como nos conectamos e interagimos digitalmente, promovendo uma experiência de usuário significativamente aprimorada e abrindo caminho para novas inovações na era da conectividade móvel [5].

IV. METODOLOGIA

O Network Simulator (NS-3) é uma das principais ferramentas de código aberto para modelagem de cenários de rede, visando simular um ambiente que se aproxima das condições do mundo

real. O NS-3 permite o uso das linguagens Python e C++ para o desenvolvimento de projetos, sendo a última utilizada neste estudo. Para desenvolver um canal simulado de 5G, foi utilizado o módulo de ondas milimétricas MmWave, desenvolvido pela Universidade de Nova York. Este módulo permite a simulação de canais com frequências de até 100 GHz, ideal para coleta de dados e análise de redes 5G[6].

Nas simulações, o cenário desenvolvido possuía uma estação base e dez usuários posicionados aleatoriamente em uma área de 500m x 500m. Os usuários foram classificados em três categorias de prioridade dependendo do tipo de aplicação: Alta, Média e Baixa. A Estação Base calculará as novas posições dos usuários considerando a força do sinal e a prioridade de cada usuário. A Tabela I abaixo mostra os parâmetros de simulação utilizados no NS-3.

TABELA I. PARÂMETROS DE SIMULAÇÃO

Parâmetro	Valor
Cenário	Urbano
Frequência	100 GHz
Largura de Banda	200 MHz
Ambiente	LoS
Mobilidade	Aleatório
Antena	Omnidirecional
Usuários	10
Estação Base	1
Tempo de Simulação	100 s

A Estação Base calcula a nova posição dos usuários usando o Algoritmo Morcego. Esse algoritmo é uma técnica de otimização que utiliza uma população de "morcegos", onde cada morcego ajusta sua posição com base na frequência de emissão e na taxa de pulsos, enquanto o volume dos ecos recebidos indica a qualidade da solução. O processo envolve a atualização das posições dos morcegos em cada iteração. A função de aptidão, que avalia a qualidade de cada solução, é crucial para o algoritmo, pois é definida de acordo com os objetivos específicos do problema, como a alocação de usuários, incorporando variáveis como intensidade do sinal e prioridade de tráfego (Tabela II).

TABELA II. PARÂMETROS DO ALGORITMO MORCEGO

Parameter	Value
Número de usuários	10
População de Morcegos	30
Máximo de Interações	100
F_Min	0
F_Max	2
Fator de Separação	0,8
Prioridade de Aplicação	Vídeo: 1.0;Voz: 0.8;Dados: 0.5

V. RESULTADOS

A solução visa oferecer uma melhor qualidade de experiência do usuário, realocando o usuário para uma posição otimizada em relação à estação base, proporcionando uma melhor intensidade de sinal e, consequentemente, uma melhor conexão. A Figura 1 a seguir mostra o cenário simulado, onde é possível observar a estação base e os 10 usuários em suas posições iniciais, bem como nas novas posições de acordo com o Algoritmo Morcego. Nas novas posições, os usuários estão mais próximos da estação base, desfrutando de um serviço de conectividade aprimorado.

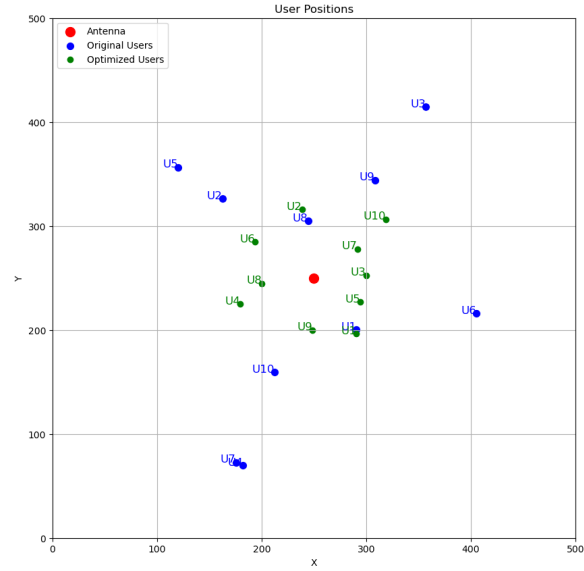


Fig. 1. User Positions

A Tabela III abaixo mostra as posições iniciais e finais de cada usuário, bem como o deslocamento que cada um teve que fazer. Após o deslocamento, todos os usuários estão mais próximos da antena, obtendo assim uma melhor experiência de conectividade. Entre os usuários de alta prioridade, o usuário 9 teve um deslocamento de 156,06m. Entre os usuários de prioridade média, o usuário 7 se moveu 235,88m. Entre os usuários de baixa prioridade, o usuário 6 teve um deslocamento de 222,11m.

TABELA III.USERS POSITION

User	Initial Distance	Final Distance	Priority	Movement
1	64.10	67.11	High	3.77
2	116.08	66.90	Low	76.82
3	196.32	50.00	Low	171.78
4	192.66	74.76	Low	155.45
5	167.92	50.00	Medium	217.11
6	158.82	65.97	Low	222.11
7	192.42	50.00	Medium	235.88
8	55.15	50.00	Medium	74.52
9	110.78	50.00	High	156.06
10	98.05	89.18	High	182.62

A Figura 2 ilustra a intensidade do sinal de cada usuário em

suas posições iniciais. Pode-se observar que apenas dois usuários (Usuários 9 e 10) têm uma intensidade de sinal acima de 30 dB, indicando uma qualidade de conexão excelente. Outros quatro usuários (Usuários 1, 2, 3 e 7) têm uma intensidade de sinal acima de 20 dB, representando uma boa qualidade de conexão. No entanto, quatro usuários (Usuários 4, 5, 6 e 8) têm uma intensidade de sinal abaixo de 20 dB, o que pode resultar em uma qualidade de conexão insatisfatória.

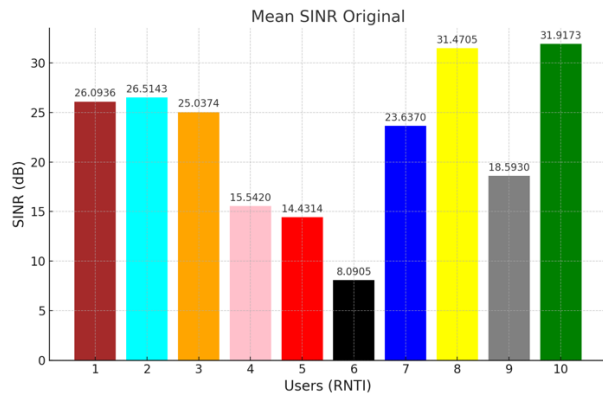


Fig. 2. Original Mean SINR

Após a otimização das posições dos usuários, é possível observar uma melhoria significativa na intensidade do sinal. Cinco usuários (Usuários 2, 3, 7, 8 e 10) agora têm uma intensidade de sinal acima de 30 dB, indicando uma qualidade de conexão excelente. Outros quatro usuários (Usuários 1, 4, 5 e 9) têm uma intensidade de sinal acima de 20 dB, representando uma boa qualidade de conexão. Apenas um usuário (Usuário 6) tem uma intensidade de sinal acima de 10 dB, mesmo assim com um ganho de 8 dB para 13 dB, apesar de ser de baixa prioridade. Essa redistribuição das posições dos usuários resultou em uma melhoria considerável na experiência de conectividade para a maioria dos usuários (Figura 3).

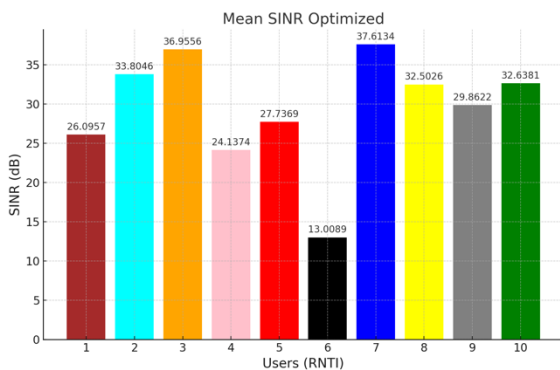


Fig. 3. Optimized Mean SINR

Ao analisar a intensidade média do sinal por prioridade de usuário, podem ser observadas melhorias significativas para os três grupos. O ganho percentual foi de 12,14% para o grupo de alta prioridade, 51,93% para o grupo de prioridade média e 54,01% para o grupo de baixa prioridade. Essas melhorias indicam que o algoritmo de otimização distribuiu os usuários

de maneira eficaz, proporcionando melhor qualidade de sinal e, conseqüentemente, uma experiência de conectividade superior para todos os usuários (Figura 4).

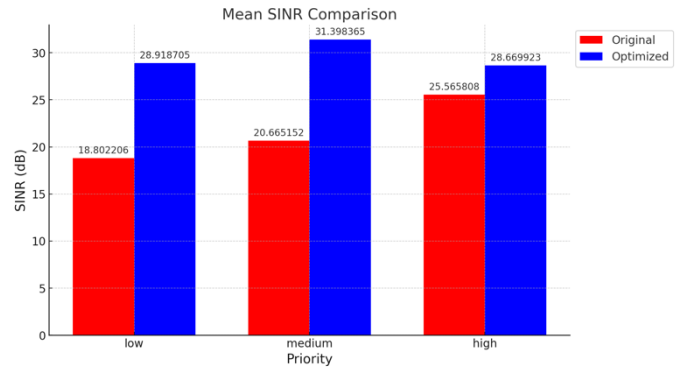


Fig. 4. Mean SINR by Priority

No gráfico abaixo, podemos ver os valores de atraso para cada usuário. É evidente que há uma variação considerável no atraso entre os diferentes usuários. Essa variação significativa nos valores de atraso pode impactar diretamente a qualidade da conexão, afetando a experiência do usuário. É importante notar que uma menor variação nos valores de atraso é desejável para garantir uma conexão mais estável e consistente (Figura 5).

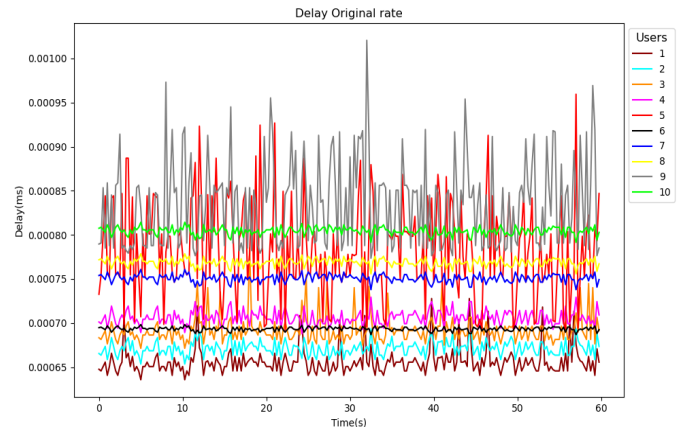


Fig. 5. Original Delay

Ao analisar os valores de atraso após otimizar as posições utilizando o Algoritmo Morcego, pode-se observar uma melhoria significativa nos valores, bem como sua estabilização. Antes da otimização, os valores de atraso apresentavam grande variabilidade, impactando negativamente a qualidade da conexão. Com a aplicação do algoritmo, houve não apenas uma redução geral nos valores de atraso, mas também maior consistência entre eles, resultando em uma experiência de conectividade mais uniforme e confiável para todos os usuários. Essas melhorias são cruciais para aplicações sensíveis ao atraso, como streaming de vídeo e videoconferência, onde manter alta qualidade de conexão é essencial (Figura 6).

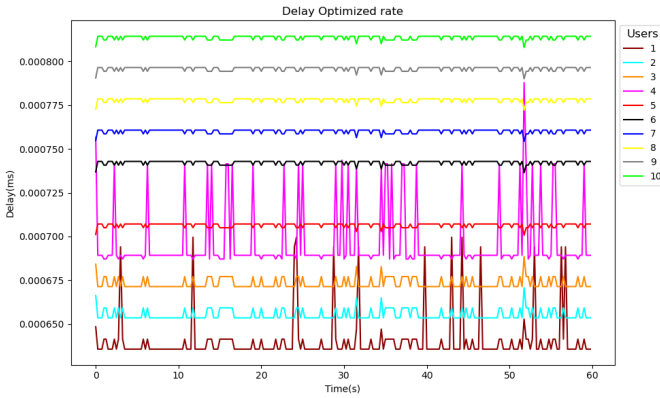


Fig. 6. Optimized Delay

Na Figura 7, é possível observar a distribuição dos valores de atraso por grupos, similar ao gráfico de relação sinal-ruído. Conclui-se que houve uma diminuição geral no atraso em todos os grupos, refletindo uma melhoria significativa na qualidade da conexão. No entanto, é importante destacar que no grupo de baixa prioridade houve um pequeno aumento no atraso. Esse aumento foi causado pela alocação de nós no Algoritmo Morcego. A proximidade entre os nós UE resultou em interferência no canal observado, o que consequentemente aumentou ligeiramente os valores de atraso. Apesar desse pequeno aumento, a melhoria geral nos valores de atraso demonstra a eficácia do algoritmo em otimizar as posições dos usuários, proporcionando uma experiência de conectividade mais estável e eficiente. Essa análise é essencial para entender o impacto das técnicas de otimização na qualidade do serviço, especialmente em cenários com alta densidade de usuários.

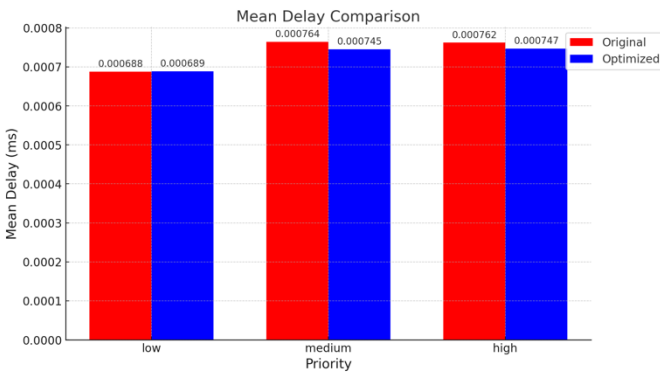


Fig. 7. Delay by Priority

VI. CONCLUSÃO

Este artigo apresentou uma abordagem inovadora para otimizar o posicionamento dos usuários em relação a uma estação base utilizando o Algoritmo Morcego. As análises demonstraram melhorias significativas na qualidade da conexão, refletidas no aumento da intensidade do sinal e na redução dos valores de atraso. A redistribuição eficaz dos usuários permitiu que a maioria dos grupos alcançasse uma qualidade de conexão superior, com ganhos substanciais observados em todos os níveis de prioridade.

