



GOIÁS FEDERAL UNIVERSITY
INSTITUTE OF INFORMATICS
POSTGRADUATE PROGRAM IN COMPUTER SCIENCE

XAVIER PAULINO SEBASTIÃO

Federated Analytics Performance Optimization for Next Generation Networks (B5G/6G)

**(Otimização de Desempenho de Análise Federada para
Redes de Próxima Geração (B5G/6G))**

Goiânia
2025



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Xavier Paulino Sebastião

3. Título do trabalho

Federated Analytics Performance Optimization for Next Generation Networks (B5G/6G)

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Antonio Carlos De Oliveira Junior, Professor do Magistério Superior**, em 01/04/2025, às 14:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Xavier Paulino Sebastião, Discente**, em 01/04/2025, às 17:15, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5279235** e o código CRC **33E26625**.

Referência: Processo nº 23070.007882/2025-72

SEI nº 5279235

XAVIER PAULINO SEBASTIÃO

Federated Analytics Performance Optimization for Next Generation Networks (B5G/6G)

(Otimização de Desempenho de Análise Federada para
Redes de Próxima Geração (B5G/6G))

Thesis presented to the Postgraduate Program in Computer Science, at the Institute of Informatics, Goiás Federal University (UFG), as a requirement for obtaining the title of Master in Computer Science.

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás, como requisito para obtenção do título de Mestre em Ciência da Computação.

Concentration Area: Computer Science

Área de Concentração: Ciência da Computação

Research Area: Computer Systems

Linha de Pesquisa: Sistemas de Computação

Supervisor/Orientador: Prof. Antonio Carlos de Oliveira Junior

Cosupervisor/Coorientadora: Prof. Maria do Rosário Campos Ribeiro

Goiânia
2025

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Sebastião, Xavier Paulino

Federated Analytics Performance Optimization for Next Generation
Networks (B5G/6G) [manuscrito] / Xavier Paulino Sebastião. - 2025.
LXXXIV, 84 f.: il.

Orientador: Prof. Dr. Antonio Carlos de Oliveira Junior; co
orientadora Dra. Maria do Rosário Campos Ribeiro.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto
de Informática (INF), Programa de Pós-Graduação em Ciência da
Computação, Goiânia, 2025.

Bibliografia. Anexos.

Inclui siglas, fotografias, abreviaturas, símbolos, gráfico, tabelas,
algoritmos, lista de figuras, lista de tabelas.

1. Optimization. 2. Federated Analytics. 3. NS-3 5G-LENA. 4.
Integration. 5. B5G/6G. I. Junior, Antonio Carlos de Oliveira, orient. II.
Título.

CDU 004



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE INFORMÁTICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **01/2025** da sessão de Defesa de Dissertação de **Xavier Paulino Sebastião**, que confere o título de Mestre em Ciência da Computação, na área de concentração em Ciência da Computação.

Aos vinte dias do mês de fevereiro de dois mil e vinte e cinco, a partir das oito horas, via webconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada “**Federated Analytics Performance Optimization for Next Generation Networks (B5G/6G)**”. Os trabalhos foram instalados pelo Orientador, Professor Doutor Antonio Carlos de Oliveira Junior (INF/UFG) com a participação dos demais membros da Banca Examinadora: Professora Doutora Maria do Rosario Campos Ribeiro (INF/UFG), coorientadora; Professor Doutor Waldir Aranha Moreira Junior (Fraunhofer Portugal AICOS), membro titular externo; Professor Doutor Victor Hugo Lázaro Lopes (IFG), membro titular interno. A realização da banca ocorreu por meio de videoconferência. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor Antonio Carlos de Oliveira Junior, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos vinte dias do mês de fevereiro de dois mil e vinte e cinco.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Antonio Carlos De Oliveira Junior, Professor do Magistério Superior**, em 20/02/2025, às 10:00, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Waldir Aranha Moreira Junior, Usuário Externo**, em 20/02/2025, às 10:00, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Do Rosario Campos Ribeiro, Professor do Magistério Superior-Substituto**, em 20/02/2025, às 10:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Victor Hugo Lázaro Lopes, Usuário Externo**, em 20/02/2025, às 10:14, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Xavier Paulino Sebastião, Discente**, em 20/02/2025, às 10:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5167720** e o código CRC **97735C4D**.

Referência: Processo nº 23070.007882/2025-72

SEI nº 5167720

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Xavier Paulino Sebastião

Graduated in Electronic Science and Communication from the University of Dar es Salaam, in Tanzania. During undergraduate studies, was a scholarship recipient of the Mozambique-Tanzania international mobility exchange program and conducted research in automation, developing an automated system for optimizing water and electricity usage for households and public spaces. In the master's program, is a scholarship recipient of the GCUB international mobility program and co-authored an article published for SBrT 2024, titled "Soluções baseadas em aprendizado por reforço profundo para implantar VANTs como gateways LoRaWAN com foco na Qualidade de Serviço IoT". Additionally, has an article accepted for the Brazilian Symposium on Computer Networks and Distributed Systems (SBRC 2025), titled "Channel-Aware Federated Analytics in B5G/6G Networks: Dynamic Power Allocation with NS-3 5G-LENA". Currently, is a university teacher at Universidade Zambeze, in Mozambique.

Acknowledgement

I thank Almighty God for granting me good health and strength throughout the period in which I worked on this research.

I am thankful to Prof. Antonio Carlos de Oliveira Junior, my supervisor, and Prof. Maria do Rosário Campos Ribeiro, my cosupervisor, for helping me with indispensable contributions whenever I needed consultation, orientation, and guidance.

I also thank my family for their support and understanding throughout this journey.

Lastly, I express my gratitude to all those who gave me assistance and the possibility to complete my research, the Institute of Informatics, the Goiás Federal University in general, my classmates, and all people who contributed direct or indirectly to the success of this work.

Abstract

Sebastião, Xavier Paulino. **Federated Analytics Performance Optimization for Next Generation Networks (B5G/6G)**. Goiânia, 2025. 83p. MSc. Thesis. Institute of Informatics, Goiás Federal University.

The rapid growth of interconnected devices across various sectors in the world has led to the generation of large volumes of data. Depending on its nature and specific needs, a significant part of this data is handled and analyzed using data science techniques to support decision-making. Alongside these advancements, as institutions rely more on data-driven systems, they also face increasing threats and security challenges that compromise the privacy of their clients or collaborators, consequently damaging their reputation. Federated Analytics (FA) is an innovative approach for preserving data security and privacy by implementing collaborative analysis of data from distributed devices without sharing raw data. However, in cases where FA operates over wireless transmission, challenges such as interference, signal degradation, and network congestion may arise. These factors can make the wireless transmission unreliable, introducing delays and causing corruption in responses and updates received at the central server, thereby compromising the quality of the final aggregated FA results. This work proposes an integrated framework for simulating FA in real 5G network conditions. The framework applies two algorithms: channel-aware power allocation algorithm to efficiently allocate transmission power for user equipments (UEs) based on distance and channel conditions, and synchronous FA-5GLENA integrated algorithm to integrate the FA with NS-3 5G-LENA and aggregate results for optimized performance within 5G network conditions. To evaluate the impact of the network on FA, three scenarios were compared: (1) uniform maximum power allocation for all UEs, (2) random power allocation, and (3) channel-aware power allocation algorithm. The simulation results show that the channel-aware algorithm outperforms the uniform and random power allocation scenarios on both the network and FA operation. On FA, the algorithm achieved statistically higher accuracy (93.17 %), precision (93.31 %), and recall (93.09 %) compared to uniform allocation (accuracy: 55.96 %, precision: 56.02 %, recall: 55.90 %) and random allocation (accuracy: 42 %, precision: 42.02 %, recall: 41.96 %), highlighting the superiority of the algorithm in enhancing FA performance.

Keywords

Optimization, Federated Analytics, NS-3 5G-LENA, Integration, B5G/6G

Resumo

Sebastião, Xavier Paulino. **Federated Analytics Performance Optimization for Next Generation Networks (B5G/6G)**. Goiânia, 2025. 83p. Dissertação de Mestrado. Institute of Informatics, Goiás Federal University.

O rápido crescimento de dispositivos interconectados em diversos setores no mundo tem levado à geração de grandes volumes de dados. Dependendo de sua natureza e necessidades específicas, uma parte significativa desses dados é processada e analisada utilizando técnicas de ciência de dados para apoiar a tomada de decisões. Paralelamente a esses avanços, à medida que as instituições dependem cada vez mais de sistemas baseados em dados, elas também enfrentam ameaças e desafios de segurança crescentes que comprometem a privacidade de seus clientes ou colaboradores, prejudicando, conseqüentemente, sua reputação. A Análise Federada (FA) é uma abordagem inovadora para preservar a segurança e a privacidade dos dados, implementando a análise colaborativa de informações provenientes de dispositivos distribuídos sem a necessidade de compartilhar os dados brutos. No entanto, quando a FA opera por meio de transmissões sem fio, podem surgir desafios como interferência, degradação do sinal e congestionamento da rede. Esses fatores podem tornar a transmissão sem fio instável, introduzindo atrasos e causando corrupção nas respostas e atualizações recebidas no servidor central, comprometendo assim a qualidade dos resultados agregados da FA. Este trabalho propõe um framework integrado para simulação da FA em condições reais de redes 5G. O framework aplica dois algoritmos: algoritmo de alocação de potência ciente do canal para alocar eficientemente a potência de transmissão aos equipamentos de usuário (User Equipments - UEs) com base na distância e nas condições do canal, e algoritmo FA-5GLENA síncrono para integrar a FA ao NS-3 5G-LENA e agregar os resultados para otimizar o desempenho dentro das condições da rede 5G. Para avaliar o impacto da rede na FA, três cenários foram comparados: (1) alocação uniforme da potência máxima para todos os UEs, (2) alocação aleatória de potência e (3) algoritmo de alocação de potência ciente do canal. Os resultados da simulação mostram que o algoritmo ciente do canal supera os cenários de alocação uniforme e aleatória de potência tanto na operação da rede quanto na FA. Na FA, o algoritmo obteve uma acurácia estatisticamente maior (93,17 %), precisão (93,31 %) e recall (93,09 %) em comparação com a alocação uniforme (acurácia: 55,96 %, precisão: 56,02 %, recall: 55,90 %) e alocação aleatória (acurácia: 42 %, precisão: 42,02 %, recall: 41,96 %), destacando a superioridade do algoritmo na melhoria do desempenho da FA.

Palavras-chave

Otimização, Análise Federada, NS-3 5G-LENA, Integração, B5G/6G

Contents

| | |
|---|----|
| List of Figures | 13 |
| List of Tables | 14 |
| List of Algorithms | 15 |
| List of Acronyms | 17 |
| 1 Introduction | 18 |
| 1.1 Motivation | 20 |
| 1.2 Problem Definition | 22 |
| 1.3 Objectives | 24 |
| 1.3.1 Main | 24 |
| 1.3.2 Specific | 24 |
| 1.4 Contribution | 24 |
| 1.5 Work Done | 25 |
| 1.6 Master's Thesis Organization | 26 |
| 2 Fundamental Concepts and Related Works | 27 |
| 2.1 Federated Analytics (FA) | 27 |
| 2.2 Benefits of FA Compared to Centralized Data Analytics and FL | 28 |
| 2.2.1 FA System Model | 29 |
| 2.2.2 FA Application Domain | 29 |
| 2.2.3 Network Model and Role of FA in 5G and Beyond | 30 |
| 2.2.4 Applications of FA and Query Types | 32 |
| 2.3 Network Simulator 3 (NS-3) | 32 |
| 2.4 The 5G-LENA NR | 33 |
| 2.5 The 5th Generation Wireless Communication Networks (5G) | 35 |
| 2.6 Beyond 5th or 6th Generation Wireless Communication Networks (B5G/6G) | 35 |
| 2.7 Wireless Links, Channels and Characteristics | 37 |
| 2.8 Propagation Models | 38 |
| 2.9 Energy Model | 40 |
| 2.10 Performance Measurement: Delay, Latency, Throughput and Loss | 41 |
| 2.11 Signal-to-Noise Ratio and Bit Error Rate | 44 |
| 2.12 Bayesian Aggregation | 45 |
| 2.13 Related Works | 47 |
| 2.13.1 Overview of Related Works | 47 |
| 2.13.2 Comparison of this Proposal with Related Works | 50 |

| | | |
|-------|--|-----------|
| 3 | Optimized FA Algorithm Proposal for 5G Network | 53 |
| 3.1 | Proposed Solution | 53 |
| 3.2 | Objective Function for Power Allocation by Channel-Aware Algorithm | 54 |
| 3.3 | Implementation of FA-5GLENA Integrated Framework | 55 |
| 3.4 | Synchronous FA-5GLENA Integrated Algorithm | 57 |
| 3.5 | The Channel-Aware Transmission Power Allocation Algorithm | 59 |
| 4 | Performance Evaluation and Results | 60 |
| 4.1 | Simulation Environment | 60 |
| 4.1.1 | NS-3 and 5G-LENA Network Configuration | 61 |
| 4.1.2 | Simulation parameters | 62 |
| 4.1.3 | FA Configuration | 63 |
| 4.2 | Simulation Results and Validation | 64 |
| 4.2.1 | Discussion of Results | 64 |
| 4.2.2 | Statistical Analysis | 71 |
| 5 | Conclusion and Future Research Works | 76 |
| 5.1 | Conclusion | 76 |
| 5.2 | Future Research Works | 77 |
| | References | 79 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | Basic FA setup. Extracted from [Wang et al. 2024] | 28 |
| 2.2 | FA setup for 6G. Extracted from [Parra-Ullauri et al. 2024] | 31 |
| 2.3 | NS-3 architecture. Extracted from [Ceballos et al. 2021] | 33 |
| 2.4 | 5G-LENA Class architecture. Extracted from [Larrañaga et al. 2023] | 33 |
| 2.5 | SNR, transmission rate, and Bit error rate relationship. Extracted from [Kurose e Ross 2021] | 45 |
| 3.1 | FA Integration with 5G-LENA Module | 57 |
| 4.1 | gNB and devices | 62 |
| 4.2 | FA performance metrics for different power scenarios. Accuracy, Precision and Recall. | 66 |
| 4.3 | Network performance metrics for different power scenarios. Delivery ratio, Throughput, Latency, Packet loss and Energy consumed. | 68 |
| 4.4 | Network performance metrics in terms of distance. Delivery ratio, Throughput, Latency, Packet loss and Energy consumed. | 70 |

List of Tables

| | | |
|------|---|----|
| 2.1 | Differences between 5G and 6G. Extracted from [Kadir et al. 2021] | 36 |
| 2.2 | Comparison of this proposal and the related works | 51 |
| 4.1 | Simulation parameters | 63 |
| 4.2 | Client Selection and Prediction Results | 65 |
| 4.3 | Statistical Significance for Accuracy | 72 |
| 4.4 | Statistical Significance for Precision | 72 |
| 4.5 | Statistical Significance for Recall | 73 |
| 4.6 | Statistical Significance for Delivery Ratio | 74 |
| 4.7 | Statistical Significance for Throughput | 74 |
| 4.8 | Statistical Significance for Latency | 74 |
| 4.9 | Statistical Significance for Packet Loss | 74 |
| 4.10 | Statistical Significance for Energy Consumed | 74 |

List of Algorithms

| | | |
|---|---|----|
| 1 | Synchronous FA - 5GLENA Integrated Algorithm | 58 |
| 2 | Channel-Aware Transmission Power Allocation Algorithm | 59 |

List of Acronyms

- 3GPP** - 3rd Generation Partnership Project
- 3GPP TR** - Third Generation Partnership Project Technical Report
- 5G-LENA** - 5th Generation - LTE/EPC Network Analyzer
- BPSK** - Binary Phase Shift Keying
- CPU** - Central Processing Unit
- EPC** - Evolved Packet Core
- FDD** - Frequency Division Duplex
- GPLv2** - GNU General Public License Version 2
- IP** - Internet Protocol
- ITU-R/IMT** - International Telecommunication Union Radiocommunication Sector's
International Mobile Telecommunications
- LAN** - Local Area Network
- LDPC** - Low-Density Parity Check
- LoRa** - Long Range
- LoRaWan** - Long Range Wide Area Network
- MAC** - Media Access Control
- MIMO** - Multiple Input Multiple Output
- MNIST** - Modified National Institute of Standards and Technology
- NIC** - Network Interface Card
- OFDMA** - Orthogonal Frequency Division Multiplexing Access
- PGW** - Packet Data Network Gateway
- QAM** - Quadrature Amplitude Modulation
- QPSK** - Quadrature Phase Shift Keying
- RAM** - Random Access Memory
- RAN** - Radio Access Network
- SGW** - Serving Gateway
- TCP** - Transport Control Protocol
- TDD** - Time Division Duplex
- TDMA** - Time Division Multiplexing Access
- TTI** - Transmission Time Interval

UAV (VANT in Portuguese) - Unmanned aerial vehicles

UDP - User Datagram Protocol

Wi-Fi - Wireless Fidelity

Introduction

The Global interconnectivity and the expansion of the internet have changed how people and organizations interact and do business [Maurseth 2018]. The advancement of digital technology, communication platforms such as email, social networks, and messaging apps has facilitated people from different parts of the world to communicate at low cost [Beksultanova, Vatyukova e Yalmaeva 2020]. For example, by benefiting from digital technologies, individuals and organizations conduct business and hold virtual meetings with collaborators and clients. This has facilitated remote work, collaborative decision-making and the promotion of entrepreneurship.