A aplicação do Algoritmo Morcego mostrou-se eficiente não apenas no aumento da intensidade do sinal, mas também na estabilização dos valores de atraso, indicando uma melhoria geral na experiência de conectividade dos usuários. Apesar de uma ligeira interferência no grupo de baixa prioridade devido à proximidade dos nós, o desempenho geral foi amplamente positivo.

Os resultados sugerem que técnicas de otimização baseadas em algoritmos bio-inspirados, como o Algoritmo Morcego, podem ser extremamente eficazes na melhoria da qualidade do serviço em redes de comunicação densas. Esse avanço não só beneficia os usuários finais ao proporcionar conexões mais rápidas e estáveis, mas também oferece insights valiosos para o desenvolvimento futuro de redes mais eficientes e resilientes. Portanto, conclui-se que a adoção de técnicas de otimização avançadas pode desempenhar um papel crucial na evolução das redes de comunicação, tornando-as mais adaptáveis e capazes de atender às crescentes demandas da conectividade moderna. Estudos futuros poderiam explorar a integração de outros algoritmos bio-inspirados e técnicas de aprendizado de máquina para aprimorar ainda mais a alocação de recursos e a gestão da rede.

REFERÊNCIAS

- [1] K. G. Krishna, A. P. Singh and A. H. S., "Wireless Networking for 5G And 6G using Millimeter-Wave Technological Innovation," 2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC), Greater Noida, India, 2023, pp. 865-869, doi: 10.1109/PEEIC59336.2023.10450826.
- [2] N. Chukhno et al., "Models, Methods, and Solutions for Multicasting in 5G/6G mmWave and Sub-THz Systems," in IEEE Communications Surveys & Tutorials, vol. 26, no. 1, pp. 119-159, Firstquarter 2024, doi: 10.1109/COMST.2023.3319354.
- [3] Z. Li, O. Muta and Y. Jitsumatsu, "Performance Evaluation of Cooperative Positioning with User-Terminal Collaboration," 2023 11th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC), Alexandria, Egypt, 2023, pp. 240-243, doi: 10.1109/JAC-ECC61002.2023.10479642.
- [4] V. -H. Nguyen, V. Corlay, N. Gresset and C. Ciochina, "Probabilistic Ray-Tracing Aided Positioning at mmWave frequencies," 2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN), Nuremberg, Germany, 2023, pp. 1-6, doi: 10.1109/IPIN57070.2023.10332518.
- [5] X. Zhou, L. Chen and Y. Ruan, "Indoor Positioning With Multibeam CSI From a Single 5G Base Station," in IEEE Sensors Letters, vol. 8, no. 1, pp. 1-4, Jan. 2024, Art no. 6000204, doi: 10.1109/LSSENS.2023.3338252.
- [6] H. Poddar, T. Yoshimura, M. Pagin, T. S. Rappaport, A. Ishii and M. Zorzi, "ns-3 Implementation of Sub-Terahertz and Millimeter Wave Drop-based NYU Channel Model (NYUSIM)" in 2023 Workshop on ns-3 (WNS3 2023), Arlington, VA, USA, June 28–29, 2023.

Technical and Regulatory Aspects of Mobile Private Networks in Brazil

1st Filipe B. L. Pimentel *, 2nd Ingrid C. de Souza, 3rd Viviane H. da Silva,
4th Rafael Viegas C. Gomes,

Sidia Institute of Science and Technology
Manaus, Brazil

Emails: (filipe.pimentel, ingrid.carvalho, viviane.silva, rafael.gomes)@sidia.com*

Abstract—The non-public or private mobile networks is a topic that has gained relevance for mobile communications since 5G deployment due to the high number of use cases. A private network is a local area network that can provides services and advantages of a public mobile network for restricted users in a limited coverage area. In Brazil, private networks are being deployed mainly for industries, enterprises, agriculture, ports and airports. The paper presents the scenario for private networks in Brazil and also discuss about the different types of deployments.

Index Terms—Private Networks, NPN, 5G, 4G, Brazil

I. INTRODUCTION

The evolution of mobile networks from 4th to 5th generation has brought new and improved features, such as Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communication (URLLC), Massive Machine Type Communication (mMTC), Mobile Edge Computing (MEC) and others. These 5G capabilities have sparked the interest of companies and industries in deploying a private network with such characteristics. A private or non-public cellular network is deployed in a limited coverage area, usually an indoor environment such as an industry or factory, in which only authorized User Equipment (UE) can access it. A characteristic of private networks is the number of deployment options, which give customers flexibility. These networks may be deployed as completely standalone networks, using its own Physical Network Functions (PNF); hosted by an MNO sharing its RAN and CORE Infrastructure; may be offered as a slice of an MNO in a 5G SA context, or even use Virtual Network Functions (VNF) from an external provider [1]. Although a private network can be implemented using LTE, the number of non-public networks (NPNs) is expected to grow exponentially with 5G SA.

II. KEY CONCEPTS

For a long time, Wi-Fi has been widely adopted by businesses as the primary wireless network, offering speeds of up to 9.6 Gbps and efficiency over short-range distances. While it has been convenient for connecting devices and facilitating internal communication, Wi-Fi's limitations in terms of Quality of Service (QoS) and reliable coverage have become evident as business demands in terms of connected devices, latency, and throughput have increased.

This work was presented as part of the result of Project AMAN, executed by Sidia Institute of Science and Technology in partnership with Samsung Eletrônica da Amazônia LTDA, according to Informatics Law n.8387/91 and Art.39 of Decree 10.521/2020.

With the growing need for more robust connectivity, private networks have started to gain prominence globally. Initially, private LTE networks were preferred for their ability to offer greater quality of service, coverage, and reliability compared to Wi-Fi. These networks allowed companies more precise control over their communications infrastructure, ensuring better performance and security.

However, with the implementation of diverse technologies across different industry sectors, it has become essential to adopt more innovative configurations to support all required features. This is where 5G emerged as the solution, meeting the growing demands of modern industries.

A. UE Usage Settings for Private Networks

The first Radio Access Technologies (RAT), such as GSM and UMTS, were designed for voice usage, where network was used mainly to transport voice traffic between one necessary person subscriber and another party, another person or a voice service. Nowadays, newest RAT are used to carry data, and the parties involved in the communication are not necessarily persons anymore: Internet of Things is playing a huge part of the traffic generated nowadays.

'UE Usage' parameter in recent mobile networks define how a device prioritizes different types of network traffic. This settings is crucial for optimizing the performance and user experience in 4G and 5G networks, specially in Internet of Things (IoT) scenarios. The two options for UE Usage are Voice Centric and Data Centric, each serving different and distinct performance characteristics.

Voice Centric: In this mode, the device prioritizes voice traffic over data traffic. A connection to a Voice Core is mandatory for the UE on current RAT and PLMN since voice communication is critical for the device, such as traditional mobile phones. If a voice connection can't be established on this RAT, a legacy RAT must be chosen for registration. On newer RAT and PLMN, voice connection is provided by an IMS core.

Data Centric: On the other hand, this option is designed to prioritize data traffic over voice traffic. This setting is ideal for there scenario where data transmission is the primary use case of devices, such as tablets and IoT devices that frequently upload and download data from the network. A connection to a Data Core is mandatory for the UE on the current RAT and

PLMN. If a data connection can't be established on this RAT and PLMN, a legacy RAT must be chosen for registration.

B. Deployment Options

Following the standards defined in 3GPP release 16, a private network can be deployed by different models, considering characteristics such as level of security, specific services, network maintenance, and investment. The first option is a stand-alone NPN: it is a fully isolated network that does not need to use the infrastructure of a traditional MNO. In this case, both Radio Access and Core network are under the management of the company itself [2].

In this deployment model described on figure 1, the NPN operator can obtain a licensed spectrum from a regulatory agency to be used in the radio network or use the unlicensed frequencies. A stand-alone NPN brings benefits such as a high security level, customized QoS rules, own data network names (DNNs), ultra low latency due to RAN and Core inside the company, among others. On the other hand, this option requires a high initial investment (CAPEX).

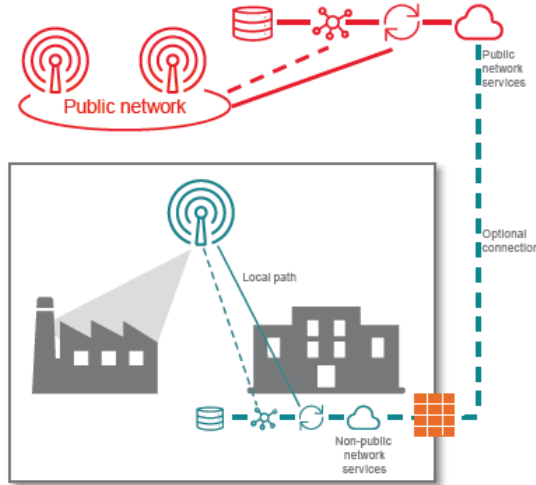


Fig. 1: Stand-alone NPN

The second option to deploy a private network is sharing parts of the RAN belonging with an MNO. In this scenario described in figure 2, only radio resources are shared between the private and public networks. The data traffic is delivered to each network separately, in which the data of the private network is forwarded to the private's Core user plane, while the public data traffic is forwarded to the MNO's core user plane. Security and privacy can still be guaranteed inside the enterprise. However, if an issue occurs on the radio element will affect the operation of the private network.

The third deployment option is sharing the RAN and the Core control plane part. In this scenario shown in figure 3, in addition to sharing radio resources, network control signaling is also carried out via the public network. In other words, functions such as access management, session management, and subscription repository are in the domain of the MNO. User plane functions and private network data traffic are still

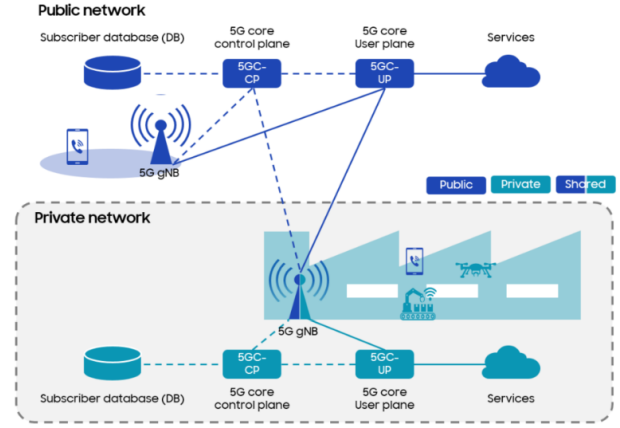


Fig. 2: Private Network with RAN sharing

kept separate from the public network using dedicated network slices, which will deliver all private data traffic to the private core user plane [3].

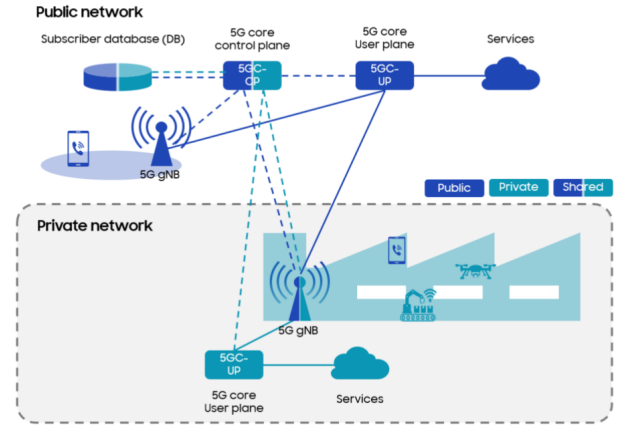


Fig. 3: Private Network with RAN and CP Sharing

The fourth deployment model is a private network fully hosted in a public PLMN. In this scenario, the radio resources as well the core functions are provided by an MNO. As a strategy to logically separate data traffic between public and private networks, end-to-end network slicing is used. This option offers lower level of security and privacy than previous options, since the service provider have full control of the network functions.

C. Legislation

Brazilian National Telecommunication Agency (ANATEL) has promulgated a series of Resolutions and Acts to regulate national telecommunications. Resolution No 617/2023 has regulated Private Networks under a SLP (Private limited Service*) license which describes the process to obtain the permission to explore the service and the radio frequencies designed for this type of network. The radio frequency bands are: 148 - 174 MHz; 225 - 270 MHz; 360 - 380 MHz; 380 - 400 MHz; 410 - 415 MHz / 420 - 425 MHz; 458 - 460 MHz / 468 - 470 MHz; 806 - 809 MHz / 851 - 854 MHz; 2.390 -

2.400 MHz; 2.485 – 2.495 MHz; 3.700 – 3.800 MHz; 27,5 – 27,9 GHz and the unlicensed frequencies [4]. According to Resolution 720/2020, prior authorization from ANATEL is needed for a company or individual to obtain an SLP license. Act No 915/2024 establishes new technical and operational requirements for the use of SLP licenses. This act repeals and replaces several previous standards, consolidating in a single document the technical requirements for the use of RF bands for these services. Main points discussed on this act are: 1. RF Power Limitations: For systems operating on 1890-1910MHz sub-bands. 2. Preference on Spectrum Utilization: Use of Low Bands (410-415MHz and 420-425MHz) for utilities services, such as water and power distribution. 3. Flexibility on Operation Mode: Possibility to use simplex or half-duplex on system previously provided only for duplex. 4. Consolidation of Governments standards, facilitating queries and consultations to documents.

D. Use cases

Private networks associated with the most suitable technology for service demand provide essential support for companies and industries, enabling them to increase their efficiency in production, planning, material control, and security. Additionally, they enable different applications in the medical field, reducing distances between doctors and patients in real-time, and can also offer solutions for Smart Cities, ensuring connectivity to different sensors and a large number of IoT devices for effective management of city areas.

These networks also facilitate a range of monitoring and management services for large areas, applicable to Agribusiness, Ports, and Airports, for the control of people and materials. For example, in agribusiness, they allow remote monitoring of crops, optimizing the use of resources such as water and fertilizers, while in ports and airports, they enable real-time tracking of cargoes and assets, improving logistical efficiency and security [5].

In the Brazilian scenario, in the year 2023, Brazilian Agency for Industrial Development (ABDI) and ANATEL awarded private network solutions in different sectors based on their innovative factor

The Industry 4.0 sector demands more connectivity from sensors and equipment and 5G is one of the main solutions.

The WEG S.A company through the project OpenLabWEG-V2COM developed an autonomous robot that can assist in the operational safety of workers in industrial plants, as it can enter and monitor risky locations and can also assist in teleinspection routines. It uses the 3.5 GHz (5G) band, which provides a quick response (low latency) to remote commands and also sends high-resolution images, with a high degree of detail, to Virtual Reality glasses used by operators [6].

Siderurgia do Futuro is a Gerdau project that operates in the 1,800 MHz (4G) and 3.5 GHz (5G) bands, with CAT-M technologies. One of its applications is raw material monitoring for automatic adjustments to the production and quality plan. And sensors to measure carbon monoxide levels and optimize truck loading.

The private network in the Agricultural sector works to increase efficiency in the operation of large areas and examples of its applications based on the management of collected and monitored data are: production efficiency, remote operation of machines, monitoring of possible fire outbreaks, occurrence of pests on plantations, livestock tracking and even weather forecasting.

Through partnerships with PUC-Rio the "Campo Conectado" is the project for a private network installed in Mato Grosso, operating in the 700 MHz band (4G). It combines different network technologies, such as NB-IoT, Wi-Fi, LoRa and satellites, to bring connectivity to Brazilian agriculture and collect and analyze agricultural data from farms.

Located in São Paulo, SemeAR-0 project from CPQD, uses 250 MHz bands (4G), Wi-Fi and LoraWAN. The project presents solutions for small producers and family farmers to monitor production, optimize management and connectivity to areas initially without connection, which made communication difficult even for product sales.

Considering at the telemedicine system, the OpenCare 5G project in partnership with the Hospital das Clínicas - USP/SP offers care for the underserved population, especially indigenous communities. The OpenRAN system operates in the 3,700 – 3,800 MHz(5G) band, offering quality medical consultations with low latency for real-time video service in remote locations.

SmartCity and Security as example, the award-winning solution was the Expansion Police Video (SSP-BA) installed in Salvador and the Metropolitan Region (BA) to assist public security forces. Operating in the 700 MHz band (4G). It is a private network of more than 1200 smart cameras and 160 mobile devices for use by police officers to assist in facial recognition, access, produce and share information in real-time audio, video and photo formats, among other features.

Another example is a solution from the City of Pato Branco/PR, where a private network of facial recognition cameras for public security and public Wi-Fi were installed in the 3,700 – 3,800 MHz (5G) bands.

To use private networks, it is essential to recognize the specific connectivity demands and the environments in which these networks are deployed. Rural areas, where coverage and range are essential to connect dispersed equipment, solutions with longer range frequency bands (4G), LoRaWAN or satellite networks are preferred. These technologies provide broader and more reliable coverage, ensuring connectivity in distant and difficult-to-reach areas.

In urban or industrial environments and high-density Internet of Things applications, where the demand is for connectivity for large numbers of devices, speed and reduced latency, 5G networks are more suitable.

This choice of technology is based on the specific needs of each context to ensure effective, high-performance connectivity

III. STATE-OF-THE-ART

The 3GPP, responsible for technical specifications for mobile networks, introduced Standalone NPNs (SNPNs) in Release 15 (R15), which operate independently of public networks,

aiming to make new services and markets viable. In Release 16 (R16), requirements for NPN were specified and industry-specific features were provided. Additionally, R16 specified improvements in identification, discovery, and access control capabilities for NPNs, such as Network ID (NID) and the closed access group (CAG) concept. NPN management was introduced in Release 17 (R17), 3GPP extended SNPN services to include authentication, authorization and accounting (AAA), user equipment onboarding for 5G connectivity, remote provisioning, and support for emergency calls and warnings of public security. Version 18 (R18) marks the beginning of 5G Advanced, bringing significant improvements in network intelligence, essential for robust industrial applications and aimed at meeting UE integration requirements.

The ITU, a United Nations agency, also sets critical standards for NPNs, focusing on interoperability, security and efficient spectrum management. ITU-T Recommendation E.212 deals with the international identification plan for public fixed and mobile networks. In July 2018, an amendment was introduced that addresses the allocation of a specific Mobile Country Code (MCC), 999, for internal use on private networks. This code is designed exclusively for use within private networks and is not routable between different networks [7].

Private network utilization is not new, but with the growth of 5G and the possibilities offered by the new generation, it has become a focus in many sectors. The technological standard imposed by the fifth generation has led many companies to invest, and it is predicted that there will be a growth of at least 400 percent in the number of implemented networks in the coming years. Private wireless networks began to be used primarily in public service and transportation sectors, even with 2G technology and Terrestrial Mobile Radio System. However, this type of network, along with Wi-Fi, no longer meets minimum requirements, particularly in terms of data capacity, QoS and security.

According to the Global Mobile Suppliers Association (GSA) report, the most used spectral bands worldwide are: 3.5 GHz to 3.7 GHz, CBRS (Citizens Broadband Radio Service) which is mainly used in the USA and 2.5 GHz.

The availability of dedicated spectrum for private networks is growing, with countries such as France, Germany, the United Kingdom and the United States leading this trend by allocating specific bands for this purpose. This allows organizations to acquire spectrum directly, without relying exclusively on mobile operators, facilitating the implementation of private networks and promoting the global adoption of these networks [8].

In addition to the technological and operational benefits, the allocation of dedicated spectrum also offers significant economic advantages. Companies in the production, transportation, healthcare and energy sectors can implement more secure and reliable communication solutions, resulting in increased productivity and reduced operational costs. In Brazil, the National Telecommunications Agency (Anatel) is responsible for regulating the spectrum bands used in private networks. The main spectrum bands specific to these services vary according to the specific use and application.

The most common spectral bands for private networks in Brazil include the 3.5 GHz Band (n77 and n78), essential for 5G services, the 2.3 GHz Band (n40), used mainly in private LTE networks, the 450 MHz, used in rural and critical telecommunications services, and in the 4.9 GHz Band, recently regulated for multimedia communications and public safety [9].

These bands are strategically chosen to balance coverage, capacity and service specificity. The advantages and properties of dedicated lanes reflect a trade-off between security, control, cost and operational complexity.

In relation to the management of Numbering Plans in private networks, ANATEL does not manage them, as there is no forecast for interconnection of these networks with public telecommunications networks. The National agency only suggests the ITU-T recommendation, facilitating future adjustments.

Today Brazil is among the top 15 countries with the highest number of private networks in the world, mainly using 4G networks, but with an increasing role in 5G. The country has been standing out on the world stage, mainly in the energy, production and agriculture segments [4].

IV. CONCLUSION

Non-public networks have gained notoriety and relevance in the telecommunications industry as a new vertical service to be offered to large customers such as industries, companies, agribusiness, among others. Due to technical aspects such as availability, reliability, low latency, security and deployment flexibility, the number of private 5G networks is expected to grow significantly in the coming years. Furthermore, the use of unlicensed spectrum together with the non-dependence on an MNO are factors in accelerating the adoption of private networks.

REFERENCES

- [1] 3rd Generation Partnership Project, "5g; service requirements for the 5g system (5gs);," *ETSI TS 122 261*, vol. 16, 2021.
- [2] —, "5g; system architecture for the 5g system (5gs); physical layer procedures," *ETSI TS 123 501*, vol. 16, 2020.
- [3] Samsung, "Private networks vol.2 architectures and features for industrial scenarios white paper," pp. 3–6, 2021.
- [4] ANATEL, "Redes privadas," url <https://www.gov.br/anatel/pt-br/regulado/radiofrequencia/redes-privativas>, Sep. 2023.
- [5] Samsung, "Private networks vol.1 transforming private networks with samsung 5g," pp. 2–3, 2021.
- [6] W. V2COM, "Open lab 5g weg-v2com," url <https://v2com.com/open-lab-weg-v2com-2/>.
- [7] ITU, "E.212 : The international identification plan for public networks and subscriptions," *Recommendation E.212*, 2018.
- [8] Ericsson, "5g spectrum for local industrial networks," 2024.
- [9] ANATEL, "Anatel destina 120 mhz para o 5g na faixa de 4,9 ghz," url <https://www.gov.br/anatel/pt-br/assuntos/noticias/anatel-destina-120-mhz-para-o-5g-na-faixa-de-4-9-ghz>, Oct. 2023.

Avaliação Tecno-econômica para ampliação de redes óticas nas bandas C, L e S

1st Sandro Ribeiro

PPGES

Universidade de Pernambuco

Recife, Brasil

sr@poli.br

2nd Dr. Carmelo Jose Albanez Bastos Filho

PPGES

Universidade de Pernambuco

Recife, Brasil

carmelofilho@poli.br

3rd Jorge Candeias do Nascimento

CIN

Universidade Federal de Pernambuco

Recife, Brasil

<https://orcid.org/0000-0002-8461-2076>

Resumo—Este estudo propõe uma metodologia para a avaliação tecno-econômica da expansão de redes óticas Multi-banda (MB), com foco na viabilidade de ampliação das redes que operam na banda C. Utilizamos dispositivos e componentes como Amplificadores Ópticos (AO), multiplexadores, demultiplexadores, wavelength selective switch (WSS), transceptores e transponders (TR) de alta capacidade, que suportam interfaces avançadas como 100G, 200G, 400G e 800G. As bandas óticas consideradas incluem a banda C (1530-1565 nm), banda L (1565-1625 nm) e banda S (1460-1530 nm). A infraestrutura da rede ótica utiliza fibras óticas (FO) do tipo single mode fiber (SMF) padrão ITU-T G652, amplamente empregadas devido à sua baixa atenuação (≈ 0.4 dB/Km). Apresentamos um processo de validação utilizando a ferramenta de simulação GNPY (Gaussian Noise simulator in Python), que fornece o GSNR (Generalized Signal Noise rate) um estimador de qualidade de transmissão (QoT-E). Validamos as interfaces e selecionamos os cenários para discussão do melhor custo-benefício. Consideramos as propostas de ampliação, expandindo a rede em até 10 vezes sua capacidade atual, de aproximadamente 4Tbps para aproximadamente 40Tbps. A avaliação tecno-econômica leva em conta não apenas o custo dos dispositivos, componentes e da infraestrutura necessária, mas também a Qualidade de Transmissão (QoT), fornecendo uma visão detalhada da viabilidade de ampliação das redes existentes com baixo Capital Expenditure (CapEx).

Palavras chave—Multibanda, AO, WSS, TR, SMF, QoT-E, GNPY, GSNR, QoT, CapEx

I. INTRODUÇÃO

De acordo com as estimativas da União Internacional de Telecomunicações (ITU-T), a banda larga fixa representa 83% de todo o tráfego de dados atuais [1]. Em uma projeção da demanda de tráfego de dados de 2023 até 2033 obtida em [2], foi demonstrada que a demanda estimada por dados até 2033 no Brasil pode potencialmente atingir de 315 a 396 exabytes por ano. Essa demanda estimada representa um crescimento de mais de 4 vezes acima da demanda registrada em 2023. Os cenários previstos preveem um crescimento anual da demanda por dados entre 13% e 18% de 2024 a 2033. O atraso na adoção de tecnologia de fibra moderna por grandes operadores, tem provocado gargalos de infraestrutura em determinadas regiões [3]. O elevado custo de ampliação da infraestrutura de redes óticas, torna inviável para as operadoras alugar ou implantar novas infraestruturas óticas. O maior desafio é atender à crescente demanda de dados, maximizando a capacidade das redes em operação, que atualmente estão no

limite da capacidade da fibra ótica. As operadoras buscam definir a melhor solução para ampliar sua capacidade e atender a crescente demanda de tráfego. Alguns questionamentos sobre o limite dos transponders de quinta geração, que estão se aproximando do limite de Shannon, pode ser observado nos trabalhos e pesquisas de ampliação das redes óticas existentes, que estão operando com capacidades de transmissão na faixa de 100Gbps por canal, e cujas redes estão sendo forçadas a evoluir para dispositivos com taxas entre 400G e 800G por canal [4].

A transmissão ótica MB tem se mostrado mais atraente, para investimentos em ampliação na capacidade das redes em operação, pois oferece a possibilidade de se utilizar todo o espectro de baixa perda, que variam de 1250nm a 1675nm, referente as faixas denominadas O, E, S, C, L e U, e com atenuações inferiores a 0.4 dB/Km [5]. Nos curto e médio prazos, as atualizações de rede podem explorar bandas além das bandas C e L, enquanto a longo prazo, diferentes soluções poderiam ser adotadas considerando em muitos caso de uma decisão econômica [6] e [7]. Shariati et al. [8] mostraram que a exploração das bandas C + L + S melhoram a capacidade da rede existente de 8% a 14%, se comparadas com a implantação de sistemas de transmissão Multifibra (MF) com três fibras de banda C convencional, considerando apenas a eficiência espectral dos cenários, sem incluir os custos relativos.

Neste artigo, apresentamos uma análise técnico-econômica na qual vários cenários de redes nas bandas C, L e S (com a banda C convencional), são avaliadas e comparadas com relação ao custo-benefício da expansão das redes, considerando os impactos financeiros dos dispositivos e componentes óticos diferentes da banda C convencional, no custo de implementação dos cenários propostos. A capacidade e qualidade das redes são investigadas, e discutimos as possíveis soluções de atualização da capacidade baseadas em sistemas de amplificação multibanda. Os sistemas são propostos com interfaces para tráfego de altas taxas de transmissão, desenvolvidas com tecnologia de transmissão coerente e modernas técnicas de modulação, que proporcionam taxas de transmissão de dados de até 800G. Foram comparadas várias abordagens de implantação de soluções, utilizando a mesma fibra e com fibras em paralelo, avaliando o desempenho e custo de implementação de cada cenário explorado. Os

resultados revelam que o custo obtido por hardware e por bit, estão diretamente relacionados a maturidade e disponibilidade dos dispositivos e componentes utilizados. As interfaces que operam na banda convencional C, possuem um custo relativamente baixo se comparadas as interfaces em outras bandas, e apresentam um custo relativo cerca de 20% e 80% maior nos amplificadores e transponders operando nas bandas L e S, respectivamente. Avaliamos a eficiência espectral e os custos relativos, da expansão de uma rede operando em sua capacidade máxima, em torno de 4.8Tbps para uma rede com capacidade de 40Tbps ou mais, ou seja, ampliando esta rede em 10 vezes sua capacidade transmissão atual.