In the academic field, it has opened opportunities for educational and research institutions and platforms to offer courses and educational resources online, making them accessible to people worldwide and promoting access to knowledge [Beksultanova, Vatyukova e Yalmaeva 2020]. In health, has opened a way to use telemedicine services that allow remote consultations and monitoring between patients and healthcare providers, in that way improving access to healthcare services [Ghafar 2024]. In agriculture, interconnected devices and sensors have fortified real-time monitoring of soil conditions, weather patterns, crop health, and insights into crop yields, pest infestations, and environmental impacts [Oliveira-Jr et al. 2020]. These applications tell just a few of the potential benefits of digital technologies in various fields.

The examples above illustrate the need to create conditions for future generations to further integrate networks and expand capabilities from different perspectives. One such need is data security and privacy. In these communications, which often involve the exchange of information, challenges arise that require a responsible approach. From privacy invasion to data misuse, constantly threaten secure information management [Yan et al. 2023]. Another need is data handling. As the growing number of connected devices in various industries generates more data than ever before, there is a rising demand for advanced data science techniques to extract valuable insights for better decision-making while ensuring efficient management and storage. Therefore, centralizing this data for analysis is not always feasible due to privacy, security, and communication efficiency consideration [Elkordy et al. 2023]. In this context, Federated Analytics (FA) emerges as

an innovative approach for preserving data security and privacy, that enables collaborative analysis and drawing of insights from distributed data owned by multiple parties, such as mobile devices, edge servers, or institutional organizations [Chen et al. 2021]. Instead of directly sharing raw data, only intermediate query responses are transmitted to the central entity, such as a service provider [Elkordy et al. 2023] [Luo et al. 2018].

This approach not only meets stringent data privacy and regulatory requirements by limiting information exposure beyond the intended insights, but also offers an efficient way to address the energy, latency, and costs associated with transferring large volumes of data. By optimizing data handling and transmission, it maximizes the utility of the data while minimizing the risks associated with its transfer, ultimately improving communication efficiency as only necessary responses are sent to the central server.

From a general perspective, FA can be defined as a scenario for data analysis where a questioner wants to respond to a data analysis query through the collaboration of multiple proprietary data devices that have their local raw data. Raw data is not exchanged or transmitted, but instead, intermediate query responses intended for aggregation at the questioner are transferred to answer the intended query. Google introduced the concept of FA in 2020 as a way to represent “collaborative data science without data collection” [Elkordy et al. 2023]. Initially, it was used in support of Federated Learning (FL) as a way of Google engineers to evaluate the quality of learned machine learning models (ML) against real-world data. Since then, its implementations have expanded to other applications. One notable use case is the identification of popular elements across distributed devices, such as frequently unrecognized vocabulary, or most popular songs recognized by mobile phones [Zhu et al. 2020].

FA shares some similarities with FL [Kairouz et al. 2021]. Both involve collaborating with data distributed across multiple sources (like devices or servers). They aim to maintain data privacy by keeping raw data local, allowing analysis or learning without exposing sensitive information. Both are decentralized, with data processing or model training happening locally on individual devices or servers instead of in a central location. But, as stated by [Kairouz et al. 2021], the primary goal of FL is to train machine learning models across distributed devices while preserving data privacy. This involves optimizing model performance using data from various sources without centralizing it. In contrast, FA is focused on data analysis and generating insights across distributed datasets [Elkordy et al. 2023]. It covers a broader range of analytical techniques beyond just machine learning.

1.1 Motivation

As technology advances rapidly in the digital era, concerns about data security and privacy increase [Maurseth 2018]. Although global interconnectivity and the expansion of the internet have modernized the world, they have also brought challenges that need to be researched and resolved, from privacy invasion to data misuse, which constantly risk secure information management [Yan et al. 2023].

The article by [Oyadomari, Costa e Ribeiro 2023], puts the government of Brazil on the list of organizations managing large volumes of personal data from citizens as part of its regular activities, such as security, taxation, and the provision of public services in the country. This data often includes sensitive information from various sources. This study shows that 40% of internet users are very concerned, and 29% are concerned about the government's use of their data, with slightly fewer concerns regarding data use by private sector entities. The article also emphasizes that biometric data, such as fingerprints and facial recognition, raises the highest level of concern, followed by health data. In this situation, protecting personal and sensitive information against cyber threats becomes an urgent priority.

The government of Brazil, in its effort, created the the legal instrument called General Data Protection Law (LGPD in Portuguese), established by Law No. 13.709/2018, to address the growing challenges of personal data processing in an increasingly digital world [Santos 2021]. The law emphasizes the importance of data privacy and security, particularly in sensitive sectors such as academia and research, where vast amounts of data are handled and analyzed. The LGPD requires that data processing be carried out with respect to principles like purpose, necessity, and transparency (Art. 6), ensuring that personal data is used solely for its intended purpose and giving individuals control over their information. Additionally, it grants individuals rights such as access, correction, deletion, and data portability (Art. 18), thereby placing greater responsibility on institutions to manage and protect personal data. By imposing strict regulations on data handling, the LGPD aims to prevent data misuse and improve privacy protection. This legal framework has created a need for innovative approaches to data analysis that can uphold privacy and security while managing the challenges of processing large datasets.

With the implementation of the LGPD, individuals gained significant control over their personal data [Miragem 2019]. It allows them to request clear and transparent information about how their data is collected, processed, and used. The law recognizes their right to correct inaccurate data and revoke consent for data processing. This reinforces individual autonomy over personal information. For organizations, compliance with the LGPD alongside the Consumer Protection Code, demands the adoption of robust security measures to prevent unauthorized access and improper data handling

[Miragem 2019]. This includes developing transparent privacy policies, conducting staff training, and appointing a Data Protection Officer (DPO) responsible for guiding compliance efforts. These regulatory requirements highlight the necessity of secure and efficient data management strategies, further reinforcing the importance of solutions that balance data utility with privacy protection in modern digital environments.

One strategy for protecting sensitive data is FA. It is an innovative approach that allows collaborative analysis and the discovery of insights from data distributed across multiple sources, such as mobile devices, edge servers, or organizations [Elkordy et al. 2023] [Luo et al. 2018]. Instead of directly sharing raw data, only intermediate query responses are transmitted to the central entity, such as a service provider for aggregation. FA addresses privacy issues by limiting the exposure of information beyond selected contents. This approach also helps improve energy efficiency and communication, as only necessary responses are transmitted to the central server. Thus, by mitigating risks associated with data handling and transfer, FA aligns directly with the principles of the LGPD, making it a viable alternative for applications that require security, privacy, and regulatory compliance.

For example, imagine a researcher in Brazil working with medical data from different hospitals and clinics. Consider the aim is to predict and diagnose pathologies such as breast cancer or cardiovascular diseases, but the data is highly sensitive. Given Brazil's General Data Protection Law (LGPD), sharing personal patient data across institutions is restricted, even if the data is anonymized. However, it would be invaluable to combine insights from different institutions to improve predictions, compare patterns, and accelerate research. Assume a hospital in São Paulo is training a model to detect breast cancer from mammogram images. Another hospital in Goiânia has its own data from breast cancer patients but cannot share these images directly due to legal concerns (LGPD). With FA, both hospitals can train their models locally, making predictions and classifications on their own datasets. After training, the hospitals can perform local predictions on their data, and then exchange only the intermediate results such as predictions or classifications for aggregation. In this way, FA can enable researchers and medical institutions to collaborate effectively without violating data protection regulations, like LGPD. This method ensures that sensitive patient data never leaves the local institution, while still allowing the aggregation of valuable insights. Therefore, federated analytics not only facilitates collaboration between different research institutions without compromising personal data privacy but also ensures that in healthcare, data can be used to predict and treat diseases more effectively. This practical example illustrates how federated analytics can be a powerful tool to protect personal data while driving significant advances in research and the treatment of pathologies.

However, when FA operates over wireless transmission, challenges can occur,

including interference, signal degradation, and network congestion, which can make the transmission media unreliable. Due to this unreliability nature, responses and updates received at the central server may be subject to delays or contain incorrect information, which can affect the quality of the final aggregated response. As pointed by [Wang et al. 2024], one of the major challenges in FA is the reliance on wireless communication, which causes delays and inconsistencies in data aggregation. Since FA operates in a distributed and dynamic environment, participants communicate over varying network conditions. In this situation, the slowest participant in the system can degrade significantly performance of the whole system, a problem known as the “straggler effect”. Still, according to [Wang et al. 2024], this issue gets worse when there are unstable data transmission rates and unreliable communication channels, leading to delays that impact the efficiency and accuracy of FA. Therefore, they recommend that addressing these challenges is a key to ensuring the scalability and effectiveness of FA in real-world applications. For this reason, we are motivated to explore techniques of optimizing network communication for FA within the context of 5G/6G networks. By addressing the challenges presented by unreliable wireless transmission, our goal is to boost the overall performance and reliability of FA in distributed environments. In this context, we aim to optimize the communication processes in FA system, to ensure that responses and updates are transmitted accurately, efficiently and timely within the system. To achieve this, we are conducting experiments using NS-3 5G-LENA for 5G network simulation in combination with TensorFlow framework for FA machine learning.

1.2 Problem Definition

The growth of interconnected devices and services across industries has led to the generation of unprecedented volumes of data, driving an increasing demand for advanced data science techniques to extract valuable insights to support decision-making processes [Elkordy et al. 2023]. Currently, the dominant approach in most data analysis involves methods of collecting and processing. In this model, the raw data generated by devices is collected and sent to a central server for processing and analysis [Luo et al. 2018]. However, centralizing such a vast volume of data for analysis is not always feasible due to privacy, security, and communication efficiency considerations.

Inefficient communication can cause delays in transmitting responses and updates. Consequently, it increases response times, and slows down the training process and the overall performance of FA [Zhao et al. 2022]. For example, slow and unreliable connections between clients and the coordinating server can prolong model training, increase waiting times, and make the system less responsive. This can affect the quality of machine learning models and the speed at which queries and responses are sent [Lin et al. 2020].

Poor communication can excessively consume network bandwidth and quickly drain device batteries. These issues compromise performance and limit the scalability of FA.

Furthermore, a communication environment involving data exchange requires secure and efficient algorithms to ensure sensitive data remains protected during transmission between clients and the central server [Han et al. 2022]. Poor communication can increase the risk of data leaks, threatening sensitive details being analyzed. Additionally, centralized data is susceptible to privacy attacks, as all information is stored in a single location [Zhao et al. 2022]. In such cases, a single attack can expose a large volume of personal or confidential information to unauthorized access.

FA takes part in the systems being implemented to address these challenges. These systems provide decentralized methods for data analysis, enabling collaboration among distributed datasets while preserving data privacy. Despite this advantage, FA faces significant challenges when implemented in wireless networks. Limited resources and unstable connections can result in significant delays in data transmission and occasionally packet loss [Chen et al. 2021]. In this context, FA performance is highly sensitive to variations in network conditions such as latency, packet loss, and resource limitations, all of which are inherent to the unreliable nature of wireless transmission channels [Zhao et al. 2022]. As a result, responses and updates received at the central server may experience delays or corruption, which can affect the quality of the final aggregated response. For this reason, this research focuses on studying real parameters of 5G networks to optimize the performance of FA.

When data updates and responses travel to the main server, they can get delayed or corrupted due to network congestion or weak signals, interference, and other factors. These problems affect how quickly and accurately FA can give final aggregated results, which is necessary for making informed decisions and conducting data analysis effectively. To solve these challenges, research is required to study real-world networks closely. The studies can include looking at how fast data travels, how much data the network can handle, and how reliable the connection is. Our research aims to develop strategies that make FA work better in networks projected for B5G/6G.

Optimizing FA in networks isn't just about theory, it involves practical experiments with different network setups. This study tries to find the best network configurations that reduce delays, data loss, energy consumption, and improve throughput while keeping smooth and reliable communication between the server and clients. This approach is important for making FA systems more robust and efficient. By adapting FA algorithms to match how real 5G networks behave, we hope to improve how quickly and accurately data insights are delivered, benefiting users across various fields.

1.3 Objectives

1.3.1 Main

The main objective of this work is to evaluate the impact of the network on FA performance and optimize FA in the context of 5G networks. To achieve this, simulations are conducted to evaluate how network conditions affect the quality of the FA final aggregated responses.

1.3.2 Specific

- Research how factors like latency, packet loss, throughput, and energy consumption impact the performance of FA across networks;
- Evaluate various 5G network scenarios using simulations conducted with NS-3 5G-LENA to understand their effectiveness in improving FA performance;
- Research and develop algorithms for the efficient allocation and utilization of network resources to optimize FA, considering device distance and channel conditions in 5G networks;
- Validate the results experimentally and by literature, and explore applications of FA in fields such as health, smart cities, agriculture, education and IoT, among others, for B5G/6G networks.