O resto do artigo está organizado da seguinte forma. Na seção II, apresentamos uma rápida abordagem dos componentes utilizados nos experimentos, as ferramentas e técnicas facilitadoras para o estudo. Na seção III, é apresentada a proposta desta pesquisa. Na seção IV, a metodologia aplicada para obtenção dos resultados. Na seção V, discutimos a análise e validação dos resultados. Concluímos o estudo com as considerações sobre as respostas obtidas, para validação das soluções que apresentaram a melhor opção tecno-econômica, para expansão da capacidade em termos de custo e desempenho.

II. FUNDAMENTAÇÃO TEÓRICA

A. Dispositivos e Componentes

Neste estudo consideramos três componentes principais da rede, para a avaliação de custo dos cenários propostos e utilização nas simulações da camada óptica, que são os transponders, os WSS e os amplificadores ópticos.

1) *Transponders*: Neste estudo serão utilizados dispositivos baseados na comunicação óptica coerente, que consiste no uso de modulação coerente na extremidade transmissora e do uso de tecnologia coerente na extremidade receptora, o que proporciona um aumento na distância de alcance destes dispositivos a nível de milhares de quilômetros [9].

2) *Amplificadores*: Os amplificadores EDFA apresentam ganho mais eficazes para as bandas C e L, e os TDFA, são amplificadores ópticos utilizados principalmente para a banda S do espectro óptico, pois quando excitado por uma fonte de bombeamento tem picos de emissão no espectro que coincide com a banda S (aproximadamente 1460-1530nm). São os componentes com maior maturidade e disponibilidade comercial, e que utilizaremos neste estudo.

3) *WSS*: Estudos recentes indicam que atualmente mais de 90% dos nós ROADM (Reconfigurable Optical Add-Drop Multiplexer) utilizam tecnologia Wavelength Selective Switch (WSS) [10]. O WSS são elementos que possuem interfaces de controle e gerência, que permitem a atenuação de forma independente de cada comprimento de onda, permitindo a equalização dos canais de forma independente, evitando a geração dos níveis de interferência que afetariam a qualidade da transmissão do sistema, devido a diferença de potência entre os canais transmitidos.

B. Qualidade de Transmissão-QoT

Em enlaces ópticos a estimativa da qualidade da transmissão (QoT) é considerada como uma etapa fundamental de controle para a configuração do caminho da luz, a fim de verificar a viabilidade adequada da capacidade de transmissão ao longo de um caminho de luz [11]. Baseado em [12], pode-se afirmar que as tecnologias baseadas em transmissão coerente, estão cada vez mais disponíveis para ampliação da capacidade das redes existentes. E para lidar com as redes baseadas nesta tecnologia, é imprescindível uma ferramenta que forneça um Estimador de Qualidade de Transmissão (QoT-E) preciso. O Telecom Infra Project (TIP) pelo grupo de trabalho Physical Simulation Environment (PSE) [13] e [14], desenvolveu uma ferramenta chamada Gaussian Noise simulation in Python (GNPy) [15], melhor detalhado na próxima seção.

C. Simulador GNPy

A ferramenta GNPy é um simulador de redes ópticas onde, as operações da estrutura de simulação são baseadas na capacidade de estimar o QoT de um ou mais canais operando caminhos ópticos em uma determinada rota de rede. O GNPy é uma biblioteca de código aberto desenvolvida pela comunidade, para a construção de ferramentas de planejamento e otimização de rotas em redes ópticas mesh do mundo real, e baseada no modelo de ruído gaussiano (GN). Em essência, esta ferramenta permite que um usuário determine a viabilidade de modulação e alocação de capacidade em uma rede multiplexada por divisão de comprimento de onda (WDM), totalmente coerente com software independente do fornecedor [16]. No GNPy a propagação específica de um sinal e suas deficiências são calculadas, e um caminho linear através da rede é usado, e a ferramenta fornece um conjunto de recursos estruturados em torno de um mecanismo central que cuida dos efeitos de propagação e da estimativa de QoT.

D. Relação $BER \times SNR$

A Qualidade da Transmissão (QoT) é um indicador crítico que avalia a capacidade de transporte das redes ópticas, permitindo analisar se a configuração em estudo atende às condições de capacidade projetadas. Em [17], são apresentados modelos para avaliar a QoT de uma rede, que permitem estimar analiticamente os níveis de Signal-to-Noise Ratio (SNR) e Bit Error Rate (BER). Os valores de BER obtidos são utilizados para avaliar a qualidade do sinal recebido, enquanto o SNR permite estimar se os parâmetros projetados garantem a qualidade da rede em estudo. Em nosso estudo adaptamos a Equação 1 obtida em [18], que relaciona SNR e BER para diferentes formatos de modulação.

$$BER = A \operatorname{erfc}[\sqrt{B * SNR}] \quad (1)$$

O SNR (Signal-to-Noise Ratio) é uma métrica fundamental para a avaliação da qualidade do sinal em sistemas ópticos. Na Equação 1 os valores de A e B é que determinam qual formato de modulação está sendo escolhido. A Tabela I é um exemplo de valores de SNR requerido, conforme cada formato

de modulação, e pode ser obtido usando a Equação 1. Essa configuração permite uma pre FEC (Forward Error Correction) de BER mínima de 0.004.

TABELA I
VALORES DE SNR OBTIDOS P/ EQ. 1

Formato de Modulação	SNR Requerida
BPSK	5.46
QPSK	8.47
8QAM	12.45
16QAM	15.13
32QAM	18.12
64QAM	21.06

III. PROPOSTA DO ESTUDO

A proposta deste estudo é explorar cenários que utilizem redes ópticas MB, empregando soluções que abranjam o crescimento das redes, aproveitando a infraestrutura em operação na banda C, expandindo a capacidade da rede com outras bandas do espectro óptico como as bandas L e S. Avaliar a eficiência espectral e a capacidade de ampliação das redes, com a análise da topologia selecionada, utilizando interfaces desenvolvidas com tecnologia de transmissão coerente, e modernas técnicas de modulação que permitem altas taxas de transmissão no canal óptico de longa distância. Avaliar os dispositivos e componentes conforme suas características, validando a viabilidade e aplicação na elaboração das soluções. Analisar e comparar os cenários com relação ao seu desempenho e custo relativo, e com os resultados obtidos possibilitar uma análise tecno-econômica confiável. Uma melhor compreensão da nossa proposta, criamos o fluxo do processo apresentado na figura 2, mostrando o passo a passo das etapas para atingirmos os objetivos do estudo, que trata da expansão de redes ópticas operando na banda C convencional, em até 10 vezes a capacidade atual. Esta proposta também pode ser aplicada para expansão da capacidade das redes conforme o cenário proposto.

No primeiro bloco: definimos a topologia de rede na camada física estudada; a ferramenta de simulação; os dispositivos e componentes utilizados no estudo; a configuração dos parâmetros da camada física; executamos as iterações para validação dos dispositivos e componentes; obtenção dos resultados; validação dos componentes e dispositivos viáveis para o estudo. No segundo bloco: a seleção dos elementos viáveis; elaboração de propostas dos cenários; cálculo dos custos de cada solução; comparação e validação dos cenários. A proposta apresentada visa obter de forma rápida e confiável, um método para avaliação de soluções que ofereçam alta capacidade de transmissão, com eficiência, qualidade e baixo custo de implementação nas bandas C, L e S, atendendo os objetivos de suprir e solucionar o problema do iminente estágio de congestionamento das redes, em virtude da crescente demanda por tráfego de dados nas operadoras de telecom.

IV. METODOLOGIA

A metodologia aplicada para obtenção dos resultados deste estudo, foi baseado no fluxo proposto na Figura 2. Na primeira

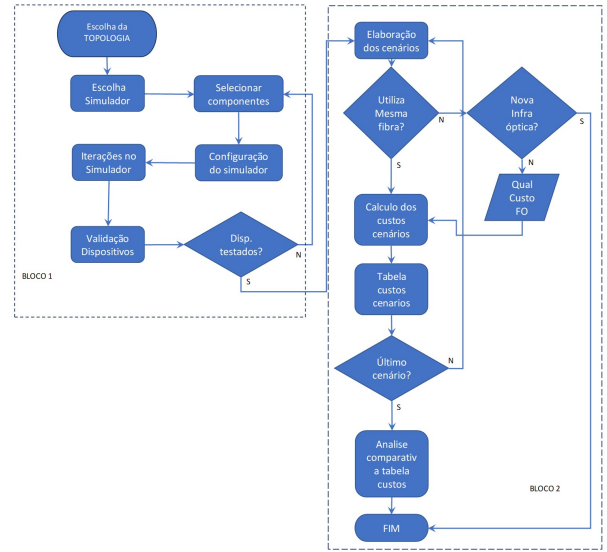


Fig. 1. Fluxograma da proposta do estudo.

etapa definimos a topologia a ser estudada, e montamos um setup experimental utilizando uma topologia de rede existente, com a camada física composta por um enlace ponto-a-ponto (origem-destino), por 2 nós ROADM interligados por uma infraestrutura de fibra óptica com cabos ópticos tipo Single Mode Fiber (SMF) padrão ITU-G652, contendo no caminho óptico 12 spans de fibra SMF, e 11 estações para regeneração e/ou amplificação do sinal, contendo 11 amplificadores tipo EDFA para banda C convencional implantados. Neste estudo consideramos apenas três elementos de rede para a avaliação da viabilidade técnica na camada óptica, que são os transponders e amplificadores ópticos, pois conforme os parâmetros especificados, quer seja de Ganho e Potência de transmissão, impactam diretamente nos resultados das simulações, através das penalidades causadas devido a variação dos parâmetros de entrada no simulador na camada física.

TABELA II
PARAMETROS TRANSPONDERS COMERCIAIS LONGA DISTÂNCIA

TR	Tipo TR	Formato Mod.	max. width=0.8 Taxa Símbolo	BW GHz	Taxa Dados	GSNR Req.
TR1	100G	PM-QPSK	32Gbaud	37.5GHz	100Gb/s	9.8dB
TR2	200G	PM-16QAM	32Gbaud	37.5GHz	200Gb/s	16.55dB
TR3	400G	PM-16QAM	63Gbaud	75GHz	400Gb/s	16.55dB
TR4	400 G	PM-64QAM	42Gbaud	50GHz	400Gb/s	22.5dB
TR5	800G	PM-16QAM	128Gbaud	150GHz	800Gb/s	16.55dB
TR6	800G	PM-32QAM	96Gbaud	112.5GHz	800Gb/s	19.5dB
TR7	800G	PM-64QAM	80Gbaud	100GHz	800Gb/s	22.5dB

Na escolha do simulador, selecionamos a ferramenta de simulação GNPY, descrita como uma biblioteca de código aberto com a capacidade de fornecer um estimador de QoT, que retorna o valor de GSNR para cada canal ao longo do caminho óptico. O GSNR foi provado ser uma métrica eficaz para estimar o QoT de uma transmissão óptica WDM não compensada moderna, coerente e modulada em vários níveis

de modulação [19] e [20].

Selecionamos os transponders listados na Tabela II. Montamos o setup experimental na ferramenta GNPY, e configuramos os parametros relacionados aos transponders selecionados. Simulamos a topologia experimental para cada transponder selecionado, para obtenção dos valores de GSNR de cada componente avaliado nas iterações realizadas. Os valores de GSNR obtidos foram utilizados para a validação e viabilidade dos componentes na topologia experimental, conforme especificações do GSNR requerido para garantir o QoT e performance de cada componente simulado. Apenas os transponders TR1, TR2 e TR3, foram classificados como validos e viáveis, conforme dados obtidos ns simulações realizadas no GNPY. Os TR1, TR2 e TR3 foram selecionados para elaboração das propostas de cenários de ampliação, gerados conforme a capacidade de transmissão de cada transponder para montagem das soluções. As propostas de cenários de ampliação da topologia experimental até 40Tbps foram elaboradas, conforme mostrado na Tabela III.

TABELA III
CENÁRIOS PROPOSTOS

Cenários	Banda	FO	Tipo TR	Capacidade
A	C	0	TR1	40 Tbps
	L	0	TR2	
B	C	0	TR2	40 Tbps
	L	0	TR1	
C	C	0	TR2	40 Tbps
	L	0	TR2	
D	C	0	TR3	40 Tbps
	L	0	TR1	
E	C	0	TR3	40 Tbps
	L	0	TR2	
F	C	0	TR2	40 Tbps
	S1	0	TR1	
G	C	0	TR2	40 Tbps
	S2	0	TR1	
H	C	0	TR3	40 Tbps
	S1	0	TR1	
I	C	0	TR3	40 Tbps
	S2	0	TR1	
J^1	C	0	TR2	40 Tbps
	C	1	TR2	
K^1	C	0	TR3	40 Tbps
	C	1	TR3	
L^1	C	0	TR3	40 Tbps
	L	1	TR1	
M^1	C	0	TR3	40 Tbps
	L	1	TR2	
N^1	C	0	TR3	40 Tbps
	S1	1	TR1	
O^1	C	0	TR3	40 Tbps
	S2	1	TR1	
P^1	C	0	TR2	40 Tbps
	L	1	TR1	

As propostas para utilização da banda S, foram elaboradas considerando as sub-bandas S1 e S2, devido as características dos amplificadores que não apresentam um ganho plano em toda a banda S. Os cenários com o índice “1” são as soluções que utilizam uma nova FO.

A Tabela IV obtida em [21] foi utilizada como referência para análise e obtenção dos custos relativos relacionados aos cenários propostos. Utilizamos o modelo de custo adaptado do modelo elaborado por Huelsermann et al. [22], para avaliar

TABELA IV
CUSTO RELATIVO COMPONENTES

Dispositivo	Custo Relativo
EDFA (C-BAND)	1
EDFA (L-BAND)	1.2
EDFA (S-BAND)	α
MUX DEMUX	0.04
1X9 WSS (C-BAND)	5
1X9 WSS (L-BAND)	6
1X9 WSS (S-BAND)	5α
1X20 WSS (C-BAND)	7.5
9X18 WSS (C-BAND)	20
9X18 WSS (L-BAND)	24
9X18 WSS (S-BAND)	20α
18X9 WSS (C-BAND)	20
TR (C-BAND)	36
TR (C-BAND)	43
TR (C-BAND)	36α
FO(fiber/Km)	0.5

os custos por dispositivos e componentes implementados em cada cenário proposto, ou seja, o somatorio do custo da fibra, dos amplificadores ópticos, dos dispositivos mux/demux, transponders e dos dispositivos OXCs. Assumimos que os valores de α e da FO foram adotados como 1.8 e 0.50, respectivamente, tomando como referência o valor relativo definido para o EDFA. O valor de alfa utilizado neste estudo é 80% maior que o custo do EDFA e o da FO 50% do custo do EDFA. O custo do EDFA no mercado depende do modelo e fabricante, bem como, das condições de projeto e contratos entre os fornecedores e as operadoras, por exemplo, este custo seria aproximadamente 4000USD se o valor usado em [?] for tomado como referência. Para uma melhor avaliação dos dados sem comprometer os resultados esperados, assumimos os custos normalizados fornecidos na Tabela IV. Outros elementos (ex. jumpers ópticos, conectores, dispositivos combinadores etc.) com custos de implementação não foram considerados, pois o custo total é considerado desprezível comparado com os custos dos elementos abordados neste estudo.

V. RESULTADOS

Os resultados obtidos para cada cenário proposto, foram compilados após a aplicação das métricas descritas, e foi gerado o gráfico apresentado na Fig. 2, onde pode-se observar que os cenários utilizando a mesma Fibra Óptica (FO) em uso e com tráfego na banda C, apresentaram um custo relativo com relação a outros cenários com a utilização de uma FO em paralelo. Os cenários com custo relativo mais baixo, apenas dois utilizam a mesma FO, contrastando com cenários que utilizam uma nova fibra óptica em paralelo, e dispositivos com altas taxas de transmissão na banda C.

A expansão da rede utilizando dispositivos na banda C, é a opção mais econômica mesmo em soluções com a utilização de uma nova FO, tanto em termos do custo por hardware implementado, quanto por bit transmitido. Observou-se também que as soluções de mesma fibra, esbarram na limitação das taxas de transmissão de dados dos dispositivos e componentes atuais, cuja maturidade dos dispositivos e componentes está

concentrado apenas em elementos disponíveis comercialmente para a banda C, com baixo crescimento na disponibilidade de dispositivos na banda L. Os resultados obtidos indicam também que, devido à baixa disponibilidade, maturidade e custos elevados dos componentes na banda S, torna os custos associados à sua implementação inviáveis no momento, se comparados a expansão de cenários na banda C e L, mesmo quando estas bandas utilizam fibras ópticas em paralelo.

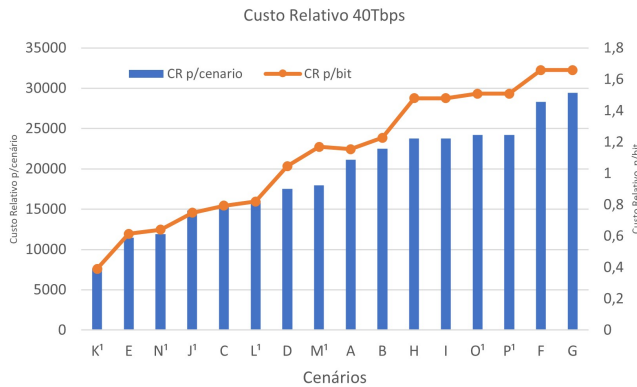


Fig. 2. Custo Relativo.

VI. CONCLUSÃO

De acordo com os resultados apresentados, a banda C ainda oferece a melhor relação custo-benefício, devido a enorme maturidade tecnológica e disponibilidade comercial de uma vasta gama de dispositivos, além de já possuírem uma infraestrutura madura e consolidada nas grandes operadoras de telecom. Pode-se observar que as bandas L e S esbarram nas limitações físicas da FO, devido a grande interferência provocada pelos efeitos não lineares, que afetam mais significativamente a qualidade de transmissão nestas faixas espectrais, fator importante para a transmissão em longa distância, que limita o alcance nos links de transmissão de dados. A maturidade tecnológica da banda C contribui para a sua predominância, enquanto o investimento no desenvolvimento de componentes para as bandas L, e especialmente a banda S permanece abaixo do esperado, e com um grau de maturidade tecnológico ainda incipiente. Concluímos que, apesar da maior largura de banda e capacidade potencial de canalização nas bandas L e S, no momento a expansão das redes na banda C, se torna mais viável do ponto de vista tecno-econômico. Embora as recentes pesquisas validando o uso de outras bandas no espectro óptico além da banda C. A aplicação de soluções de ampliação em outras bandas tais como, as bandas L e S, esbarram no alto custo dos dispositivos, e no elevado custo de investimento para o desenvolvimento de novos componentes nas referidas bandas ópticas. Portanto, a eficiência e a economicidade associadas a banda C, fazem que as soluções utilizando esta banda óptica sejam a escolha mais prática, econômica e mais viável no momento, quando se trata da ampliação de redes ópticas no cenário atual.

REFERENCIAS

- [1] International Telecommunications Union - ITU. 2024. Internet traffic. Available at: <https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-traffic/>
- [2] PRADO, Tiago S. 2023. "Value-added services and the future of telecommunications: empirical evidence to inform the network fees debate in Brazil".
- [3] Guilherme Reis, José. Guarany, Marcelo. "Projeções da demanda por tráfego de dados no Brasil: uma atualização". Maio, 2024.
- [4] Ningning Guo, Kai Zhang, Yongcheng Li, Ning Deng, Sanjay K. Bose, and Gangxiang Shen, "Impact of the band upgrade sequence on the capacity and capital expenditure of multi-band optical networks," Vol. 15, No. 10 / October 2023 / Journal of Optical Communications and Networking.
- [5] <https://www.infinera.com/blog/all-you-ever-wanted-to-know-about-optical-transmission-bands/tag/optical/>
- [6] Nicola Sambo, Bruno Correia, Antonio Napoli, João Pedro, Leily Kiani, Piero Castoldi, AND Vittorio Curri. "Network upgrade exploiting multi band: S- or E-band?". Vol. 14, No. 9 / September 2022 / Journal of Optical Communications and Networking
- [7] R. Schmogrow, "Solving for scalability from multi-band to multi-rail core networks," J. Lightwave Technol. 40, 3406–3414 (2022).
- [8] Shariati, Behnam. Khodashenas, Pouria Sayyad. Rivas-Moscoso, José Manuel. Ben-Ezra, Shalva. Klonidis, Dimitrios. Jiménez, Felipe. Velasco, Luis. Tomkos, Ioannis. "Investigation of Mid-term Network Migration Scenarios Comparing Multi-Band and Multi-Fiber Deployments".
- [9] <https://www.fibermall.com/pt/blog/coherent-optical-communication.htm>
- [10] <https://www.ciena.com/insights/what-is/what-is-roadm.html>
- [11] M. Devigili, P. Nadimi Goki, N. Sambo, P. Castoldi, L. Pot'1, A. D'Amico, V. Curri. "Multi-band Optical Network Assisted by GNPY: an Experimental Demonstration".
- [12] Alessio Ferrari, Mark Filer, Karthikeyan Balasubramanian, Yawei Yin, Esther Le Rouzic, Jan Kundrat, Gert Grammel, Gabriele Galimberti, Vittorio Curri. "Experimental Validation of an Open-Source Quality of Transmission Estimator for Open Optical Networks".
- [13] G. Grammel et al, "Physical simulation environment of the telecommunications infrastructure project (TIP)", OFC, pp. M1D-3, 2018.
- [14] J. Auge et al, "Open optical network planning demonstration.", OFC, pp. M3Z-9, 2019.
- [15] <https://github.com/Telecominfraproject/oopt-gnpy>
- [16] Alessio Ferrari, Mark Filer, Karthikeyan Balasubramanian, Yawei Yin, Esther Le Rouzic, Jan Kundrat, Gert Grammel, Gabriele Galimberti, Vittorio Curri. "GNPY: an opensource application for physical layer aware open optical networks".
- [17] Pedro Jorge de Carvalho Souza, "Uma nova estratégia para resolver o problema RMLSA considerando efeitos da camada física num tráfego estatico", Tese (Mestrado) – Universidade Federal de Pernambuco, 2018.
- [18] Cho, K.; Yoon, D. "On the general ber expression of one-and two-dimensional amplitude modulations". IEEE Transactions on Communications, IEEE, v. 50, n. 7, p. 1074–1080, 2002. 46]
- [19] Alessio Ferrari, Mark Filer, Karthikeyan Balasubramanian, Yawei Yin, Esther Le Rouzic, Jan Kundrat, Gert Grammel, Gabriele Galimberti, AND Vittorio Curri. "GNPY: an open-source application for physical layer aware open optical networks".
- [20] M. Filer, M. Cantono, A. Ferrari, G. Grammel, G. Galimberti, and V. Curri, "Multi-vendor experimental validation of an open source QoT estimator for optical networks," J. lightwave Technol. 36, 3073–3082 (2018).
- [21] Masahiro Nakagawa, Takeshi Seki, Takashi Miyamura. "Techno-Economic Potential of Wavelength-Selective Band Switchable OXC in S+C+L Band Optical Networks".
- [22] R. Huelsermann, M. Gunkel, C. Meusburger, D. A. Schupke, "Cost modeling and evaluation of capital expenditures in optical multilayer networks", Journal of Optical Networking, v. 7, n. 9, p. 814–833, Setembro 2008.
- [23] R. K. Jana, A. Srivastava, A. Lord, and A. Mitra, "Optical cable deployment versus fiber leasing: an operator's perspective on CapEx savings for capacity upgrade in an elastic optical core network," J. Opt. Commun. Netw. 15, C179–C191 (2023).

Antena *patch* para comunicação *wireless* bioinspirada na flor *Hoya Pubicalyx*

Joanes Martins Galvão
Faculdade de Engenharia Elétrica
Universidade Federal do Pará
Tucuruí, Pará, Brasil
joanes.galvao@tucurui.ufpa.br

Leonardo França da Cruz
Faculdade de Engenharia Elétrica
Universidade Federal do Pará
Tucuruí, Pará, Brasil
leonardo.cruz@tucurui.ufpa.br

Ryan da Costa Sousa
Faculdade de Engenharia Elétrica
Universidade Federal do Pará
Tucuruí, Pará, Brasil
ryan.sousa@tucurui.ufpa.br

Andrécia Pereira da Costa
Faculdade de Engenharia Elétrica
Universidade Federal do Pará
Tucuruí, Pará, Brasil
andreciacosta@ufpa.br

André Felipe Souza da Cruz
Faculdade de Engenharia Elétrica
Universidade Federal do Pará
Tucuruí, Pará, Brasil
andcruz@ufpa.br

Rafael Suzuki Bayma
Faculdade de Engenharia Elétrica
Universidade Federal do Pará
Tucuruí, Pará, Brasil
rafaelbayma@ufpa.br

Resumo—Este artigo apresenta o projeto de uma antena *patch* bioinspirada na geometria da flor *Hoya Pubicalyx*, popularmente conhecida como flor de cera. A antena possui dimensões de $42,8 \times 42,8 \times 1,5$ mm³, fabricada em um substrato FR4, com permissividade relativa de 4,4 e tangente de perdas de 0,02, alimentada por um cabo coaxial, com impedância de entrada de 50 Ω . Para a caracterização numérica, foi utilizado o *software* *Ansys Electronics Desktop* (HFSS) e, para a caracterização experimental, foi utilizado o LiteVNA-64. Os resultados obtidos indicam que a antena proposta apresenta largura de banda de perda de retorno $S_{11} \leq -10$ dB e VSWR abaixo de 2 dB, considerando a frequência de ressonância simulada e medida de 2,46 GHz e 2,42 GHz. Os diagramas de irradiação simulados possuem características direcionais, com ganho máximo de 5,91 dBi na frequência de 2,46 GHz. Os resultados experimentais mostram uma boa concordância com os resultados simulados.

Palavras-chave—Antena *patch*, cabo coaxial, FR4, flor *Hoya Pubicalyx*, geometria bioinspirada.

I. INTRODUÇÃO

A crescente utilização da comunicação sem fio tem despertado nos pesquisadores a busca de novos dispositivos com características atrativas, tais como: baixo peso, leveza, custo de produção reduzido, bons níveis de qualidade de serviço, além de atender os requisitos necessários sem perda de desempenho, [1],[2]. Para atender essa demanda, podem ser utilizadas as antenas em microfita, as quais são estruturas metálicas que se baseiam em uma placa metálica chamada *patch*, posicionada sobre um substrato dielétrico, que por sua vez é colocado sobre um plano de terra [3]. Possui várias técnicas de alimentação, as mais utilizadas são por linha de microfita ou cabo coaxial conectado ao plano de terra [4].

As antenas *patch* em microfita apresentam diversas vantagens em relação às de micro-ondas convencionais, uma vez que permitem diferentes possíveis geometrias que podem ser utilizadas em circuitos e/ou dispositivos de alta frequência, atendendo às exigências de uma grande variedade de serviços e à mobilidade [1], [5], [6].

Neste sentido, projetos de antenas em microfita bioinspiradas na natureza têm sido propostos na literatura, e têm apresentado resultados aceitáveis em várias faixas de frequência. As plantas apresentam um bom modelo de estudo para projetos que utilizam o espectro eletromagnético, devido sua geometria ser favorável à captação para fotossíntese, de forma semelhante às antenas no espectro de recepção de ondas eletromagnéticas [1]. Além disso, existe a busca para descrever geometrias da natureza por modelos matemáticos, onde algumas técnicas podem ser citadas, como a razão áurea, a sequência de Fibonacci, fractais, super fórmula de Gielis, sistemas de Lindenmayer, equações polares, entre outras [7].

Assim, neste trabalho é apresentado o projeto e caracterização experimental de uma antena *patch* bioinspirada na flor de cera (*Hoya Pubicalyx*). A antena foi idealizada para operar na faixa de frequência 2,4 GHz, onde operam diversos sistemas de comunicação *wireless* (sem fio). Foi realizada a caracterização numérica utilizando o *software* *High Frequency Structure Simulator* (HFSS) da Ansys [8], e para a caracterização experimental foi utilizado um analisador de rede vetorial portátil do tipo LiteVNA-64.