1.4 Contribution

This work emphasizes the promoting FA as a critical component in the transition to B5G/6G networks. The research aims to optimize FA under specific conditions in current 5G networks, with a particular focus on effectively managing transmission power to maximize performance in future networks. We propose an integrated framework and specific algorithms that enable realistic FA simulation, capturing contemporary and emerging network conditions. This systematic approach allows for a detailed assessment of the impact of changing network conditions on FA performance, using the 5G-LENA simulator to create a less complex network environment, development of FA algorithms and integration with 5G-LENA. Finally, we validate the results through an application in the domain of collaborative quality assessment of trained ML models on classification problems, ensuring the relevance and applicability of the developed solutions. The main contributions of this work are as follows:

- **Network Adaptive FA Algorithm:** Implementation of network and FA algorithms specifically adapted to the characteristics of 5G network environments.

- **FA with NS-3 Simulator:** Development of a framework that integrates FA with NS-3 5G-LENA simulator. This facilitates realistic and scalable simulations of FA under diverse network conditions.
- **Performance Evaluation:** Validation of FA-5GLENA integrated algorithm through an application in the domain of collaborative quality assessment of trained machine learning models on classification problems, specifically evaluating the impact of latency, packet loss, throughput and energy consumption on FA accuracy, precision and recall.
- **Validation and Documentation:** Detailed analysis evaluating how various network conditions, such as latency, packet loss, throughput, and energy consumption affect FA's accuracy, precision, and recall, demonstrating their effectiveness in addressing challenges related to 5G networks.
- Detailed documentation for implementing FA in 5G networks, serving as a reference for future researchers and developers in B5G/6G networks.
- **Source Code Sharing:** The source code ¹ has been made accessible to enable replication and validation of results, also serving as a valuable resource for future researchers and developers in the FA domain for B5G/6G networks. Available at <https://github.com/LABORA-INF-UFG/fa-5glena>.

1.5 Work Done

Various research areas have been explored during the studies, which helped shape the direction of the current work. The study began with a comprehensive literature review on IoT, B5G/6G networks, FL and FA, and key tools such as NS-3, TensorFlow, PyTorch, Q-Learning, and Deep Q-Learning. This review provided a theoretical foundation for further research and experimentation. Building on this knowledge, many experiments were conducted in reinforcement learning, focusing on Q-Learning and Deep Q-Learning, to evaluate how these techniques can be integrated and utilized to enhance network performance. Additionally, FL and FA extensive experiments were also explored with NS-3 integration to assess the feasibility of these integrations and evaluate how network conditions influence their performance. These works directly contributed to the development of the current project, providing insights into how network environments shape and affect the efficiency of distributed computational techniques. These efforts led to the following outcomes:

¹ Available at <https://github.com/LABORA-INF-UFG/fa-5glena>

- An article accepted for the Brazilian Symposium on Computer Networks and Distributed Systems (SBRC 2025) titled "Channel-Aware Federated Analytics in B5G/6G Networks: Dynamic Power Allocation with NS-3 5G-LENA". This project involves the development of an integrated framework for simulating Federated Analytics (FA) in real 5G network conditions. The framework utilizes two algorithms: a channel-aware power allocation algorithm designed to efficiently allocate transmission power to user equipment (UEs) based on distance and channel conditions, and a synchronous FA-5GLENA integrated algorithm that combines FA with the NS-3 5G-LENA simulator to aggregate results and optimize performance within 5G network environments.
- Coauthored the article titled "Soluções baseadas em aprendizado por reforço profundo para implantar VANTs como gateways LoRaWAN com foco na Qualidade de Serviço IoT" (<http://dx.doi.org/10.14209/sbrt.2024.1571036460>), published for the XLII Brazilian Symposium on Telecommunications and Signal Processing (SBRT 2024). This work connected the slicing of 3GPP networks with Quality of Service (QoS) parameters from LoRa non-3GPP technology to ensure adequate performance for IoT devices. It integrated various technologies like Mixed Integer Linear Programming (MILP), Markov Decision Process (MDP), Deep Q-Network (DQN), Actor-Critic (A2C) and Network Simulator 3 (NS-3) for on-line UAV positioning. Results obtained demonstrate significant QoS improvements compared to other state-of-the-art solutions.

1.6 Master's Thesis Organization

The thesis is organized as follows: In Chapter 1, we introduce the project. Then, we illustrate fundamental concepts and related works in Chapter 2. In Chapter 3, we outline the master's thesis proposal (Optimized FA Algorithm Proposal for Networks), while Chapter 4 covers performance evaluation and preliminary results. Then, we present the work plan and next steps in Chapter 5. Finally, we conclude the thesis in Chapter 6.

Fundamental Concepts and Related Works

The first part of this chapter introduces fundamental concepts that form the theoretical foundation of this work, describing practical and architectural aspects of various technologies used in the implementation. The second part discusses related works to understand the current state and identify opportunities for new contributions in the field of FA.

2.1 Federated Analytics (FA)

From a general point of view, FA can be understood as a scenario for data analysis where a questioner wants to respond to a data analysis query through the collaboration of multiple proprietary data devices that have their local raw data. In this system, raw data is not exchanged or transmitted, but instead, intermediate query responses intended for aggregation at the questioner are transferred to answer the intended query. The term FA was first coined by Google in 2020 to represent “collaborative data science without data collection” [Wang et al. 2022]. It was first explored in support of Federated Learning (FL) as a way for Google engineers to evaluate the quality of learned Machine Learning (ML) models against real-world data [Elkordy et al. 2023]. Since then, FA use cases have spanned a variety of domains that significantly benefit from the ability to perform data analysis without compromising the privacy of participants [Elkordy et al. 2023]. In technology, it allows to evaluate the accuracy of prediction models like those from Gboard, using anonymous typing data. In healthcare, it facilitates diagnosis and research with large volumes of data, without exposing sensitive patient information, supporting studies on treatments and symptoms. In marketing, FA helps advertisers understand the effectiveness of their ads through anonymous viewing data, optimizing advertising strategies without compromising users’ privacy. Figure 2.1 shows the basic setup of FA.

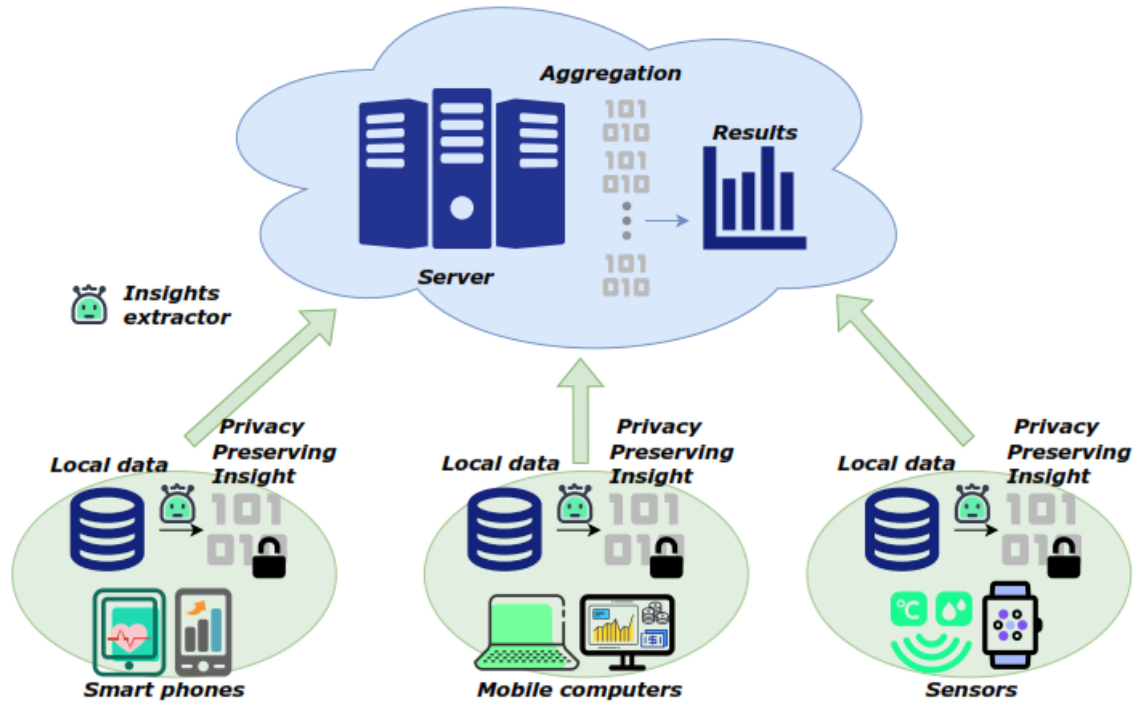


Figure 2.1: Basic FA setup. Extracted from [Wang et al. 2024]

Federated Analytics (FA) shares some similarities with Federated Learning (FL). Both involve collaboration with data distributed across multiple sources (such as devices or servers). They aim to preserve data privacy by keeping the raw data locally, allowing analysis or learning to occur without exposing sensitive information [Elkordy et al. 2023]. Both operate in a decentralized platform, with data processing or model training happening locally on individual devices or servers rather than in a centralized location. However, the primary goal of FL is to train machine learning models on distributed devices while preserving data privacy [Fachola et al. 2023]. This involves optimizing the model's performance using data from multiple sources without centralizing it. On the other hand, FA is focused on data analysis and generating insights from distributed data [Elkordy et al. 2023]. It includes a wider range of analysis techniques beyond just machine learning.

2.2 Benefits of FA Compared to Centralized Data Analytics and FL

From [Wang et al. 2024], if we compare with centralized data analytics and FL technique, FA offers the following benefits:

- **Communication cost reduction:** FA reduces communication costs by transmitting only valuable insights derived from client data, rather than raw data. These insights

are structured to contain essential information for the analytics tasks and are typically independent of the number of data samples held by the clients.

- **Privacy preservation:** FA preserves privacy by preventing the transmission of raw data. Instead, only abstracted insights, which are less sensitive, are transmitted. Unlike traditional data analytics, which require raw data to be shared, FA improves privacy by both eliminating the need to transmit raw data and incorporating additional privacy mechanisms.
- **Broader task coverage:** FA provides expansive task coverage by addressing a wide range of data science tasks, beyond just neural network-based predictive tasks, which are the focus of FL. FA encompasses both basic calculations and more complex tasks that involve advanced data structures and computation procedures. By combining FA with FL, it effectively covers the full spectrum of learning and analytics tasks, enabling all data-driven activities to be performed in a federated environment.

2.2.1 FA System Model

The central figure in FA is the querier, often referred to as the questioner, who seeks to obtain insights or answers from data that is distributed across various clients, also known as parties. Each client holds a segment of the overall data, termed their local dataset. FA essentially operates as a collaborative data analysis framework where the querier aims to resolve a data analytics query through the cooperation of multiple data owners who manage their own local raw data. Rather than exchanging or transmitting raw data, intermediate responses to queries are shared. These responses are subsequently aggregated by the querier to provide a comprehensive answer to the posed question. From this generalized idea, by [Elkordy et al. 2023], if we consider $D = \{D_i\}_{i=1}^k$ as the set of local datasets from each client, the goal of FA will be for a central querier to answer the query Q as given in formula

$$Q(D) = F_{\omega}(D_1, D_2, \dots, D_k) \quad (2-1)$$

where F_{ω} is the parameterized function on the data describing the target query.

2.2.2 FA Application Domain

The results are validated through an application in the domain of ML model evaluation on classification problems. This motivation arises from the growing interest in FA, particularly in the collaborative assessment of the quality of ML models in relation to data from practical applications. This data, collected from real everyday situations

and not from simulated or controlled environments, offers a more reliable and complex representation of the conditions and variability that the models will face in real use.

As defined by [Elkordy et al. 2023], suppose there is a pre-trained ML classification model defined by ω . Then, in FA, a typical query to evaluate the model's accuracy on distributed datasets can be formulated as

$$Q_{\omega}(D) = Acc(\omega; \{D_1, D_2, \dots, D_N\}) = \sum_{i=1}^N \frac{|D_i|}{\sum_{i=1}^N |D_i|} Acc(\omega; D_i), \quad (2-2)$$

where the answer to the query is determined by calculating the weighted average of each client's local test accuracy, denoted as $Acc(\omega; D_i)$. The term $\sum_{i=1}^N |D_i|$ serves as the normalizer. To find the local accuracy, each client uses its model to evaluate its own labeled dataset and locally computes the ratio of correct classifications. Following this idea, in this article we use a weighted averaging algorithm to calculate the aggregated accuracy and extend the method to aggregation of precision and recall, all with incorporation of reliability score (confidence).

2.2.3 Network Model and Role of FA in 5G and Beyond

Consider a 5G network composed of a Base Station (gNB) connected directly to a central FA server, as proposed in the *Multi-Access Edge Computing* (MEC) approach [ETSI 2022]. Suppose the existence of a set $S = \{i_1, i_2, \dots, i_N\}$ with N active 5G devices before starting an analysis session. Each client has a dataset \mathcal{D}_i sample stored on their respective local devices. These devices are connected to the gNB via a 5G connection and are capable of collecting data and performing pre-processing necessary for specific federated analysis tasks.

Instead of transmitting raw data to the central server, devices perform local calculations to respond to specific queries issued by the FA server. Intermediate responses which may include summary statistics are transmitted to the gNBs. The aggregation of these intermediate responses is performed by the server to compile a consolidated analysis or a final result, maintaining the confidentiality of the original data from each device. To reinforce security and privacy, techniques such as Differential Privacy (DP) or Secure Multiparty Computing (MPC) can be integrated into the FA model for 5G networks. This arrangement minimizes the need for bandwidth for data transmission and increases privacy and security, taking advantage of the high capacity and low latency of 5G networks to perform efficient and secure data analysis.

As anticipated by [Parra-Ullauri et al. 2024], future-generation networks (B5G/6G) systems will still face challenges regarding privacy and confidentiality due to their distribution (with more interconnected nodes, such as smartphones, IoT devices, and

edge servers), capacity and data heterogeneity, and varied ownership (administrators). In these scenarios, the role of FA will be to enable collaborative data analysis while maintaining privacy. In addition, they detail that a key factor that will make FA effective in B5G/6G networks is having a framework that integrates abstraction, supports various functions, and accelerates the creation of new applications, as illustrated in Figure 2.2.