II. ANTENA BIOINSPIRADA NA FLOR DE CERA

O procedimento do projeto da antena *patch* proposta partiu da análise da geometria da antena flor de cera, Fig.1a.

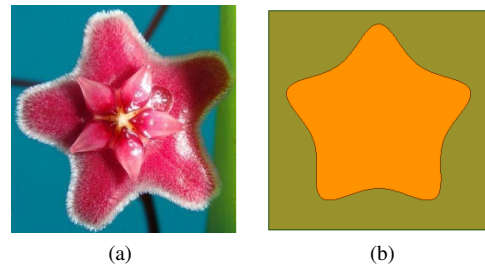


Fig. 1. (a) Foto da flor de cera [10]; (b) Modelo gerado no MATLAB®

As dimensões para um projeto inicial podem ser aproximadas por frações do comprimento de onda (λ) para a primeira ressonância correspondente a operação na faixa de frequência de 2,4 GHz [9]. Para o dimensionamento do elemento irradiante da antena proposta, foi necessário determinar a largura W e o comprimento L aproximados da geometria da flor de cera. Para isso, foi utilizada a superfórmula de Gielis [9]. Ao implementar a superfórmula, e realizar o ajuste dos parâmetros no *software* MATLAB®, foi possível obter o caminho de contorno da geometria desejada, como é mostrado na Fig. 1(b).

A partir do resultado gráfico da Fig. 1(b), a imagem foi gerada no formato PNG (*Portable Network Graphics*) e em seguida, formatada em arquivo DXF (*Drawing Exchange Format*), para posterior fabricação. Ao implementar a superfórmula de Gielis no *software* HFSS, foi possível construir a geometria e executar as simulações eletromagnéticas. Na Fig. 2 são mostrados os detalhes da geometria do *patch* (com contorno de 5 pontas suaves simétricas ao centro), do *ground* (plano de terra) e da alimentação por cabo coaxial.

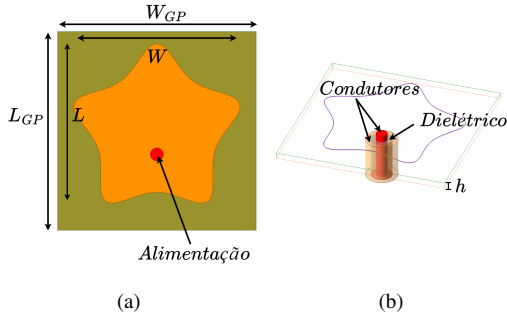


Fig. 2. (a) Geometria da antena bioinspirada na flor de cera; (b) Detalhe da alimentação por conector

Na Tab. I são mostradas as dimensões das partes funcionais da antena *patch* proposta.

TABELA I
DIMENSÕES DA ANTENA PROPOSTA

	Substrato e plano terra	Patch	Alimentação (diâmetro condutor)		
Parâmetro	$W_{GP} = L_{GP}$	$W = L$	Raio	Raio	Cilindro
Valor	4,28	3,57	0,14	0,4	0,26

Para a fabricação da antena proposta foi utilizado o substrato dielétrico de fibra de vidro FR4 com espessura $h = 0,15$ cm, constante dielétrica de 4,4 e tangente de perdas de 0,02. A antena foi alimentada por um conector SMA inserido por baixo do plano de terra. Para a modelagem do cabo coaxial de alimentação, admitiu-se o condutor interno com 0,14 cm de diâmetro, condutor externo 0,4 cm e o cilindro dielétrico de 0,26 cm com permissividade relativa de 2,25.

Para encontrar o ponto de alimentação com impedância de 50 *ohms*, foi utilizado o *software* HFSS para realizar uma varredura em uma região de simetria do *patch*, onde observou-se menor perda de retorno na posição $x = 0,507$ cm do centro da antena (na direção do comprimento L).

O protótipo da antena foi confeccionado utilizando um adesivo de vinil depositado sobre o condutor virgem da placa de FR4. Por meio da fabricação digital, foi possível utilizar uma CNC laser para executar o corte computadorizado da geometria da flor de cera sobre o vinil depositado. Após a execução do corte, parte do vinil foi removido, e então deu-se início ao processo de ataque químico por corrosão de perclorato de ferro, para extrair o condutor indesejado da placa, mantendo apenas o *patch* com geometria da flor de cera. Por fim, utilizando um furador de placas e uma estação de solda, foi realizada a soldagem do conector SMA.

Todo o processo de fabricação foi realizado no Laboratório de Fabricação - FABLAB do Campus Universitário de Tucuruí, da Universidade Federal do Pará. As etapas do processo de fabricação são demonstradas na Fig. 3.

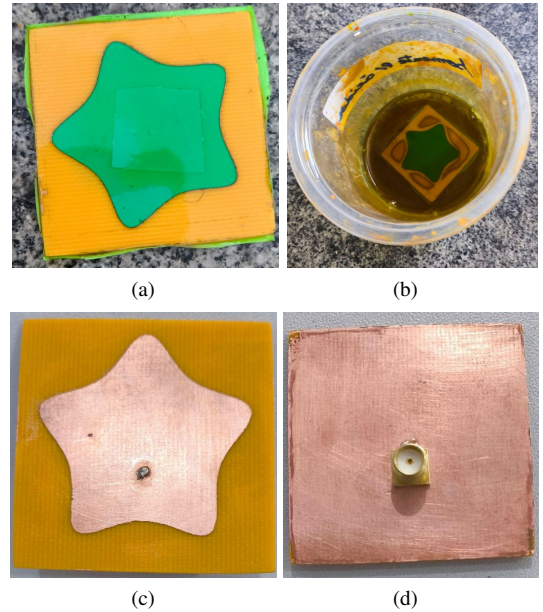


Fig. 3. Processo de fabricação da antena proposta: (a) Antena com máscara adesiva; (b) Placa submersa no perclorato de ferro; (c) Antena fabricada (vista de topo); (d) Vista inferior (plano de terra)

III. RESULTADOS SIMULADOS E MEDIDOS

Na análise de resultados da antena proposta são apresentados a perda de retorno (parâmetro S_{11}), a VSWR (*Voltage Standing Wave Ratio*), a largura de banda de -10 dB, a Carta de *Smith*, mostrando a impedância de entrada da antena, a densidade superficial de corrente e campo distante no *patch* da antena, e os diagramas de irradiação 2D.

Os processos de simulação e caracterização experimental foram realizados no Laboratório de Eletromagnetismo Aplicado - LEMAG, do Campus Universitário de Tucuruí, da Universidade Federal do Pará. Após as simulações no *software* HFSS, deu-se início o processo de caracterização experimental utilizando um analisador de rede vetorial portátil de baixo custo, do tipo LiteVNA-64, na faixa de medição 50 kHz a 6,3 GHz.

A leitura dos resultados medidos pode ser visualizada na tela do notebook, sendo possível obter os dados em arquivo no formato .csv ou .slp a partir do *software* executável VNA-QT [11]. O *setup* de medição é mostrado na Fig. 4.

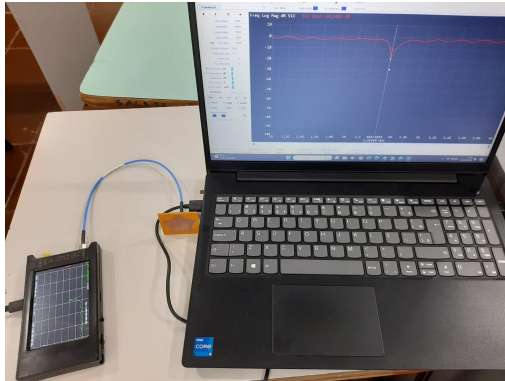


Fig. 4. Setup de medição

Uma análise comparativa dos resultados simulados e medidos na faixa de frequência de 1,5 GHz a 3,5 GHz são observados nas Fig. 5 e Fig. 6.

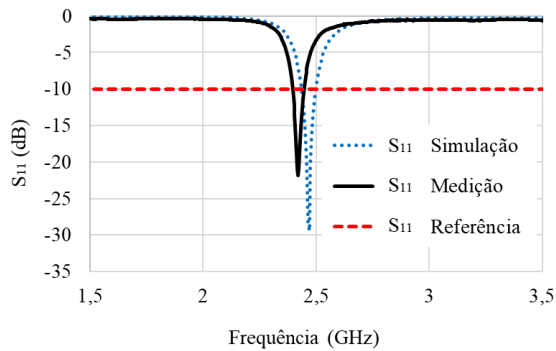


Fig. 5. Resultado simulado e medido do S_{11} para a antenna proposta

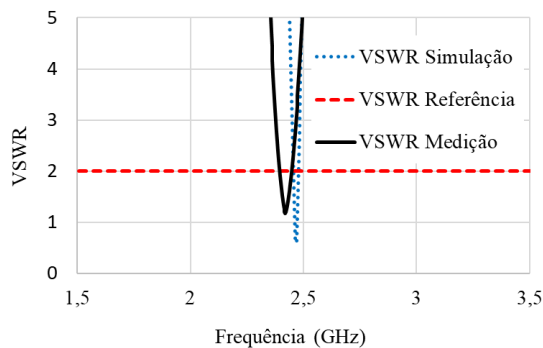


Fig. 6. Resultado simulado e medido do VSWR para a antenna proposta

Na Fig. 5 é mostrada a perda de retorno (parâmetro S_{11} , ou coeficiente de reflexão Γ da antenna) obtidos via simulação, e medição experimental, comparados com o nível -10 dB.

Desta forma, o resultado simulado da antenna projetada apresenta perda de retorno de $S_{11} = -29,38$ dB, frequência de ressonância em 2,46GHz, largura de banda de 60 MHz. Enquanto isso, os resultados obtidos via medição mostram perda de retorno de $S_{11} = -21,8$ dB, frequência de ressonância em 2,42 GHz, largura de banda de 50 MHz. A partir da Fig. 6 verifica-se que a antenna simulada apresentou VSWR de 0,58 dB, enquanto a medição experimental foi de 1,17 dB, ambos em suas respectivas frequências de ressonância, estando abaixo da faixa crítica de 2 dB. É perceptível através da comparação dos resultados de S_{11} e VSWR, simulados e medidos, que estão em concordância e dentro dos limites aceitáveis, indicando que a antenna apresenta mínimas reflexões e o sinal injetado nela pode ser transmitido.

Os valores simulados e medidos para a impedância de entrada da antenna, traçados sobre a carta de Smith, podem ser visualizados na Fig. 7, em que o centro da carta indica o casamento de impedâncias em 50 ohms.

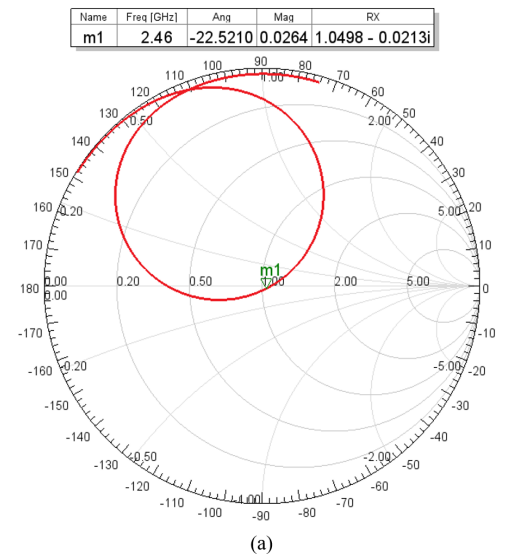
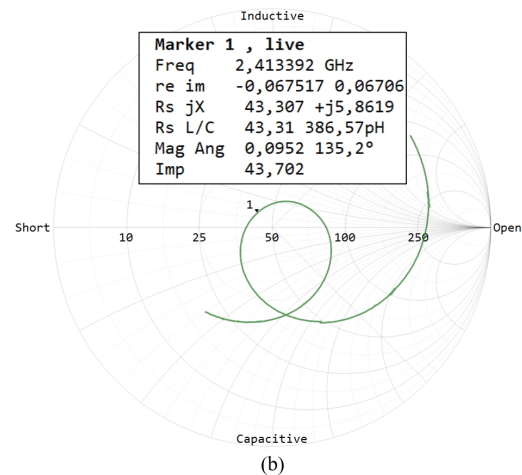


Fig. 7. Gráficos dos valores de impedância de entrada da antenna bioinspirada na flor de cera sobre a carta de Smith: (a) Simulação em 2,46 GHz; (b) Medição em 2,41 GHz

A partir dos resultados de carta de *Smith* na Fi. 7(b), verificou-se que a impedância de entrada medida para a antena foi de $43,30 \Omega + j5,8619 m\Omega$ na frequência próxima de 2,42 GHz, demonstrando uma mínima parte reativa e indutiva.

Os diagramas de radiação simulados para a antena proposta (ganhos direcionais no plano E e H) são mostrados na Fig. 8, considerando a frequência de ressonância em 2,46 GHz.

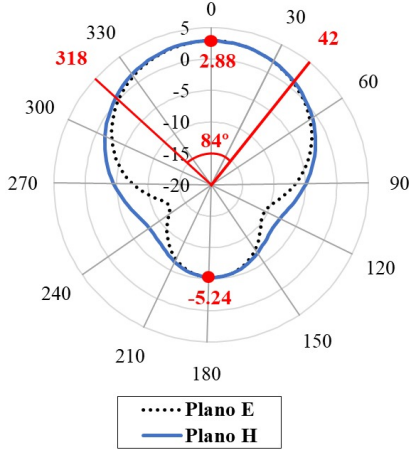


Fig. 8. Diagrama de irradiação no plano E, e no plano H, ambos na frequência de ressonância de 2,46 GHz

Pode-se observar que o diagrama da antena proposta no plano E e plano H apresenta máxima irradiação na direção $\theta = 0^\circ$, com ganho máximo de 2,88 dB. Além disso, a antena tem uma largura de feixe de meia potência (HPBW) de aproximadamente 84° em ambos os planos, demonstrando simetria de radiação, com um pequeno lóbulo traseiro de $-5,24$ dB.

Para demonstrar o modo fundamental ressonante da antena, na Fig. 9 é mostrada a distribuição da densidade superficial de corrente (J , A/m) no *patch* da antena.