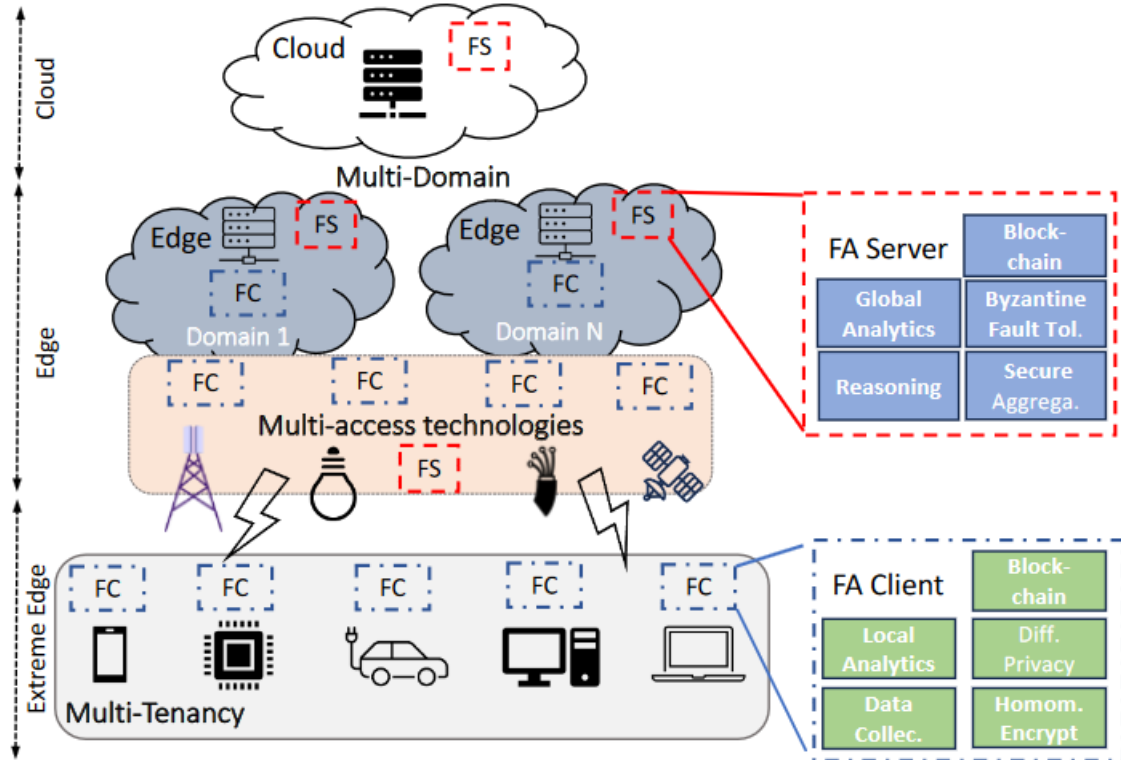


Figure 2.2: FA setup for 6G. Extracted from [Parra-Ullauri et al. 2024]

FA will become increasingly important for both end users and key industries, as it enables large-scale data analysis across networks without the need to share raw data. This will benefit B5G/6G networks in at least two ways [Parra-Ullauri et al. 2024]. Firstly, FA will allow analysis within different administrative areas like private networks and various edge device providers. In other words, it will facilitate analysis across different administrative areas, such as private networks and various edge device providers, by allowing each area to gather insights locally and then combine them for broader conclusions. Secondly, in the multi-tenancy setup of B5G/6G, where multiple users share the network infrastructure, strict limits are expected on data sharing between users. FA will address this challenge by providing network management insights across all users without exposing sensitive data. This makes FA a promising method for ensuring privacy during data analytics in the B5G/6G era [Parra-Ullauri et al. 2024].

2.2.4 Applications of FA and Query Types

Since FA was introduced, FA applications have expanded to various important fields. Based on [Elkordy et al. 2023] and [Wang et al. 2022], the following are some of the common applications fields:

- Evaluation analytics for machine learning models – Used by Google engineers to evaluate the accuracy of prediction models like those from Gboard by using captured data from anonymous users' typing activities on their phones.
- Analytics for medical studies and precision healthcare – Used for precision medicine to facilitate diagnosis and research with large volumes of data, without exposing sensitive patient information, supporting studies on treatments and symptoms without the data leaving the local medical institutions.
- Guiding advertisement tactics – Used for advertisers to understand the effectiveness of their ads on the potential customers, through analysis of anonymous viewing data, for future advertising strategies optimization, without compromising users' privacy.

As stated by [Elkordy et al. 2023], FA query can be understood as a general class that includes any question by a querier on distributed private datasets. They categorize FA queries into three groups and discuss their potential applications as given below:

- Statistical testing queries – applied in data science queries used to discover key statistical properties of the distributed private data. Examples can be the estimation of the mean median, heavy hitters, key-valued data frequencies, and hypothesis testing.
- Set queries – applied in analytics for discovering data associations such as set intersection, set union, and intersection cardinality, etc.
- Matrix transformation queries – applied in operations such as dimensionality reduction using methods such as principal component analysis, and projections.

2.3 Network Simulator 3 (NS-3)

NS-3 is an open-source discrete-event network simulator used for simulating internet systems, licensed under the GNU GPLv2 [Ceballos et al. 2021]. It is a powerful tool for research, designed to model network elements, simulate complex network behaviors, and analyze the performance of networking protocols in a controlled environment [Hapanchak e Costa 2022]. The NS-3 simulator works with scripts in C++ and its architecture is composed of a set of modules, containing the abstraction, the core, and the compiler of the NS-3 simulator. The NS-3 models are more realistic and more efficient

in the simulation context. NS-3 uses a real-time emulator, real network functions from the computer and can connect to external devices [Hapanchak e Costa 2022]. Figure 2.3 gives an architecture of NS-3.

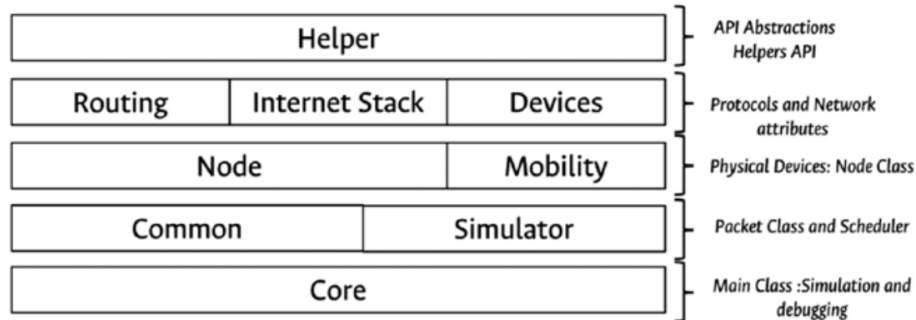


Figure 2.3: NS-3 architecture. Extracted from [Ceballos et al. 2021]

2.4 The 5G-LENA NR

The Centre Tecnològic de Telecomunicacions de Catalunya [CTTC-LENA 2024], describes 5G-LENA as an open-source simulator for 5G New Radio (NR) networks, built as a module for the NS-3 simulator, to allow performing extensive and more advanced research. Initially funded by InterDigital, it is available under GPLv2 license to facilitate collaboration, faster adoption, contributions from industry and academia, and results reproducibility. The simulator supports end-to-end simulations, covering from the application to the physical layers. 5G-LENA provides valuable insights for designing 5G solutions and evaluating the potential performance of such solutions. Discussed by [Larrañaga et al. 2023], similar to NS-3, 5G-LENA applies C++ classes to simulate the 5G functionalities. Figure 2.4 gives its end-to-end class architecture.

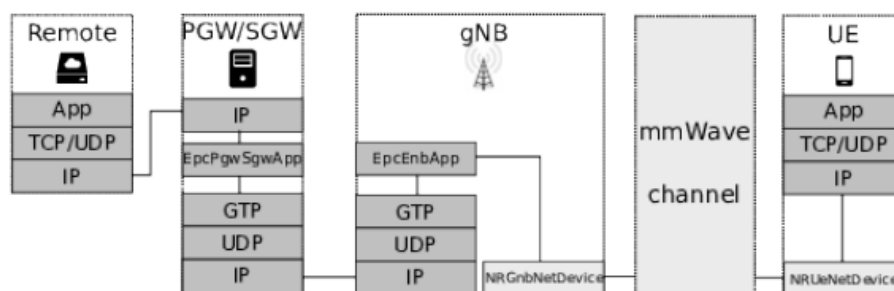


Figure 2.4: 5G-LENA Class architecture. Extracted from [Larrañaga et al. 2023]

According to [Koutlia et al. 2022], the 5G-LENA simulator offers key features, including support for frequencies from 400 MHz to 100 GHz using the 3GPP TR 38.901 channel model [38.901 2017]. It uses a non-standalone architecture combining 5G RAN

and 4G EPC and provides flexible NR frame structure configuration with multiple numerologies. The simulator supports OFDMA and TDMA access schemes, variable TTIs, and redesigned MAC layers with advanced schedulers for time- and frequency-domain resources. It features uplink grant-based access, FDD and TDD with configurable patterns, NR-compliant processing timings, and advanced Bandwidth Part (BWP) management. The physical layer includes LDPC coding, 256-QAM, and beamforming with dual-polarized MIMO and rank adaptation, supporting both directional and isotropic radiation.

Resource scheduling plays a crucial role in the efficient operation of 5G networks, ensuring optimal bandwidth utilization and fairness in resource allocation among user equipment (UEs). From [CTTC-LENA 2024] through the NR Module Release 3.3, in the NS-3 5G-LENA, resource scheduling takes place in the MAC layer of the gNB (base station), where resource blocks (RBs) are dynamically assigned to UEs based on channel conditions, QoS requirements, and network load. This mechanism directly impacts metrics such as throughput, latency, and fairness in spectrum access.

To optimize spectrum usage, as detailed in [CTTC-LENA 2024], 5G introduces Bandwidth Parts (BWPs), which allow for the segmentation of bandwidth into different ranges for various applications and traffic profiles. This approach is particularly useful in scenarios where different services require specific resource allocations, such as high-quality video traffic, ultra-reliable low-latency communication (URLLC), and massive IoT services.

By implementing a centralized scheduling mechanism at the gNB, the scheduler makes periodic decisions to allocate RBs in both the downlink (DL) and uplink (UL). This process takes into account factors such as:

- CQI (Channel Quality Indicator) Reports: UEs periodically send channel quality metrics to the gNB, which uses this information to make resource allocation decisions.
- HARQ (Hybrid Automatic Repeat Request) Retransmissions: Packets that were not correctly delivered are retransmitted to ensure reliability.
- MCS (Adaptive Modulation and Coding): The modulation and coding scheme is dynamically adjusted to optimize spectral efficiency and reduce transmission errors.

The allocation of RBs within the high throughput BWP, depending on the requirement, is managed by different scheduling strategies, with the main ones being:

- Round-Robin (RR): Distributes RBs equally among all UEs, ensuring fairness but without considering channel quality.
- Maximum Throughput (MT): Prioritizes UEs with better channel conditions, maximizing overall network throughput, but potentially excluding UEs with weak signals.

- Proportional Fair (PF): A compromise between efficiency and fairness, allocating more RBs to UEs with good channel conditions, while ensuring that UEs with lower throughput also receive resources.
- Token Bank Fair Queueing (TBFQ): Implements a credit system, allowing UEs to accumulate transmission opportunities over time to ensure fairness.

2.5 The 5th Generation Wireless Communication Networks (5G)

The evolution of each subsequent generation of wireless communication networks represents an advancement over existing technologies and incorporates new functionalities to meet future demands [Qadir et al. 2023]. From [Imoize et al. 2021], commenced in 2019, 5G commercialization opened space for many innovative technologies including the Internet of Things (IoT), immersive gaming, virtual reality [Kurt et al. 2021]. 5G has introduced significant improvements such as mobile broadband, massive machine-type communications, and ultra-reliable low latency communications, among others, aimed at addressing the limitations of 4G [Kurt et al. 2021]. In comparison to its predecessor, 5G networks boast capabilities such as providing up to 20 Gbps data rates, tripling spectral use efficiency, a 100 times increase in energy efficiency, and a gigabit-per-second user experience with end-to-end latency as low as 1 ms [Damsgaard et al. 2023]. Additionally, 5G is designed to sustain uninterrupted connectivity for devices with mobility of 500 km/h, support a connection density of 1 million devices per square kilometer, and provide an area traffic capacity of 10 Mbps per square meter [Luo, Li e Wang 2018]. Despite its potential to enable a wide range of smart Internet of Everything (IoE) services, as anticipated in [You et al. 2021], 5G alone may not be sufficient to fulfill the continuously expanding demands of wireless communications of future smart communities.

2.6 Beyond 5th or 6th Generation Wireless Communication Networks (B5G/6G)

Given the increased need for advanced services and the growing number of interconnected devices, 5G wireless communication systems will need to upgrade to guarantee high throughput, low latency, continuous connectivity, and security [Salahdine, Han e Zhang 2023]. Still from [Salahdine, Han e Zhang 2023], the next generation B5G and 6G are expected to be designed with features to empower wireless communication networks, to achieve high-quality connectivity in environments of high

demand of connected devices, such as industries, and fulfill the needs of diverse applications. These features will lead to integration of more intelligent and sophisticated applications, enabling technologies that operate on very high frequencies and broader bands [Salahdine, Han e Zhang 2023].

B5G/6G networks are expected in future to facilitate the creation of a fully automated smart environment that will cover a large number of people. To realize this, several technological advancements will be made, covering software-defined networking (SDN), massive MIMO, network function virtualization (NFV), vehicular-to-everything (V2X), mobile edge computing (MEC), network slicing, terahertz communications, visible light communication, virtualization of network infrastructure to the implementation of an intelligent communication environment [Salahdine, Han e Zhang 2023]. This work falls within the goal of the development of an intelligent data security and privacy communication environment for next-generation wireless networks.

The following table 2.1 illustrates the differences between 5G and B5G/6G wireless networks with respect to with respect to the supporting technologies, data features, and enabling applications.

Table 2.1: Differences between 5G and 6G. Extracted from [Kadir et al. 2021]

| Issue | 4G | 5G | 6G |
|------------------------------|-----------------|-----------------|---------------------|
| Per Device Peak Data Rate | 1 Gbps | 10 Gbps | 1 Tbps |
| End to End (E2E) Latency | 100 ms | 10 ms | 1 ms |
| Maximum Spectral Efficiency | 15 bps/Hz | 30 bps/Hz | 100 bps/Hz |
| Mobility Support | Up to 350 km/hr | Up to 500 km/hr | Up to 1000 km/hr |
| Satellite Integration | No | No | Fully |
| Artificial Intelligence (AI) | No | Partial | Fully |
| Autonomous Vehicle | No | Partial | Fully |
| Extended Reliability (XR) | No | Partial | Fully |
| Haptic Communication | No | Partial | Fully |
| THz Communication | No | Very Limited | Widely |
| Service Level | Video | VR, AR | Tactile |
| Architecture | MIMO | Massive MIMO | Intelligent Surface |
| Maximum Frequency | 6 GHZ | 90 GHZ | 10 THz |

2.7 Wireless Links, Channels and Characteristics

Wireless links enable communication between devices without physical connections, using electromagnetic waves to transmit data over the air. These links are fundamental to modern networks, including Wi-Fi, cellular (4G, 5G, and beyond), and IoT systems. Unlike wired connections, wireless links are subject to several unique characteristics, such as signal factors like distance, obstacles, and environmental conditions influence the channel's reliability and performance, which in turn affect the link's performance [Kurose e Ross 2021]. Understanding these characteristics is crucial for optimizing network efficiency, ensuring robust connectivity, and developing advanced wireless communication technologies.

According to [Kurose e Ross 2021], wireless links differ from their wired counterparts in a number of important ways:

- **Signal attenuation** - Electromagnetic radiation attenuates as it passes through matter (e.g., a radio signal passing through a wall). Even in free space, the signal will disperse, resulting in decreased signal strength (also referred to as path loss) as the distance between sender and receiver increases.
- **Interference from other sources** - Radio sources transmitting in the same frequency band will interfere with each other. For example, 2.4 GHz wireless phones and 802.11b wireless LANs transmit in the same frequency band. Thus, the interference can occur between them. In addition to interference from transmitting sources, electromagnetic noise within the environment (e.g., a nearby motor, a microwave) can result in interference.
- **Multipath propagation** - Occurs when portions of the electromagnetic wave reflect off objects and the ground, taking paths of different lengths between a sender and receiver. This results in the blurring of the received signal at the receiver. Moving objects between the sender and receiver can cause multipath propagation to change over time.