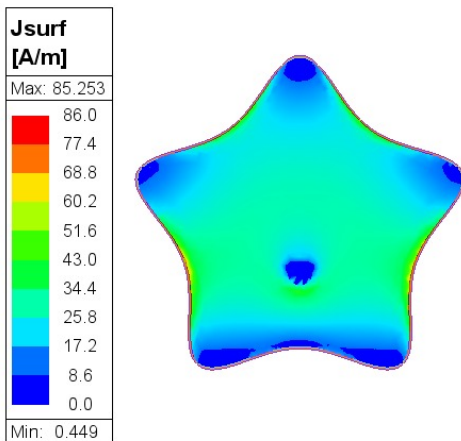


Fig. 9. Distribuição de densidade de corrente superficial (J , A/m).

Nota-se uma maior concentração nas bordas da antena e na região próxima à alimentação. Verifica-se que a a distribuição é menor nas pontas da geometria da flor de cera.

Por fim, na Fig. 10 é mostrado o diagrama de campo distante, concêntrico com a geometria da antena *patch* proposta.

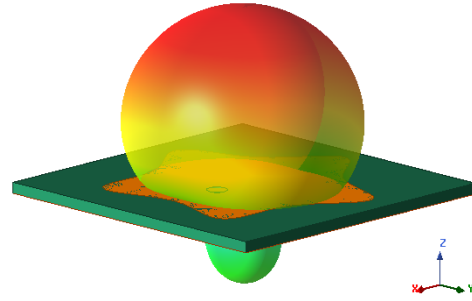


Fig. 10. Diagrama de campo distante da antena *patch* proposta.

A partir da Fig. 10 constata-se que o diagrama de radiação da antena proposta é simétrico nos planos E e H, semelhante aos diagramas apresentados por antenas *patches* circulares. Ao comparar a antena proposta, que possui um *patch* com diâmetro efetivo de 3,57 cm, com uma antena de *patch* circular com diâmetro 3,38 cm, verifica-se uma melhora na antena proposta com diminuição do lóbulo traseiro, e aumento do ganho de diretividade.

IV. CONCLUSÃO

Neste artigo foi apresentado a caracterização numérica e experimental de uma antena de microfita do tipo *patch*, bioinspirada na flor *Hoya Pubicalyx* (popular flor de cera). A antena foi projetada para operar na frequência de ressonância de 2,4 GHz, no entanto, através das medições realizadas utilizando um *setup* com um analisador de rede vetorial portátil, o protótipo fabricado ressoou em 2,42 GHz, apresentando perda de retorno $-21,8$ dB e largura de banda de 50 MHz. Em comparação as antenas *patches* convencionais, a antena apresentou melhor ganho de diretividade. Logo, pode-se inferir que a antena projetada apresenta boa eficiência de irradiação, o que a torna uma boa escolha para aplicações em sistemas de comunicação sem fio.

V. AGRADECIMENTOS

Os autores agradecem à Fundação Amazônia Paraense de Amparo à Pesquisa - FAPESPA pelo financiamento deste projeto de pesquisa e agradecem, ainda, a contribuição dos colaboradores do Laboratório de Eletromagnetismo Aplicado - LEMAG, e do Laboratório de Fabricação Digital - FABLAB, ambos localizados no Campus Universitário de Tucuruí, da Universidade Federal do Pará.

REFERÊNCIAS

- [1] E. E. C. Oliveira, R. B. Bezerra, T. A. C. Silva, R. C. F. Medeiros. "Antena patch bioinspirada em um trevo de três folhas para a faixa de frequência de comunicação sem fio," em *Revista Científica Semana Acadêmica*. Fortaleza, 2017. Disponível: <https://semanaacademica.org.br/artigo/antena-patch-bioinspirada-em-um-trevo-de-tres-folhas-para-faixa-de-frequencia-de-comunicacao>. Acesso Jul. 26, 2024.

- [2] A. Alsudani, H. M. Marhoo, "Design and Enhancement of Microstrip Patch Antenna Utilizing Mushroom Like-EBG for 5G Communications," *Journal of Communications*. vol. 18, no. 3, March 2023, pp.156-163, 2023.
- [3] C. A. Balanis, *Antena Theory – Analysis e Design*, 4th Edition, John Wiley & Sons, Inc. USA, 2016.
- [4] R. Garg, P. Bhartia, I. Bahl, A. Ittipiboon, "Microstrip Antenna Design Handbook. *Artech House Antennas and Propagation Library*, 2000, pp.14–19.
- [5] K. L. Wong, *Compact and broadband microstrip antennas*, New York, John Wiley & Sons, Inc. USA, 2002.
- [6] A. S. M. Z. Kausar, A. W. Reza, A. A. Latef e M. H. K. M. E. Ulla, "Optical nanoantenas: state of the art, scope and challenges as a biosensor along with human exposure to nano-toxicology," *Sensors*, vol. 15, pp. 8787-8831, 2015.
- [7] M. A. de Oliveira et al., "Using polar transformation to design a dissimilar antenna array inspired on four-leaf clover," 2018 *IEEE Radio and Wireless Symposium (RWS)*, Anaheim, CA, USA, 2018, pp. 228-230, doi: 10.1109/RWS.2018.8304994.
- [8] <http://www.ansys.com>
- [9] P. F. S. Júnior; J. C. Silva, P. H. F. Silva. "Projeto de uma Antena de Folha para 700 MHz". Em: MOMAG 2014: 16º SBMO - *Simpósio Brasileiro de Micro-ondas e Optoeletrônica e 11º CBMag* - Congresso Brasileiro de Eletromagnetismo. Curitiba, pp.758-763, 2014.
- [10] <https://garden.org/plants/view/679235/Wax-Plant-Hoya-pubicalyx-Bright-One/>
- [11] <https://nanorfe.com/nanovna-v2-software.html>

Heuristic Approach to Optimize VNF Placement and Routing in 6G Networks

Matheus Pantoja
Universidade Federal do Pará
Belém, Brasil
matheus.pantoja@itec.ufpa.br

Albert Santos
Universidade Federal do Pará
Belém, Brasil
albert.santos@itec.ufpa.br

Rafael Vieira
Universidade do Estado do Amapá
Macapá, Brasil
rafael.vieira@ueap.edu.br

Carlos Natalino
Chalmers University of Technology
Chalmers, Suécia
carlos.natalino@chalmers.se

Diego Cardoso
Universidade Federal do Pará
Belém, Brasil
diego@ufpa.br

Abstract—This paper proposes the Efficient Node Activation Heuristic (ENAH) to solve the Virtual Network Function Placement and Routing Problem (VNF-PR) in 6G networks. ENAH effectively manages the On/Off state of servers, activating only the necessary servers before allocating Virtual Network Functions (VNFs). This strategic activation minimizes both latency and energy consumption while maintaining high Quality of Service (QoS). Extensive simulations in high-demand scenarios demonstrated that ENAH significantly reduces latency by up to 51% and energy consumption by up to 68%. These improvements contribute substantially to the efficiency, sustainability and performance of modern telecommunication networks.

Index Terms—VNF, NFV, Service Function Chain, VNF Placement and Routing, 6G

I. INTRODUCTION

In the current telecommunications scenario, driven by advances in cloud computing and virtualization technology, network service providers have increasingly adopted the implementation of Virtual Network Functions (VNFs) [1]. VNFs are software components that can be optimally incorporated into general computing platforms. This approach, by avoiding the need for dedicated and specialized hardware, not only reduces operational costs but also increases flexibility and agility in deploying new network services [2]. However, the diversity of available services, each with its own requirements in terms of resources represents a significant challenge. These services are provided as an ordered sequence of VNFs, forming a service function chaining (SFC) [3]. Thus, the effective deployment and resource allocation for these VNF chains represent a complex yet crucial challenge that needs to be overcome [4].

Network Function Virtualization (NFV) is a well-established concept proposed by the European Telecommunications Standards Institute (ETSI) to simplify the deployment and management of network services using virtualization technologies [4]. Compared to traditional hardware appliances, NFV allows service deployment on virtual machines to enhance flexibility and scalability. The basic idea behind NFV is to decouple network functions, such as firewall, proxy, or intrusion detection systems from dedicated hardware, allowing

them to run as software [3]. This approach enables network functions to be implemented as dynamic applications, activated according to specific demands and locations. This technology facilitates efficient traffic management, allowing computational scalability and high-precision connection management [2].

The energy consumed in data centers is a growing concern, representing a significant portion of operational costs, with forecasts that it may exceed capital expenditures (CAPEX) in the near future [1]. In the context of NFV and VNF, energy consumption impacts the overall efficiency and sustainability of network operations. High energy consumption not only increases operational costs but can also affect the Quality of Service (QoS) by limiting the resources available for network functions. In NFV-enabled networks, computing capacity is designed for traffic peaks, making it unnecessary to activate all servers during low traffic periods [5]. Controlling the number of active nodes is crucial for efficient resource utilization and reducing operational costs while ensuring low latency to guarantee QoS and end-user satisfaction [6]. In NFV-enabled networks, latency can be influenced by the location of servers, and an efficient solution for VNF placement on servers, considering the selection of the most suitable servers to be activated, can significantly minimize latency and improve service quality.

Considering resource constraints in the network infrastructure, a key feature for the success of NFV is the economical placement of VNFs and the routing of service demands [3]. This study focuses on the VNF Placement and Routing Problem (VNF-PR), which aims to optimize the activation of the necessary servers and subsequently the placement of VNFs to use resources effectively. The problem consists of finding (i) the optimal placement of VNFs on nodes and (ii) feasible routing paths, respecting node latency and capacity constraints [2].

The main motivation of this work is to develop an efficient approach for controlling the activation of network nodes that minimizes latency and energy consumption while maintaining QoS and network availability, especially in 6G networks.

The primary contribution is the proposal of a heuristic that improves resource utilization by selectively activating network nodes based on demand, thereby reducing operational costs and service demand latency. As a consequence, this selective activation leads to an optimized VNF allocation. To validate the approach, we conducted tests and analyzed two classic objectives: energy consumption and QoS (minimization of total latency). The remainder of the paper is structured as follows: Section II presents a review of related work. Section III defines the VNF-PR. The heuristic is introduced in Section IV, followed by the discussion of results in Section V. Finally, Section VI concludes this work.

II. RELATED WORK

The strategy of turning servers on and off to optimize energy efficiency is widely accepted as an effective solution. The main difference between approaches lies in how to determine which servers should be activated or deactivated, which can significantly impact energy savings and network efficiency.

The solution proposed in [7] analyzed scheduling policies for NFV in 6G networks, considering server reliability and boot time. A Markov chain model was proposed to account for the on/off states of servers and their impact on system reliability, along with an algorithm to set activation/deactivation thresholds aimed at optimizing energy consumption while ensuring the desired reliability. The results showed that the proposed approach can significantly improve energy efficiency while maintaining system reliability.

The architecture proposed in [6] for 6G networks integrates holistic network virtualization and pervasive network intelligence. This architecture includes network slicing management and the control of on/off states of nodes and links in software-defined networks (SDN) to save energy. Resource virtualization and network intelligence are addressed together to optimize network performance, including energy efficiency and resource allocation. The results indicate that the integrated approach can lead to significant energy savings and better utilization of network resources.

The approach in [5] for energy efficiency in different modes of SDN/NFV networks, including partial and hybrid networks for 6G, proposes a general system model and a multi-stage graph-based algorithm to optimize energy consumption. The algorithm considers the routing and placement of network functions as well as the on/off states of nodes and links to maximize energy efficiency during the transition to fully virtualized networks. Simulation results showed a significant reduction in energy consumption, highlighting the effectiveness of the proposed approach for transitional networks.

The solution proposed in [8] optimizes the embedding and scheduling of SFC in 6G networks with NFV. The approach selects physical nodes with abundant resource capacities to accommodate the SFCs, linking the necessary nodes and optimizing the total scheduling time of the SFC. This process involves the dynamic activation and deactivation of servers to ensure efficient resource use. Experimental results show an

SFC acceptance rate at least 10% higher than typical heuristics, highlighting its efficiency in dynamic server management and resource optimization.

As can be observed, most existing approaches do not consider the joint optimization of latency and energy consumption, nor the activation of nodes based on demand. Additionally, these works propose specific algorithms that may not be easily scalable for large-scale networks due to their complexity. This work aims to fill these gaps by proposing an efficient and scalable heuristic for VNF-PR in 6G networks, which will be detailed in the following sections.

III. PROBLEM DEFINITION

The VNF-PR involves finding the placement of virtual network functions on nodes and the routing paths so that the sum of the network function installation costs and node activation costs is minimized. Solutions must consider node and network function capacity constraints, as well as flow precedence and routing [2]. A solution to the problem is considered feasible if it satisfies all required constraints and is called optimal if the associated cost is minimized. The associated cost is defined according to each specific objective, which can be the sum of node activation costs, energy consumption, or even the total system latency. The problem is NP-Hard, even for a single flow and without additional constraints [3].

A. System Model

The telecommunications network is modeled as a bidirectional graph $G(N, L)$, where N represents the set of nodes (servers or switches) and L is the set of links interconnecting these nodes. Each node $i \in N$ can be equipped with computational resources, such as CPU, memory, and storage, necessary to host VNFs. The links L are characterized by bandwidth capacity and latency.

Client demands are modeled as a tuple $d(d_o, d_d, T_d, B_d, c_d)$, where d_o is the demand's origin node, d_d is the destination node, T_d is the allowable end-to-end latency, B_d is the required bandwidth, and c_d is the ordered sequence of VNFs that the demand must traverse. Each VNF $v \in c_d$ must be allocated on a network node, and the links between nodes must be selected to minimize total latency.

B. Energy Consumption Model

To model energy consumption in the network, we consider switches and servers (PMs). The energy consumption of switches has a constant component and a variable component related to the number of active ports [9]. Thus, the total energy consumed by a switch is:

$$P_{sw} = P_{const} + P_{port} \cdot N_{port} \quad (1)$$

where P_{const} is the idle state energy consumption and P_{port} is the energy consumed by each active port. N_{port} represents the total number of active ports. If a physical link $l_{i,j} \in L$ between nodes i and j is active, it requires two ports. Thus, the total network energy consumption is given by:

$$P_{network}^T = P_{const} \sum_{i \in N} y_i + 2P_{port} \sum_{(i,j) \in L} l_{i,j} \quad (2)$$

Here, $l_{i,j}$ is 1 if the link between i and j is active, and y_i indicates if switch i is active.