Husheng Li defines a communication channel as the medium through which information is sent from the source to the destination [Li 2016]. It is an abstraction of practical communication systems and can be modeled using mathematics. Once the mathematical model is set up for the communication channel, it can be possible to estimate the capacity of the channel. A wireless channel model provides a simplified representation of real-world conditions, capturing essential aspects of wireless propagation that impact the performance of specific wireless technologies [Tarokh et al. 2009]. In wireless channels, information is transmitted through electromagnetic waves in open space. Common wireless communication systems, such as cellular networks and Wi-Fi, enable mobile connectivity, although, these channels tend to be unreliable and typically offer li-

mitted capacity [Li 2016]. The authors in [Tarokh et al. 2009], based on ITU-R IMT-2000 program, classify terrestrial mobile and personal communication systems into three main propagation environments. Picocells represent indoor settings where the distance between the transmitter and receiver is typically within a few hundred meters. Microcells are outdoor environments where both the base station and remote terminal antennas are positioned below the local rooftop level, often within the same street or nearby streets, with a maximum separation of about 1 km. In these cases, a line-of-sight (LOS) connection is often present. Macrocells, on the other hand, involve outdoor environments where the base station antenna is significantly elevated above the rooftops, while the remote terminal remains below, with distances extending up to several kilometers. These links usually operate under non-line-of-sight (NLOS) conditions.

As discussed by [Tarokh et al. 2009], path loss is the most fundamental measure of channel quality. In decibels, path loss, if represented by PL , is defined as

$$PL = P_t + G_t + G_r - P_r \quad (2-3)$$

where P_t and P_r are the time-averaged power levels (in dBm) at the output of the transmitter and the input of the receiver, respectively, and G_t and G_r are the gains (in dBi) of the transmitting and receiving antennas. The relationship between path loss and the distance, d , between the transmitter and receiver generally follows a power-law relation and can be described by

$$L_d = L_0 + 10n \log_2 \left(\frac{d}{d_0} \right) + X_\sigma \quad (2-4)$$

where L_0 is the value of path loss (in dB) at the reference distance d_0 , n is the distance exponent and X_σ is a zero-mean Gaussian random variable with standard deviation σ . The random variable X accounts for the location variability or shadow fading that is generally attributed to differences in the degree to which the path is obstructed at different points throughout the coverage area.

2.8 Propagation Models

The study of radio signal propagation path loss (PL) is important for planning, designing and evaluating the performance of radio communication networks [Khaled, Ajib e Mcheick 2022]. The attenuation of signals due to complex environmental conditions, such as vegetation in rural areas, buildings in urban areas and indoor obstacles may lead to radio signal path loss (PL), which limits coverage range and quality of service (QoS) [Khaled, Ajib e Mcheick 2022]. To address this challenge, PL models are developed and refined based on the specific environment under rese-

arch, aiding in the planning, designing, and performance evaluation of wireless systems [Khaled, Ajib e Mcheick 2022]. Many PL models are available in the literature, such as the free space, log distance (LD), 3gpp models, among others.

Generally, propagation path loss models are designed to calculate the difference between transmitted power and received power by employing empirical methods, deterministic methods, or a combination of both. By nature, empirical models frequently need the adjustment of certain parameters based on field strength measurements taken within a specific environment [Fernández et al. 2012]. For that reason, the log-distance path loss model assumes that the mean power loss varies logarithmically with distance and makes use of a parameter called propagation exponent that must be modified for each environment to be studied [Fernández et al. 2012]. In this project, to simulate the transmission channel with noise and interference, propagation models 2-5 and 2-6 are applied namely, UMi-street canyon line-of-sight and non-line-of-sight path loss models [38.901 2017] to model microcell urban environment studied in this project.

In urban microcell (UMi) environments, particularly in street canyon scenarios, is considered that buildings line both sides of the streets, signal propagation is significantly affected by reflections, diffraction, and shadowing [38.901 2017]. The 3GPP TR 38.901 standard provides path loss models to estimate signal attenuation in both Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) conditions for these environments. The LOS model characterizes scenarios where the direct signal path between the transmitter and receiver remains unobstructed, while the NLOS model accounts for conditions where obstacles (e.g., buildings, vehicles) can obstruct direct transmission, leading to increased path loss.

For environment of a clear direct path existence between the transmitter and receiver, the LOS path loss model in Equation 2-5 is applied:

$$L_1 = 32.4 + 21 \cdot \log_{10} \left(\frac{d}{1m} \right) + 20 \cdot \log_{10} \left(\frac{f}{1GHz} \right) + 4.0 \quad (2-5)$$

where from [38.901 2017], the value 32.4 in dB is the offset that accounts for the baseline free-space loss in urban environments. The distance-dependent term $21 \cdot \log_{10}(d)$ indicates path loss relation with the logarithm of distance. The frequency-dependent term $20 \cdot \log_{10}(f)$ reflects the increased attenuation experienced at higher frequencies due to atmospheric absorption and material penetration losses. Finally, 4.0 in dB is a correction factor added to fine-tune the model based on empirical measurements.

For environment of a direct path nonexistence (blocked) between the transmitter and receiver, the NLOS path loss model in Equation 2-6 is applied:

$$L_2(d, f) = 22.4 + 35.3 \cdot \log_{10} \left(\frac{d}{1m} \right) + 21.3 \cdot \log_{10} \left(\frac{f}{1GHz} \right) - 0.3 \cdot hUt - 1.5 + 7.82 \quad (2-6)$$

The baseline offset of 22.4 in dB represents the initial signal loss in an NLOS setting, which is expected to be lower than that of the LOS model due to additional loss factors. The distance-dependent term $35.3 \cdot \log_{10}(d)$ indicates the path loss relation with distance compared to the LOS model, as the signal undergoes multiple reflections and diffractions. The frequency-dependent term $21.3 \cdot \log_{10}(f)$ is slightly higher than in the LOS model, as higher frequencies are more susceptible to absorption by obstacles. An additional term, $-0.3 \cdot (hUt - 1.5)$, introduces height dependency, to account for the influence of the height of the user terminal (hUt), for example, 1.5 meters used in this project, to overall loss. Finally, a correction factor of 7.82 in dB is added to refine the model for real-world deployments.

2.9 Energy Model

Energy consumption is an important metric used to evaluate the performance of wireless network protocols. For example, in an energy-aware protocol, where the operation depends on the energy level, an energy model becomes fundamental for effective simulation [Energy Model - Nsnam 2024]. According to [Energy Model - Nsnam 2024], in NS-3, a node can only have a single energy source, that represents the total energy reserve at the node. If it's required, multiple device energy models can be configured on a node, but each represents a different device. In that case, there is no dependence between the energy models, each device energy model will only notify the energy source of the energy consumed by that device, and update the remaining energy of the source [Energy Model - Nsnam 2024]. When energy is drained, the energy source will notify all device energy models connected to it.

In NS-3, the libraries were created using object-oriented programming, where their blocks are created using classes. Within the main energy model, there is a device energy model which is a class responsible for monitoring the state of a device, and accordingly calculate its energy consumption [Energy Model - Nsnam 2024]. It provides interfaces for updating the remaining energy in the energy source and handles the notification from the energy source when node energy is depleted. It also maintains a record of the total energy consumption of the device. Continuing with [Energy Model - Nsnam 2024], following the nature of NS-3, the child classes of the device energy consumption class implement energy consumption schemes of specific devices, such as radio device. For example, one typical child implementation of this class is the radio energy model class that represents the energy model for WI-FI radio devices with states. The Wi-Fi Radio Energy Model in NS-3 quantifies the energy consumption of a Wi-Fi NIC based on its operational states (transmit, receive, idle, switch, sleep, etc) [Energy Model - Nsnam 2024]. If we consider E_x as representing energy consumption during transmission, then E_x is

calculated by the formula

$$E_x = P_x \cdot t_x \quad (2-7)$$

where:

- P_x → is the power consumed per packet during transmission (dBm – dependent on NIC), measured in NS-3 5G-LENA
- t_x → is the time in seconds spent on transmitting, measured in NS-3 5G-LENA
- E_x → is the transmission energy (J)

The formula 2-7 is used because it directly links the power required for transmission with the time it takes to transmit the data, which allows us to understand the energy efficiency of the network. By focusing on successfully transmitted packets, we exclude retransmissions and packet drops, which helps isolate the energy used for actual data delivery. This method is appropriate as it aligns with the goal of measuring and optimizing energy consumption for effective data transmission, providing a clear analysis of how network conditions impact the energy efficiency of the successful data transfer process. For a broader analysis, future work could consider including retransmissions and dropped packets to provide a more comprehensive evaluation of energy usage, reflecting the full impact of network inefficiencies. This would likely lead to higher energy measurements due to the additional energy required for handling transmissions.

2.10 Performance Measurement: Delay, Latency, Throughput and Loss

The quality of the network significantly influences data transmission. The propagation speed of a link varies based on the physical characteristics of the medium used for transmission. As packets travel from one node (host or router) to another across the network medium, they encounter various delays at each node, including nodal processing, queuing, transmission, and propagation delays. These factors collectively contribute to the overall nodal delay [Kurose e Ross 2021]. The performance of many Internet applications and services can be greatly affected by network delays. Delays vary based on infrastructure and geographic location of communicating computers. To understand computer networks, it's necessary to understand the nature and importance of these delays. Delays are divided into various types depending on the reasons causing them. According to [Kurose e Ross 2021] and [Comer 2011], these types are:

- Processing Delay - the time required to examine the packet's header and determine where to direct the packet, including other factors such as checking for bit-level errors in the packet that occurred during transmission.

- **Queuing Delay** – the time a packet spends in the memory of a switch or router waiting to be selected for transmission. The size of a queue depends entirely on the amount of traffic that has arrived recently on the device. Queuing delays account for most delays in the Internet. When queuing delays become large, it is said that the network is congested.
- **Transmission Delay** - this is the amount of time required to transmit all bits of a packet onto a communication link. In practical scenarios, transmission delays usually range from microseconds to milliseconds.
- **Propagation Delay** - the time it takes for a signal to travel through a transmission medium and it's dependent on the distance covered.
- **Access Delay** – the time needed to obtain an access to a transmission medium. It occurs in networks that use shared media, where devices sharing a medium must contend for access. Access delays depend on the number of devices that contend for access and the amount of traffic each device sends.
- **Switching Delay** – the time required by a device to forward a packet in a network. This is the time required to compute a next hop and begin transmission.
- **Server Delay** – the time required for a server to respond to a request and send a response. This is the time required for a server to examine a request and compute and send a response. Since servers normally queue incoming requests, it means that server delay is variable and depends on the current load. In many cases, a user's perception of Internet delay arises from server delay rather than network delays.

The preceding paragraphs discuss how various factors such as transmission and propagation conditions contribute to the overall performance of a network. In that context, while delays impact data transmission, from [Kurose e Ross 2021] and [Comer 2011], another issue arises related to the finite capacity of queues preceding the links, which introduces the possibility of packet loss as traffic intensity increases. When a queue reaches maximum capacity and a new packet arrives, the router will drop the packet, resulting in loss. Knowledge about this behavior is important in network performance assessment, where the focus is not only on minimizing delays, but also on managing the probability of packet loss, especially under high-traffic conditions. This highlights the need for effective congestion control and packet management strategies to maintain reliable data transmission across networks.

Regarding latency, [Comer 2011] defines it as the time required to transfer data across a network. It specifies how long it takes for data to travel across a network from one computer to another, and it is also measured in seconds. Still from [Comer 2011], throughput can be understood as the amount of data that can be transferred per unit time. Since the time taken to transfer data in some part is influenced by delays and latency, then higher delay or latency may increase the time taken for data to reach its

destination, thereby reducing the rate at which data can be effectively transmitted. This lowers throughput.

In this project, we measured average end-to-end throughput for received packets, obtained by the Formula 2-8 extracted from [nstram]

$$\text{Throughput} = \frac{\text{rxBytes} \times 8.0}{\text{flowDuration} \times 10^6} \quad (2-8)$$

where

- rxBytes → represents the total number of bytes successfully received at the destination node. Multiplication by 8.0 converts bytes to bits.
- flowDuration → refers to the total time (in seconds) taken for the flow to complete. Division by 10^6 converts bits per second (bps) to Megabits per second (Mbps).

The end-to-end throughput quantifies the data rate at the receiver, considering only successfully received data. Unlike link throughput, it excludes retransmitted or dropped packets to provide a measure of delivered data. Since the flow duration used in the Formula 2-8 represents the total transmission period, variations in link conditions, queuing delays, and protocol mechanisms contribute to changes in the measured throughput.

In the same way, we measured average end-to-end latency for received packets, using the Formula 2-9 extracted from [nstram]

$$\text{Latency} = \frac{1000 \times \sum_{i=1}^N \text{Delay}_i}{N} \quad (2-9)$$

where

- $\sum_{i=1}^N \text{Delay}_i$ → calculates the total of all individual delays.
- N → is the total number of received packets. Dividing by N gives the average delay for the N packets.
- The factor 1000 → is used to convert the delay from seconds to milliseconds.

The end-to-end latency quantifies the average delay per successfully received packet. It is influenced by factors including queuing delays, propagation delays, and processing delays, all of which affect packet delivery times. Like throughput, it is impacted by network congestion, link conditions, and protocol mechanisms, which can contribute to increased delays in data transmission. These factors also impact the overall data delivery efficiency in real-world networks. For further analysis, link throughput could be included in future works to provide a more comprehensive evaluation of network performance. Furthermore, the analysis could be extended to include round-trip time (RTT) to help assess latency improvements and provide deeper insights into network responsiveness.

2.11 Signal-to-Noise Ratio and Bit Error Rate

In wireless communications, when the base station (gNB in 5G) transmits a signal, the user equipment (UE –in 5G) receives an electromagnetic signal that consists of a degraded version of the original transmission (due to attenuation, multipath propagation effects, and other factors), combined with background noise present in the environment. This behavior results in what is known as the Signal-to-Noise Ratio (SNR). From [Kurose e Ross 2021], the SNR is a relative measure of the strength of the received signal (i.e., the information being transmitted) and this noise. The unit of measurement of SNR is decibels (dB). Mathematically, for example, [Kurose e Ross 2021] explain that the SNR measured in dB, is 20 times the ratio of the base-10 logarithm of the amplitude of the received signal to the amplitude of the noise. It is important to know that a larger SNR makes it easier for the receiver to extract the transmitted signal from the background noise.

Still discussed by [Kurose e Ross 2021], the bit error rate (BER) is the probability that a transmitted bit is received incorrectly. The Figure 2.5 shows that BER is directly related to the SNR. That is, for a given modulation scheme, a higher SNR results in a lower BER. Therefore, increasing the SNR helps reduce the BER. A sender can increase the SNR by increasing its transmission power, that leads to decrease of the probability that a frame is received incorrectly. However, there are also disadvantages associated with increasing the transmission power, as more energy can be consumed by the sender, and the sender's transmissions can suffer more interference with the transmissions of other senders on the network. In this context, For a given SNR, modulation schemes with higher bit transmission rates tend to have higher BER. For example, with an SNR of 10 dB, BPSK modulation at 1 Mbps can achieve a BER below 10^{-7} . In contrast, QAM16 modulation at 4 Mbps may have a BER as high as 10^{-1} , which is too high for practical use. With an SNR of 20 dB, QAM16 can operate at 4 Mbps with a BER of 10^{-7} , and BPSK at 1 Mbps achieves an even lower BER. For this reason, [Kurose e Ross 2021] try to demonstrate that choosing the right modulation technique and parameters depends on network performance requirements and channel conditions. View in Figure 2.5 performance of BSPK, QAM16 and QAM256.

In this project, the SNR is used to determine the minimum transmission power required to guarantee efficient communication while maintaining a target delivery ratio (DR). It is compared to the noise power to assess the quality of the received signal. The SNR influences the Bit Error Rate (BER) and Packet Error Rate (PER), higher SNR values result in lower error rates, which in turn improves the probability of successful packet delivery. Therefore, SNR is acting as central to the dynamic adjustment of transmission power in response to changes in DR. If the DR falls below the target, transmission power is increased, while if the DR exceeds the target, power is reduced. This adjustment process

relies on the SNR to guide the necessary changes in power to maintain the desired DR. The optimization algorithm iteratively recalculates the SNR, BER, PER, and DR, adjusting transmission power within set limits. In this way, SNR plays a fundamental role in ensuring that the communication system operates within the desired performance parameters. Note that for future works, incorporating the Signal-to-Interference-plus-Noise Ratio (SINR) could provide more detailed information about the quality of the received signal by considering both interference and noise.

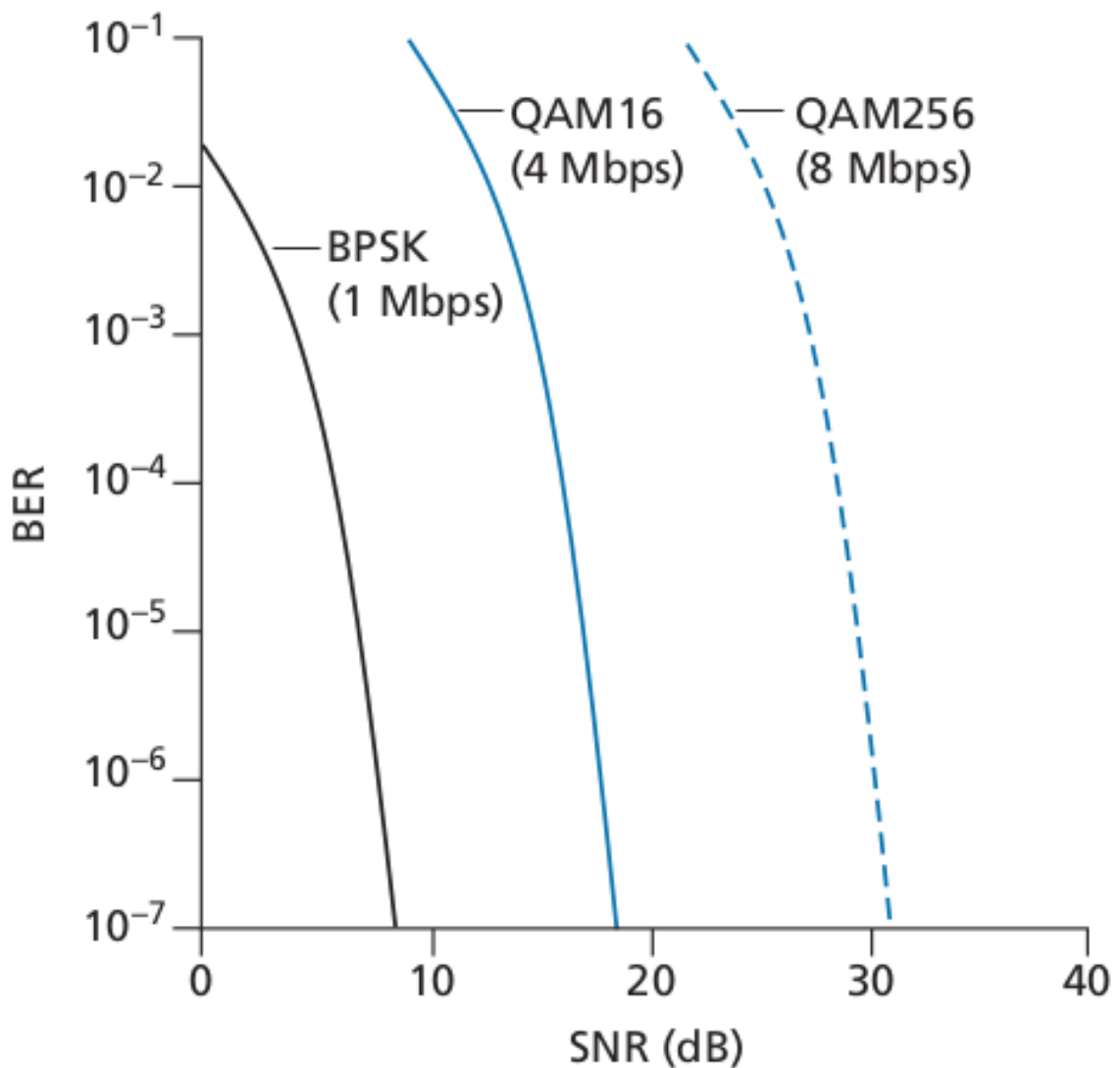


Figure 2.5: SNR, transmission rate, and Bit error rate relationship. Extracted from [Kurose e Ross 2021]

2.12 Bayesian Aggregation

According to [Ramachandran e Tsokos 2009], the Bayesian approach seeks to combine information from two sources: new information contained in the data and

knowledge from theory or opinion formed at the beginning of the research to form the basis for the statistical procedure. The prior distribution represents initial belief, while the information in the data is expressed by the likelihood function. By combining the prior distribution and likelihood function, it's possible to obtain the posterior distribution which expresses revised uncertainty in light of the data. The main difference between the Bayesian approach and the classical approach is that in the Bayesian setting, the parameter is viewed as random variables, while the classical approach considers the parameter to be fixed but unknown [Ramachandran e Tsokos 2009]. The parameter is random in the sense that it can be assigned a subjective probability distribution that describes the confidence about the actual value of the parameter.

Bayesian statistics formalizes the way to express the probability of events incorporating data as well as prior beliefs or knowledge, and its definition is strongly related to conditional probabilities [Ramachandran e Tsokos 2009], given as

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{and} \quad P(A|B) = \frac{P(B \cap A)}{P(B)} \quad (2-10)$$

Noting that $P(A \cap B) = P(B \cap A)$ since $A \cap B$ and $B \cap A$ mean observing events A and B at the same time, we can substitute one expression into another. Taking $P(A \cap B) = P(A|B)P(B)$ into $P(B \cap A)$ we obtain

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2-11)$$

Based on [Ramachandran e Tsokos 2009], the Bayesian formula becomes important if we want to express how well our beliefs or hypotheses are matched with the data. If we use H to denote our hypothesis and E the evidence (data), then Bayesian theorem becomes

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad (2-12)$$

where:

- $P(H|E) \rightarrow$ is the posterior probability. This is the value to be found.
- $P(H) \rightarrow$ is called the prior and reflects the prior knowledge or hypothesis before looking at the data.
- $P(E|H) \rightarrow$ is called the likelihood. For a given hypothesis, this value describes the probability of observing the data previously established.
- $P(E) \rightarrow$ is called the evidence. It becomes a normalization constant and always incorporates the probability to observe the data.

When a specific value for the evidence is not available, but is necessary for calculations, the evidence can be decomposed and obtained by using the total law of

probabilities. As defined by [Ramachandran e Tsokos 2009], assuming that $S = E_1 \cup E_2 \cup \dots \cup E_n$ represents the space containing all events E , where $P(E_i) > 0$, $i = 1, 2, 3, \dots, n$ and $E_i \cap E_j = \emptyset$, then the total law of probabilities of events E is given by

$$P(E) = \sum_{i=1}^n P(E|H_i)P(H_i) \quad (2-13)$$

Substituting in formula 2-12, the formula used by the server to aggregate predictions from clients and calculate the final answer in our project is obtained, given as

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{i=1}^n P(E|H_i)P(H_i)} \quad (2-14)$$

2.13 Related Works

Existing literature has demonstrated the potential of FA in improving data privacy, reducing communication overhead, and improving scalability across distributed networks. Despite this success, specific implementations and optimizations projected for B5G/6G networks remain an area of active research and development.

2.13.1 Overview of Related Works

In recent literature, studies related to FA have been developed. These studies have explored different ways to address the challenges and practical uses of the FA approach. The study by [Zhao et al. 2022] introduced a semi-hierarchical FA framework that integrates FL with edge computing architectures. It used multiple edge servers to aggregate updates from IoT devices and combined learned model weights without relying on the cloud or a central server. The authors analyzed the convergence of this approach and its performance under variable parameters, unreliable networks, and packet loss. Experimental results from this study showed that the method offers efficient communication, robustness, and fault-tolerant data analysis for IoT networks. Like our project, their work included an evaluation of network conditions on FA performance through packet loss analysis. However, it differs in its focus on edge computing for IoT, and our work incorporates more network metrics in the analysis.

The work done by [Shi et al. 2022] proposed a framework called ‘‘Federated Anomaly Analytics enhanced Distributed Learning (FAA-DL)’’. This framework allowed clients and the server to collaborate in analyzing anomalies. First, FAA-DL identified all trained local models and separated the potential malicious ones. Then, it checked each potential malicious local model using functional encryption. After that, it eliminated the verified anomalies and aggregated the remaining models to generate the global model.

To evaluate FAA-DL, they trained classifiers on MNIST and Fashion-MNIST datasets under various local model poisoning attacks. Experimental results indicated that FAA-DL improves the accuracy of the global learned model by up to 6.90 times under strong attacks and outperforms state-of-the-art defense methods with a guarantee of robustness. This work illustrates the application of FA on anomaly detection, while our work projects the application of FA in future generation networks. Differently, their study does not include the analysis of network impact in FA performance.

Applying FA techniques, [Yue, Kontar e Gómez 2024] developed a “Federated Data Analysis (FDA)” approach to improve the statistical linear regression model. Their approach was based on hierarchical modeling, allowing the combination of information across multiple groups. To achieve this, the authors introduced two federated hierarchical model structures that facilitated information sharing between devices, thereby promoting collaborative analysis. The proposed frameworks offered capabilities such as uncertainty quantification, variable selection, hypothesis testing, and rapid adaptation to new unseen data. They validated their methods in various real-life applications, including monitoring aircraft engine conditions. The results demonstrated that the FDA approach for linear models could serve as a competitive benchmark model for the future development of federated algorithms. Apart from demonstrating the application of FA, in their study, they made use of statistical prediction methods. In our work, we also apply statistical methods such as the Bayesian aggregation method. However, their study does not include the analysis of network impact on FA performance.

The study done by [Fachola et al. 2023], demonstrated how FA, using quantitative methodologies with data from educational platforms, student performance metrics, and device usage patterns, among others, can effectively address educational challenges such as identifying user profiles, shaping personalized learning trajectories, and predicting early dropout risks, among others. This study focused on applying Federated Learning (FL) methods to a fundamental learning analysis problem, like predicting student dropout rates. Through experimentation, they found that their proposed solutions showed performance results comparable to centralized approaches. As our work, their project discussed a use case of FA. They apply FA in education, while our work is projected for future network generations. Yet, their work also does not include the analysis of network impact on FA performance.

Designed for decentralized data analytics in connected vehicles, [Zhao et al. 2024] extended the concept of federated learning (FL) to support decentralized model training directly on vehicles, which helped eliminate the need for a centralized server. To improve communication efficiency, they introduced a federated regularized nonlinear acceleration-based local training method aimed at minimizing communication rounds. Also, they applied a random broadcast gossip-based mechanism

to lower the computational complexity per iteration. Experimental results illustrated that their approach significantly reduces communication overhead when compared to conventional gradient descent and momentum-based FL methods. These results suggested its potential for improving data analytics efficiency within autonomous vehicle settings. Related to our work, they concentrated their study on communication efficiency. Our work also studies the communication efficiency but on a different perspective, which is network.

The research by [Toka et al. 2023] from the article "5G on the Roads: Latency-Optimized Federated Analytics in the Vehicular Edge" focused on the coordination of vehicles in autonomous driving, with a particular emphasis on communication and reducing latency in data collection, processing, and sharing. Their work studied how federated analytics can be used to preserve data privacy while maintaining up-to-date, high-definition maps through crowd-sourcing. They analyzed latency optimization techniques and proposed a system that applies a two-tiered infrastructure consisting of mobile vehicular nodes and static nodes, utilizing 5G's low-latency and high-throughput capabilities. Their results showed that careful system planning and the integration of 5G infrastructure can significantly decrease latency in data distribution, especially in densely populated metropolitan areas. Similar to our work, they targeted latency optimization and information delivery for 5G networks. However, while their study focused only on latency, our research extends the optimization to include additional metrics for 5G networks, aiming to minimize the overall impact on FA performance. Also, their project uses both mobile and fixed nodes, but our work only uses fixed nodes.

While the majority of these studies discuss the applications of FA, and a few discuss the impacts of network conditions such as unreliable channels, packet loss and latency, none of them evaluate in detail how the unreliable nature of wireless transmission link affects FA. They neither integrate NS-3 for simulations nor perform network optimization for FA. Only one work by [Toka et al. 2023] attempts to apply network latency optimization on FA. The studies primarily focus on machine learning and the application of FA itself. The low quality of the wireless network can significantly impact communication and FA performance.

Another aspect not discussed is the network bandwidth. Bandwidth limitations can also pose challenges for FA, especially in transferring updates and query responses between distributed clients and servers. Understanding how bandwidth limitations affect data transmission and model updates is important for improving FA processes. Another concern of distributed environments is resource allocation and utilization. In environments where network conditions vary, it's important to allocate resources effectively to ensure fair participation and efficient use of computational resources. Examining how wireless network conditions affect resource allocation decisions and the scalability of FA can

provide insights to enhance system performance and scalability.

Similarly, these studies have not explored the evaluation of the impact of noise and interference in wireless networks on FA performance. This parameter, like others, is necessary because it helps understand how unfavorable conditions in the communication medium, such as real-world noise and interference, influence FA performance in actual network conditions. This information is important for developing strategies that optimize the robustness and reliability of FA systems, especially in wireless communications. All these aspects lack a deep approach and experimentation to gain a real understanding of how the nature of wireless networks impacts FA performance, and based on the obtained knowledge develop optimization strategies and algorithms for FA.

2.13.2 Comparison of this Proposal with Related Works

The works introduced by [Zhao et al. 2022], [Zhao et al. 2024] and [Toka et al. 2023] relate to this proposal when they concentrate on achieving efficient communication on FA. However, the strategies and technologies used differ from our work. The first work includes network impact analysis on FA performance, but it only considers packet loss, while our work includes more network metrics like throughput, latency and energy consumption. The third work also includes network analysis, but only concentrated on latency optimization. The second work does not include network parameters at all, it only concentrates on machine learning. The remaining studies developed by [Fachola et al. 2023], [Shi et al. 2022] and [Yue, Kontar e Gómez 2024] relate to our work on demonstrating the application of FA in various fields, but neither integrate NS-3 nor make network analysis on FA, nor try to apply any FA or network optimization technique. Their focus is only on machine learning. Table 2.2 gives more details on the differences between this proposal and the related work.