In data centers, the energy consumed by computing nodes mainly depends on the CPU, which is the largest energy consumer. Therefore, the energy consumption model for PMs considers CPU utilization. The energy consumption of a PM is proportional to CPU load, varying from a minimum consumption in idle state to a maximum at full load [10]:

$$P_{pm} = P_{idle} + (P_{max} - P_{idle}) \times \theta_{cpu} \quad (3)$$

where P_{idle} is the energy consumed when the CPU is at 0

$$\theta_{cpu} = \frac{C_{util}}{C_{total}} \quad (4)$$

Considering C_{util} as the utilized resource, the total energy consumption of a PM, P_{pm}^T , is described as:

$$P_{pm}^T = \sum_{i \in N} \left(P_{idle} x_i + (P_{max} - P_{idle}) \sum_{f \in F} \sum_{C_{util} \in R} \frac{C_{f,C_{util}}}{C_{i,C_{util}}^T} z_{i,f} \right) \quad (5)$$

where x_i is a binary variable that indicates if the PM is active, and $z_{i,f}$ is an integer variable representing the number of instances of function f allocated to the PM at node i .

C. System Latency

Eq. 6 below calculates the total system latency through two components: (i) the total delay introduced to all services by VNF processing, and (ii) the total delay due to data transmission between network nodes. The overall system latency is defined as:

$$T_{total} = T_{proc} + T_{trans} \quad (6)$$

T_{proc} represents the total processing latency of the system and is expressed as:

$$T_{proc} = \sum_{v \in V} \sum_{i \in N} \sum_{d \in D} u_{i,v,d} \cdot T_v \quad (7)$$

where $u_{i,v,d}$ is a binary variable that indicates if demand d is being processed by VNF v allocated at node i , and T_v represents the processing latency of VNF v .

T_{trans} represents the total transmission latency of the system and is expressed as:

$$T_{trans} = \sum_{(i,j) \in L} \sum_{(k,l) \in F} \sum_{d \in D} w_{i,j,k,l,d} \cdot T_{i,j} \quad (8)$$

where $w_{i,j,k,l,d}$ is a binary variable that indicates if virtual link (k,l) of demand d uses physical link (i,j) , and $T_{i,j}$ represents the transmission latency of physical link (i,j) .

Algorithm 1 Efficient Node Activation Heuristic (ENAH)

```

0: Input: Graph  $graph$ , Set of Nodes  $N$ , Set of Links  $l$ ,
   Distance Vector  $D$  (initialized to zeros and of size equal
   to the number of nodes in the graph)
0: Output: Index of the node to be activated
0: if all nodes are off then
0:   for each node  $n \in N$  do
0:     for each node  $i \in N$  do
0:       if the link between  $n$  and  $i$  is on then
0:          $D[n] \leftarrow D[n] + \text{euclidean distance}(n, i)$ 
0:       else
0:          $D[n] \leftarrow D[n] + \text{euclidean distance}(n, i) \times 2$ 
0:       {Assign a higher weight to unconnected links}
0:     end if
0:   end for
0:   end for
0:   Node  $n$  with the smallest sum is turned on
0: else {Calculate the distances of nodes relative to the con-
   nected nodes}
0:   Initialize the list of connected nodes:
   ConnectedNodes  $\leftarrow []$ 
0:   for each node  $n \in N$  do
0:     if  $n.on = \text{True}$  then
0:       Add  $n$  to the list ConnectedNodes
0:     end if
0:   end for
0:   for each node  $n \in N$  do
0:     for each node  $i \in \text{ConnectedNodes}$  do
0:       if the link between  $n$  and  $i$  is on then
0:          $D[n] \leftarrow D[n] + \text{euclidean distance}(n, i)$ 
0:       else
0:          $D[n] \leftarrow D[n] + \text{euclidean distance}(n, i) \times 2$ 
0:       {Assign a higher weight to unconnected links}
0:     end if
0:   end for
0:   end for
0:   Node with the smallest distance is turned on
0: end if=0

```

IV. HEURISTIC FORMULATION

We consider that all servers have the same processing capacity. In a scenario where all servers are turned on, demand allocation becomes less complex, as any available server can meet the requests.

The Efficient Node Activation Heuristic (ENAH) aims to minimize latency by activating servers based on network topology and current demand. By iteratively applying these steps, ENAH efficiently activates servers based on evolving network demands, ensuring efficient resource use and minimizing latency. This results in more efficient demand allocation, directing traffic to servers based on the shortest total distance. ENAH operates in two distinct scenarios:

1. All Nodes Off: When all servers are initially turned off, the algorithm calculates the sum of distances from each

TABLE I
CONSIDERED SERVICES [11]

Service Type	VNF Chain	Bandwidth	Delay	Traffic %
Web Service	NAT-FW-TM-WOC-IDPS	100 kbps	500 ms	18.2%
VoIP	NAT-FW-TM-FW-NAT	64 kbps	100 ms	11.8%
Video Streaming	NAT-FW-TM-VOC-IDPS	4 Mbps	100 ms	69.9%
Online Games	NAT-FW-VOC-WOC-IDPS	50 kbps	60 ms	0.1%

node to all other nodes in the network. If a link between two nodes is already active, the Euclidean distance is used. If the link is inactive, the distance is multiplied by a penalty factor (in this case, 2) to discourage the use of links that would require activation. The node with the shortest distance, which in this scenario equals the smallest sum of distances, is then activated. The idea in this scenario is to identify and activate the most central node in the network, which has the shortest accumulated distance to all other nodes, thereby minimizing the average network latency.

2. Partially Connected Nodes: When some servers are already on, ENAH focuses on the nodes that are not yet active. It calculates the sum of distances from each inactive node to the nodes that are already active. If a link between an inactive node and an active node is already active, the Euclidean distance is used. If the link is inactive, the distance is multiplied by a penalty factor (in this case, 2) to discourage the use of links that would require activation. The inactive node with the smallest sum of distances to the active nodes is chosen to be activated. In this scenario, the idea is to prioritize nodes that are already connected or closer to connected nodes, taking into account the state of the links to minimize the total path length. Here, the distance is influenced by the state of the node, not just by its smallest sum.

After selecting the server node, the routing of the service flow is carried out efficiently, activating the links between the demand's origin node and the chosen server node by the heuristic. This ensures that the activated nodes minimize latency, considering both the distances between the nodes and the current state of the network links.

V. RESULTS

In this section, we analyze and compare the proposed heuristic with two different approaches: the Random approach, where nodes are activated randomly, and the All-On approach, where all nodes are always on. The comparison will be made in terms of latency and energy consumption. The heuristic was implemented in Python 3.8.17, and the simulations were run on a machine equipped with an Intel® Core(TM) i7-7700 @ 3.60GHz, 16GB of RAM, running Linux Mint. The experiments were repeated 30 times for different workloads, and the average results are presented.

We define service flows as video streaming (VS), web services (WS), voice over IP (VoIP), and online gaming (OG), as shown in Table I. Each service requires a specific set of functions and bandwidth [11].

For the tests, we considered six types of VNFs: Network Address Translation (NAT), Firewall (FW), Traffic Monitor

(TM), WAN Optimization Controller (WOC), Video Optimization Controller (VOC), and Intrusion Detection System (IDS). Different sets of demands were generated and distributed between origin and destination nodes. Each demand was assigned to a service type based on the traffic percentage presented in Table I. We assumed that all VNFs have a processing capacity of 200 Mb/s (Megabits per second) and a processing time of 10 ms [11].

The tests were performed using the DFN topology from the SNDLib with 10 nodes and 45 links [12]. The capacity of each physical link was set to 1 GB/s, and the propagation delay of optical fiber is approximately 10 ms/km [11]. We tested five different scenarios, varying the number of demands per node from 5 to 10, allowing a detailed analysis of the heuristic's behavior under different workloads.

Figure 1 shows the system's energy consumption profile under different server demand scenarios. The ENAH and Random approaches demonstrate almost identical energy consumption at all demand levels, as evidenced by the overlapping shaded areas representing standard deviation. This proximity is due to both approaches using the same number of servers and different amounts of active links. Most energy is consumed by the servers, making the demand distribution almost irrelevant in terms of energy consumed.

On the other hand, the All-On approach, where all servers are kept active, shows significantly higher consumption, highlighting the inefficiency of this approach and the importance of selectively activating servers to optimize energy consumption. For scenarios with 5, 8, and 10 demands per server, ENAH shows savings of approximately 68%, 53%, and 52%, respectively, compared to the All-On approach. Therefore, the ENAH heuristic proves efficient in optimizing energy consumption, providing significant reductions. The results indicate that ENAH is robust and effective, maintaining consistent energy consumption even with varying demands. This reduction in energy consumption aligns with one of the key objectives of the VNF-PR, which is to minimize energy usage while ensuring network performance. Such improvements are crucial for data centers and other environments where energy efficiency is essential, resulting in lower operational costs.

Figure 2 illustrates the system's latency under different server demand scenarios. The results demonstrate the need for a solution to determine which nodes to activate, as the "Random" approach, which activates servers randomly, shows higher latencies, especially in low to medium demand scenarios. For the scenario with 5 demands per node, ENAH presents a reduction of approximately 44% compared to the Random approach. For the scenario with 8 demands per node, ENAH

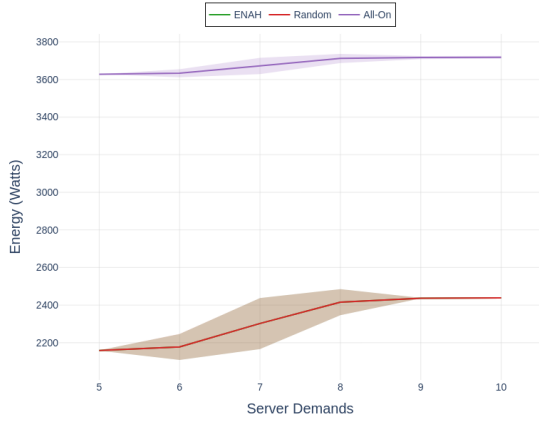


Fig. 1. Energy

presents a reduction of approximately 47% compared to the Random approach. For 10 demands per node, ENAH's latency is reduced by about 51% compared to the Random approach.

The All-On approach shows higher latencies compared to ENAH. For scenarios with 5, 8, and 10 demands per node, ENAH presents reductions of approximately 27%, 29%, and 33%, respectively, compared to the All-On approach. Analyzing the shaded areas in the graph, which represent the standard deviation of measurements, shows that the ENAH heuristic not only reduces average latency but also presents lower variability in results, suggesting greater consistency. This reduction in latency aligns with one of the key objectives of the VNF-PR, which is to minimize system latency while ensuring network performance. This is crucial for time-sensitive applications such as online gaming and VoIP, where perceptible delays can significantly impact user experience.



Fig. 2. Latency

VI. CONCLUSION

In this work, we addressed the VNF-PR problem, focusing on its specific objectives of minimizing energy consumption and system latency while maintaining QoS. We proposed the

ENAH heuristic to optimize server activation in virtualized network environments. ENAH proved to be an effective and robust solution, excelling in both latency reduction and energy consumption. Compared to alternative approaches, the heuristic reduced average latency by up to 51% and energy consumption by up to 68% in high-demand scenarios. The main results indicate that ENAH manages the state of server activation, turning on only the necessary servers, resulting in an optimal allocation of VNFs as a consequence. This balance between energy efficiency and performance is crucial for 6G networks, especially for latency-sensitive applications such as online gaming and VoIP.

For future work, we suggest investigating the implementation of scenarios with servers of different processing capacities and the inclusion of multiple servers per node. Another promising direction is to combine the ENAH heuristic with other solutions specifically focused on demand routing. Additionally, exploring reinforcement learning techniques could enable the creation of a model that learns from network behavior, orchestrating resources more automatically and efficiently.

REFERENCES

- [1] C. R. De Mendoza, B. Bakhshi, E. Zeydan, and J. Mangues-Bafalluy, "Near optimal vnf placement in edge-enabled 6g networks," in *2022 25th Conference on Innovation in Clouds, Internet and Networks (ICIN)*. IEEE, 2022, pp. 136–140.
- [2] A. Mouaci, É. Gourdin, I. Ljubić, and N. Perrot, "Virtual network functions placement and routing problem: Path formulation," in *2020 IFIP Networking Conference (Networking)*. IEEE, 2020, pp. 55–63.
- [3] J. Li, X. Qi, J. Li, Z. Su, Y. Su, and L. Liu, "Fault diagnosis in the network function virtualization: A survey, taxonomy and future directions," *IEEE Internet of Things Journal*, 2024.
- [4] A. Mouaci, É. Gourdin, I. Ljubić, and N. Perrot, "Two extended formulations for the virtual network function placement and routing problem," *Networks*, 2023.
- [5] R. Moosavi, S. Parsaeefard, M. A. Maddah-Ali, V. Shah-Mansouri, B. H. Khalaj, and M. Bennis, "Energy efficiency through joint routing and function placement in different modes of sdn/nfv networks," *Computer Networks*, vol. 200, p. 108492, 2021.
- [6] X. Shen, J. Gao, W. Wu, M. Li, C. Zhou, and W. Zhuang, "Holistic network virtualization and pervasive network intelligence for 6g," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 1–30, 2021.
- [7] J. Ortin, P. Serrano, J. Garcia-Reinoso, and A. Banchs, "Analysis of scaling policies for nfv providing 5g/6g reliability levels with fallible servers," *IEEE Transactions on Network and Service Management*, vol. 19, no. 2, pp. 1287–1305, 2022.
- [8] H. Cao, J. Du, H. Zhao, D. X. Luo, N. Kumar, L. Yang, and F. R. Yu, "Resource-ability assisted service function chain embedding and scheduling for 6g networks with virtualization," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3846–3859, 2021.
- [9] W. Wu, Y. Li, L. Chen, B. Zhang, W. Wang, Y. Zhao, and J. Zhang, "Service function chain mapping based on joint load balancing in computing power network," in *2023 Opto-Electronics and Communications Conference (OECC)*. IEEE, 2023, pp. 1–4.
- [10] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," *Future generation computer systems*, vol. 28, no. 5, pp. 755–768, 2012.
- [11] A. Varasteh, M. De Andrade, C. M. Machuca, L. Wosinska, and W. Kellerer, "Power-aware virtual network function placement and routing using an abstraction technique," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.
- [12] Zuse Institute Berlin, "Sndlib: The sndlib network optimization library," <http://sndlib.zib.de/home.action>, 2024, accessed: 2024-05-24.