Table 2.2: Comparison of this proposal and the related works

| Work | Intervention | Metric | Contribution | Network Analysis | Integration with NS-3 or 5G-LENA |
|----------------------------|---|--|--|------------------|----------------------------------|
| [Zhao et al. 2024] | Communication for connected vehicles | FA communication cost | FA framework | ✗ | ✗ |
| [Yue, Kontar e Gómez 2024] | Federated hierarchical linear model structures | FA efficiency | Federated data analysis method | ✗ | ✗ |
| [Toka et al. 2023] | Communication for 5G-vehicular networks | Latency | Analytical model and optimization algorithm | ✓ | ✗ |
| [Fachola et al. 2023] | Application of FL methods to learning Analytics | <ul style="list-style-type: none"> • FA accuracy • FA running time | Horizontal FL techniques to apply FA to data distributed across different educational institutions | ✗ | ✗ |
| [Zhao et al. 2022] | Communication for IoT | <ul style="list-style-type: none"> • FA convergence • Network packet loss | FA framework | ✓ | ✗ |
| [Shi et al. 2022] | Local models anomaly detection and elimination | <ul style="list-style-type: none"> • FA accuracy, robustness, efficiency | FA framework | ✗ | ✗ |
| This proposal | Communication for B5G/6G networks | <ul style="list-style-type: none"> • FA accuracy, precision, recall • Throughput, latency, packet loss, delivery ratio, energy consumption | FA framework and optimization algorithm | ✓ | ✓ |

This proposal stands out as the most comprehensive project integrating the network with FA. Unlike related studies, which are primarily machine learning-focused, this work incorporates network performance evaluation as a core component for FA optimization. Besides using a different network technology (NS-3 with 5G-LENA), it considers a broader set of metrics, including throughput, latency, delivery ratio, and energy consumption. These metrics provide a clearer picture of network behavior from multiple perspectives, to ensure that network quality is maximized. This guarantees the efficient transmission of client responses to the server, which significantly contributes to FA performance optimization. See Table 2.2, for a detailed comparison between this work and related studies.

Optimized FA Algorithm Proposal for 5G Network

3.1 Proposed Solution

To solve the problems discussed in section 1.2, this work proposes a detailed approach to optimize FA performance through network improvements. The proposed solution introduces a framework that analyzes how varying network conditions, like latency, packet loss, throughput and energy consumption affect the final response quality, and overall performance of FA in the context of the real 5G network. The focus is on identifying factors that can improve the efficiency and scalability of FA systems in wireless channel conditions. Upon a deep understanding of the impacts of network on FA performance, two adaptive optimization algorithms were developed to improve the performance of FA systems. The algorithms are: channel-aware transmission power allocation algorithm and FA-5GLENA integrated algorithm. The first was designed to optimize and efficiently allocate the transmission power for UEs based on distance, and channel conditions. The second was adapted to integrate the FA and 5G-LENA simulator, to allow simulating the FA in real 5G network conditions.

Through detailed simulations and evaluations, the effectiveness of the proposed solutions in real-world scenarios were evaluated. We expect that this research will open the way for advancements in distributed data analysis systems projected for future B5G/6G networks. By improving communication efficiency, FA systems can operate more effectively and deliver timely insights while minimizing resource utilization and operational costs.

3.2 Objective Function for Power Allocation by Channel-Aware Algorithm

The main objective of this project is to improve communication between FA clients and the server in a 5G network by optimizing transmission power. The algorithm aims to dynamically allocate power for each UE to balance an increased delivery ratio and throughput with reduced latency and energy consumption. To achieve these goals based on transmission power as the main variable, the objective function is formulated as

The main objective of this project is to improve communication between FA clients and the server in a 5G network by optimizing transmission power. To achieve this, we vary the transmission power to ensure an optimal trade-off between power efficiency and network performance. The algorithm is designed to dynamically allocate transmission power for each UE to balance increased delivery ratio and throughput with reduced latency and energy consumption. With transmission power as the primary variable, the objective function is formulated as

$$\min \sum_{i=1}^N (P_i^{(t)}) \quad (3-1)$$

Subject to

- $0 \leq P_i^{(t)} \leq 30 \text{ dBm}, \forall i \in \{1, 2, 3, \dots, N\}$
- $\mathfrak{T}_i^{(t)}(P_i^{(t)}) \geq 100 \text{ Mbps}, \forall i$
- $\mathcal{L}_i^{(t)}(P_i^{(t)}) < 1 \text{ ms}, \forall i$
- $\mathcal{DR}_i^{(t)}(P_i^{(t)}) \geq 99 \%, \forall i$

Where

- $P_i^{(t)}$ is the transmission power for the i – th UE at iteration t .
- $\mathfrak{T}_i^{(t)}(P_i^{(t)})$ is the average end-to-end throughput for received packets from the i – th UE, as a function of transmission power, measured in 5G-LENA.
- $\mathcal{L}_i^{(t)}(P_i^{(t)})$ is the average end-to-end latency for received packets from the i – th UE, as a function of transmission power, measured in 5G-LENA.
- $\mathcal{DR}(P_i^{(t)})$ is the delivery ratio for the i – th UE, as a function of transmission power, calculated by the formula $\mathcal{DR}_i^{(t)}(P_i^{(t)}) = 1 - \text{PER}_i^{(t)}(P_i^{(t)})$.

Energy efficiency is important in wireless communication systems, especially in the context of 5G networks where multiple UEs are involved. Higher transmission power increases energy usage, as energy consumption is directly proportional to transmission power. Therefore, optimizing power helps in extending battery life for mobile devices, which enhances the sustainability of the network. Reducing energy consumption, reduces

operational costs, and minimizes interference with other network elements. Lower energy consumption also reduces the environmental impact. In mobile networks, where devices run on batteries, energy efficiency is a key factor.

Throughput is another most important performance metric in communication systems. It refers to the rate at which data is successfully delivered over the network and therefore directly impacts the overall FA performance. The constraint $\mathfrak{T}_i^{(t)}(P_i^{(t)}) \geq 100 \text{ Mbps}$ ensures that the throughput for each UE meets a desired performance for FA operation. Increasing transmission power can enhance throughput by improving signal quality, thereby enabling higher data rates. Maximizing throughput is essential for high-performance communication. However, increasing power beyond a certain level could lead to negative returns and interference. Optimization seeks the best balance.

In 5G networks, low latency is critical for FA to facilitate timely communication, updates and responses. The constraint $\mathcal{L}_i^{(t)}(P_i^{(t)}) < 1 \text{ ms}$ guarantees that latency remains within acceptable limits for 5G ultra-low latency (ULL) standards, thereby enhancing FA performance. Higher transmission power can reduce latency by improving signal quality, leading to faster data transmission and processing. This reduction in latency is essential, as excessive delays can result in packet loss, response corruption, and transmission failures.

The \mathcal{DR} is the proportion of transmitted data packets that are successfully received by the destination without errors. It is a crucial metric for network reliability and the overall performance of the communication system. The constraint $\mathcal{DR}(P_i^{(t)}) \geq 99 \%$ makes sure that the \mathcal{DR} meets the required threshold for reliable communication. Like other metrics, \mathcal{DR} is dependent on transmission power, therefore, higher transmission power may increase the signal strength, which can improve the delivery ratio by reducing the probability of packet loss. Maximizing the delivery ratio ensures that the network provides reliable communication, which is essential for critical applications in 5G networks. A high delivery ratio implies fewer retransmissions, lower delays, and a more stable network, which automatically improves FA performance and reduces resource waste.

3.3 Implementation of FA-5GLENA Integrated Framework

We have developed an integrated FA and NS-3 5G-LENA framework to analyze how different network conditions such as latency, packet loss, throughput, and power consumption affect the quality of the final response and the performance of FA in the context of real-world network conditions of 5G networks. NS-3 is an open source discrete event network simulator, licensed under the GNU GPLv2 [Ceballos et al. 2021]. It is a powerful research tool designed to model network elements, simulate complex

network behaviors, and analyze the performance of network protocols in a controlled environment [Hapanchak e Costa 2022]. 5G-LENA is a module of NS-3 that simulates 5G New Radio (NR) networks, which allows the simulation and evaluation of the potential of 5G solutions [Larrañaga et al. 2023].

The integration was made through a shared memory technique, allowing the FA to access and use the network results including the commands, and make the FA commands available for the 5G-LENA simulator. Both the server and clients operate synchronously in this setup. The central server, acting as the querier, coordinates queries and manages interactions among distributed clients engaged in data analysis tasks. Each simulation round starts with the FA server connecting to the NS-3 5G-LENA simulator via gNB, where it sends a network simulation request command (netrun) including initial parameters such as the number and list of selected clients, and the query. The server oversees the start, progression, and conclusion of the simulation for that round. After dispatching the initial parameters, the server awaits responses. Once the network request command is received, the 5G-LENA simulator performs the simulation, outputting network statistics like latency, packet loss, delivery ratio, throughput, and energy consumption for each client. To signal the end of the network simulation, 5G-LENA sends a command (netend) for the FA server. The network statistics are then utilized by the FA server in conjunction with FA algorithms to select clients for subsequent rounds and to aggregate FA results at the end of the simulation.

The use of shared memory enables real-time data exchange and interaction between the FA and 5G-LENA. When the project runs, the FA server writes the command netrun to the shared memory space, signaling the 5G-LENA simulator to start network simulation. Upon completion, the simulator summarizes the DRs for each UE, writes them to the shared memory, and signals the end of the simulation by writing the command netend. The FA server then reads the network command and DRs to analyze the network conditions and select clients for the current round of training, prediction, and results aggregation. This bidirectional communication setup via shared memory makes it possible for both systems to operate in a synchronized manner, similar to how real-time interactions and adaptations occur between FA processes and 5G network conditions. See in Figure 3.1 for detailed integration.

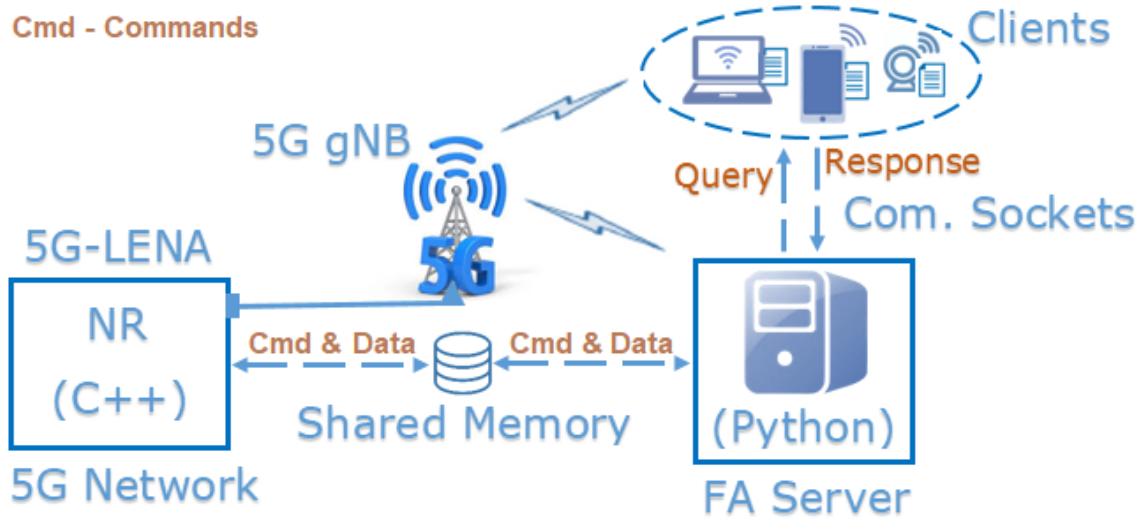


Figure 3.1: FA Integration with 5G-LENA Module

Following the server's commands, clients perform local model training and result prediction based on the local MNIST dataset in independent and identically distributed (IID) format, to guarantee that each client receives a statistically representative sample of the entire dataset. The simulation ends when the selected clients send their responses to the server, which then aggregates all the responses using Bayesian aggregation and weighted average techniques. To ensure security during transmission, secure communication protocols are applied between the clients and the central server, thereby guaranteeing data privacy and integrity.

3.4 Synchronous FA-5GLENA Integrated Algorithm

The algorithm is composed of two methods, Bayesian aggregation based on reliability score (prior), for determining the final prediction answer from the results set sent by clients, and weighted average for aggregating the performance metrics (accuracy, precision, recall) from the clients. The algorithm first simulates the NS-3 5G-LENA network to collect performance statistics (packet loss, throughput, latency, delivery ratio, and energy consumption) for 20 clients. Then, the statistics are used in the FA process, where the server selects clients with a delivery ratio $DR \geq 90\%$ to participate in the round. These selected clients train the model, then make classifications on MNIST based on received query, and send responses along with performance metrics to the server. Non-selected clients send error responses. The server aggregates all responses and determines the final answer to the query after each round. The aggregation is done with two methods, Bayesian aggregation, for determining the final prediction answer (class), and weighted average for aggregating the performance metrics (accuracy, precision, recall) from the clients. Considering that clients might have different dataset sizes, the two methods

were chosen because they take into account the contribution of each client based on its confidence (reliability) and dataset size as the weight, instead of treating all clients equally. The confidence value for each client is calculated from the function predict, which is used to classify images from the MNIST dataset. This function returns an array of probabilities and associated classes (0 to 9) as output. The prediction class associated with the maximum probability in the output array becomes the answer and its probability is the confidence value for a specific client.

The Bayesian aggregation method by applying parameters α for number of correct predictions (success) and β for number of incorrect predictions (failure), both from the beta distribution, updates and stores cumulative probabilities in a dictionary (class_probabilities) for each class. It then calculates total weight (total_weight) and confidence (reliability) across all clients. Finally, the method normalizes the class_probabilities to obtain the final aggregated probabilities, which are used to determine the final aggregated prediction (class).

The weighted averaging method uses dataset size from each client as the weight to generate the final weighted performance metrics average. A larger dataset size implies a higher weight for the client's contribution compared to others. For this process, the Equation 2-2 is applied. See Algorithm 1 for detailed operation.

Algorithm 1 Synchronous FA - 5GLENA Integrated Algorithm

- 1: **NS-3 5G-LENA simulates the network and collects statistical results of 20 clients**
 - 2: **FA Server reads NS-3 network statistical results**
 - 3: **for** each client i on iteration t in network **do**
 - 4: **Server selects a subset $S_j^{(t)}$ of clients with $DR_i^{(t)} \geq 90$ %**
 - 5: **Server sends a query $q^{(t)}$ to the selected clients in $S_j^{(t)}$**
 - 6: **for** each client j in subset $S_j^{(t)}$ **do**
 - 7: **Client j performs local training of model $\omega_i^{(t)}$ for n epochs**
 - 8: **Client j responds $r_i^{(t)}$ to the query $q^{(t)}$ including performance metrics**
 - 9: **end for**
 - 10: **for** each client i not in subset $S_j^{(t)}$ **do**
 - 11: **Client i sends error response $e_i^{(t)}$ and zero for performance metrics**
 - 12: **end for**
 - 13: **Server aggregates all responses and performance metrics from all clients**
 - 14: **end for**
 - 15: **Server determines final response**
-

3.5 The Channel-Aware Transmission Power Allocation Algorithm

We developed an adaptive optimization algorithm that efficiently manages and allocates transmission power according to the UE's distance from the gNB and channel conditions. This algorithm dynamically calculates the minimum transmission power required for a UE to achieve a delivery ratio $DR \geq 99\%$. It begins by estimating the initial power based on network parameters such as distance, frequency, bandwidth, and path loss (considering both Line-of-Sight and Non-Line-of-Sight conditions) of 3gpp UMi models [38.901 2017]. By applying feedback technique, iteratively adjusts the power, calculating the DR using SNR, BER, and PER. If the DR is below 99%, it increases the power; if the DR exceeds 99%, it reduces the power. The iterative process continues until the DR reaches 99% (convergence) or the maximum number of iterations is reached. The final output is the minimum power required to meet $DR \geq 99\%$, while ensuring efficient power usage within limits. See algorithm 2 for detailed operation.

Algorithm 2 Channel-Aware Transmission Power Allocation Algorithm

```

1: Initialize for Each UE  $i$ :
2: Calculate path loss models (LoS and NLoS)  $L_i^{(t)}$  based on 3gpp UMi loss models.
3: Calculate receiver sensitivity using noise power and target SNR.
4: Initial Transmit Power Calculation for each UE  $i$ :
5: Compute the initial transmit power using receiver sensitivity, path loss, and margin.
6: Main Iteration  $t$  for Each UE  $i$ :
7: for each UE  $i$  do
8:   Set  $P_i^{(t)}$  to the initial calculated value.
9:   While Loop: Continue adjusting  $P_i^{(t)}$  until convergence or maximum iterations.
10:  while  $P_i^{(t)} >$  minimum transmit power and iteration  $t <$  maximum do
11:    Calculate  $DR_i^{(t)}$  based on  $P_i^{(t)}$  and channel conditions.
12:    if  $DR_i^{(t)} < 99\%$  then
13:      Increase  $P_i^{(t)}$  to improve  $DR_i^{(t)}$ .
14:    else
15:      Decrease  $P_i^{(t)}$  to obtain minimum satisfying  $DR_i^{(t)} \geq 99\%$ .
16:    end if
17:    Adjust the step size as the  $DR_i^{(t)}$  approaches 99%, after feedback.
18:    if  $DR_i^{(t)}$  is sufficiently close to 99% then
19:      Exit loop early due to convergence.
20:    end if
21:    Increment the iteration  $t$  counter.
22:  end while
23:  Final Transmit Power  $P_i^{(t)}$  for UE  $i$ : Return the calculated  $P_i^{(t)}$  for UE  $i$ .
24: end for
25: Return:  $P_i^{(t)}$  for UE  $i$ .

```

Performance Evaluation and Results

4.1 Simulation Environment

The project is ran on the Ubuntu Linux 64-bit system machine with 8GB of RAM, i5 Quad-core and 2.30GHz CPU frequency. The architecture of the framework consists of integrating two main software blocks, FA developed in python version 3.8.20 and NS-3 5G-LENA version 3.3 created in C++.

We configured 20 nodes to represent communication within a medium-sized institution with multiple departments or groups of clients and collaborators geographically distributed across various locations. In such organizations, communication may be typically both decentralized and structured, with various departments or groups of clients and collaborators exchanging important information. Using 20 nodes provides a realistic model of this type of organizational structure. This allows us to model different interaction methods, such as batch data transmission and priority messaging, helping us to analyze the impact of these variations on network performance and the efficiency of information exchange. It also introduces enough complexity to simulate communication challenges such as delays, reliability issues, and packet loss, while still keeping the simulation manageable.

The choice of 20 nodes is not limited to simulating an institution but also reflects a scenario where a group of institutions or individual researchers, located in different places, are collaborating on a specific project. For instance, in healthcare research, multiple hospitals or medical institutions in various cities could be working together on a project to predict and diagnose diseases such as cancer or cardiovascular conditions. For example, given Brazil's General Data Protection Law (LGPD), these institutions or researchers cannot share sensitive patient data directly, even in anonymized form. With FA, each institution or researcher can train their model locally on their own dataset, and only intermediate results, such as predictions or classifications are exchanged for aggregation. In this way, the 20 nodes represent the different institutions or researchers, each working on their own data but contributing to a collective effort without compromising data privacy.

We used the UMI-Street Canyon model, designed to simulate wireless communication in urban areas. As detailed in [38.901 2017], this model accounts for factors like buildings and obstacles that affect signal strength and quality in real-world urban environments. Specifically, the model incorporates both Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) conditions to reflect different communication scenarios that can occur in complex environments. LOS conditions occur when there is a direct path between the transmitter and receiver, resulting in clearer signals, while NLOS conditions arise when obstacles such as buildings block the signal path, causing attenuation and possible signal degradation. The model is well-suited for simulating decentralized communication structures, like those found in institutions with multiple departments, and it offers a more accurate representation of the challenges faced in wireless communication in these urban environments.

4.1.1 NS-3 and 5G-LENA Network Configuration

The NS-3 5G-LENA block is composed of 21 nodes, being 20 clients (UEs) and 1 base station (gNB) operating as the gateway to the FA server. As discussed in sections 2.2 and 2.3, NS-3 and 5G-LENA provide built-in models, protocols, and libraries for simulating different network scenarios. In our experiments, MobilityHelper class is used to position UEs and a gNB (acting as gateway). NodeContainer class manages simulation nodes, including gNB, UEs and other network elements. ThreeGppChannelModel simulates realistic propagation scenarios with path loss, shadowing and fading to accurately model the radio environment. NrHelper and NrPointToPointEpcHelper classes configure 5G New Radio (NR) devices and the core network (EPC), facilitating seamless communication between the gNB and the UEs. Devices are deployed using NetDeviceContainer, InternetStackHelper configures the IP stack for the devices, and IP addresses are assigned through Ipv4InterfaceContainer. The application layer is organized with a UdpServerHelper on a server attached to gNB and UdpClientHelper on remote clients (UEs). For traffic analysis, we use FlowMonitorHelper to collect performance metrics.

For resource scheduling, we configured a specific Bandwidth Part (BWP) for high throughput applications with the following parameter: `bwpIdForHighThroughput` to ensure that high throughput traffic is prioritized within a dedicated BWP, allowing for more efficient spectrum usage and reducing competition with other types of traffic. In addition, we configured the 5G-LENA to use a default scheduler, PF (Proportional Fair) scheduler, designed for high-throughput BWP, which guarantees a balance between throughput and fairness. This approach enables dynamic and efficient resource allocation within the BWP, to benefit throughput-sensitive applications. The implementation of BWPs and adaptive scheduling in our project, among others, directly impacts the fol-

lowing aspects:

- Higher throughput - as resources are concentrated and optimized for premium traffic.
- Reduced packet loss - due to dynamic adaptation to channel conditions.
- Improved spectral efficiency, ensuring optimized resource allocation.

4.1.2 Simulation parameters

The simulation uses various parameters that define the scenario, traffic, and network configurations. As shown in Figure 4.1, we position a 5G gNB in a fixed location where the FA server is installed and randomly distribute 20 UEs within a radius of 200 meters that must transmit data to the server. Other parameters are presented in Table 4.1.

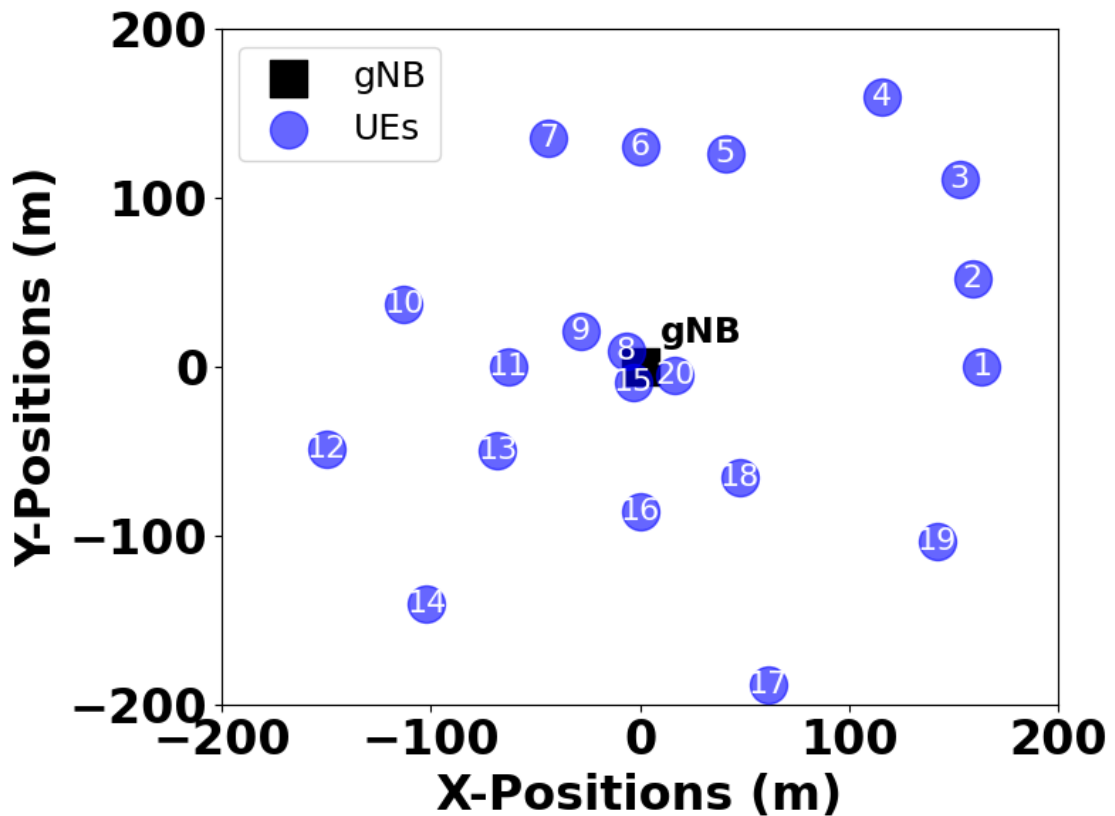


Figure 4.1: gNB and devices

In the simulation, each UE corresponds to a different department within the institution. The UEs are randomly placed in a 2-dimensional space around a fixed gNB, which is located at the center (origin) of the Cartesian coordinate system, at a height of 10 meters. This random placement of UEs simulates the geographic distribution of the institution's departments, reflecting their separation across different locations. This

Table 4.1: Simulation parameters

| Parameter | Value |
|-----------------------|-------------|
| OperationalBand | Single (HT) |
| UdpPacketSize | 1500 bytes |
| lambdaHT | 20000 pps |
| simTime | 1000 ms |
| udpAppStartTime | 400 ms |
| bandwidthBand | 400e6 Hz |
| numerologyBwp1 | 3 |
| centralFrequencyBand1 | 6.8e9 Hz |
| TxPower | [0, 30] dBm |

approach guarantees that all devices stay at the same position throughout the simulation and allows better control and fair comparison of the generated results in each configured scenario, while also accurately modeling the geographic layout of the departments.

The simulation varies transmission power levels between 0 and 30 dBm across three scenarios: (1) uniform maximum power allocation for all UEs, regardless of the distance to the gNB and channel conditions; (2) random power allocation for all UEs, irrespective of the distance to gNB and channel conditions; and (3) channel-aware power allocation algorithm. Transmission power values adhere to Anatel resolution [[Anatel - Resolução N° 747 2021](#)], which regulates wireless communication standards in Brazil. Each scenario undergoes 15 Monte Carlo simulations to ensure statistical reliability. Performance metrics (throughput, latency, packet loss, delivery ratio, and energy consumption) are collected, with averages and confidence intervals calculated and plotted for clear visualization. These results are used to assess the impact of network condition variations on the performance and quality of the final aggregated responses of the FA model, as well as the response times.

4.1.3 FA Configuration

The FA block consists of a central server that coordinates the system activities, including generating queries for clients, selecting clients for each round and aggregating responses received from clients. The FA server connects to the NS-3 core network through a RemoteHost, with the SGW handling data routing and acting as a mobility anchor for UEs, while the PGW serving as the gateway to external networks. The bidirectional communication between the server and clients in the network is established using communication sockets with parameters AF_INET for IPv4 and SOCK_STREAM for TCP on both sides. Two FA subalgorithms are implemented: one integrates with NS-3 5G-LENA to process network data and make decisions on clients selection, the other

includes Bayesian aggregation method for final predictions and a weighted averaging method to aggregate client performance results based on reliability.

The FA clients' models are trained using a neural network structured with two fully connected dense layers adapted for MNIST digits classification. The first layer consists of 512 neurons, and the second layer has 256 neurons. These neurons apply rectified linear unit (ReLU) activation functions within the hidden layers. During operation, ReLU transforms negative input values to zero while keeping positive values. This improves the model's ability to learn efficiently from the MNIST dataset. The final layer of the neural network has 10 neurons using a softmax activation function. This function is incorporated for the output layer of MNIST digits classification tasks, to transform the network's outputs into a probability distribution across the ten-digit classes (0-9). Softmax ensures that the sum of these probabilities equals one (normalize), enabling the model to provide calibrated probability predictions for each digit class. This configuration optimizes the neural network's performance, allowing it to accurately classify handwritten digits and effectively handle the complexities present in the MNIST dataset. The model is compiled using the Adam optimizer and the sparse categorical cross-entropy loss function, with accuracy, precision and recall as performance metrics. Differential privacy technique is applied to maintain privacy between the clients and the server, making sure that the information transmitted during communication does not reveal sensitive data about any individual client.

4.2 Simulation Results and Validation

The simulations produce two types of results: network performance statistics and FA prediction and performance results. The prediction results include number of selected clients and the final aggregated predicted class for each scenario. For testing, the classification query requires all clients to predict the image at index 193 of the MNIST dataset, which contains the handwritten number 9. The maximum number of clients in the system is 20. The network performance metrics include packet loss, throughput, latency, delivery ratio, and energy consumption. The FA performance results consist of accuracy, precision, and recall, while the prediction result is the output class.

4.2.1 Discussion of Results

To evaluate the impact of network conditions on FA performance, we compare both network and FA results across three established scenarios: (1) uniform maximum power allocation for all UEs, (2) random power allocation for all UEs, and (3) a channel-aware power allocation algorithm. In addition to comparing the scenarios, we validate the

results against the existing literature on FA and 5G networks. The following discussion presents a detailed analysis of the results obtained in the experiment.

Table 4.2: Client Selection and Prediction Results

| Scenario | Selected Clients | Selection Percentage | Predicted Class | Result |
|-----------|------------------|----------------------|-----------------|---------|
| Maximum | 12 | 60 % | -1 | Failure |
| Random | 9 | 45 % | -1 | Failure |
| Optimized | 20 | 100 % | 9 | Success |

The results of client selection and the corresponding aggregated predictions are presented in Table 4.2. In scenario 1, with a uniform maximum transmission power allocated, only 12 clients are selected (60 %), leading to a failure in class prediction. In scenario 2, the random power allocation strategy further degrades the performance, with only 9 clients selected (45 %) and a similar failure in prediction. The lack of consideration for channel conditions and UE distances likely resulted in weaker signals, increased packet loss, and ultimately, the transmission failure of a significant number of clients.

In scenario 3, applying a channel-aware power allocation algorithm that adjusts transmission power based on UE distance and channel conditions, results in the selection of 20 clients (100 %), thus covering the total number of clients in the system. The aggregated result correctly predicted the target class (9), leading to a successful FA outcome. These results suggest that this approach contributed, to some extent, to better signal strength, reduced packet loss, and enhanced overall FA performance.

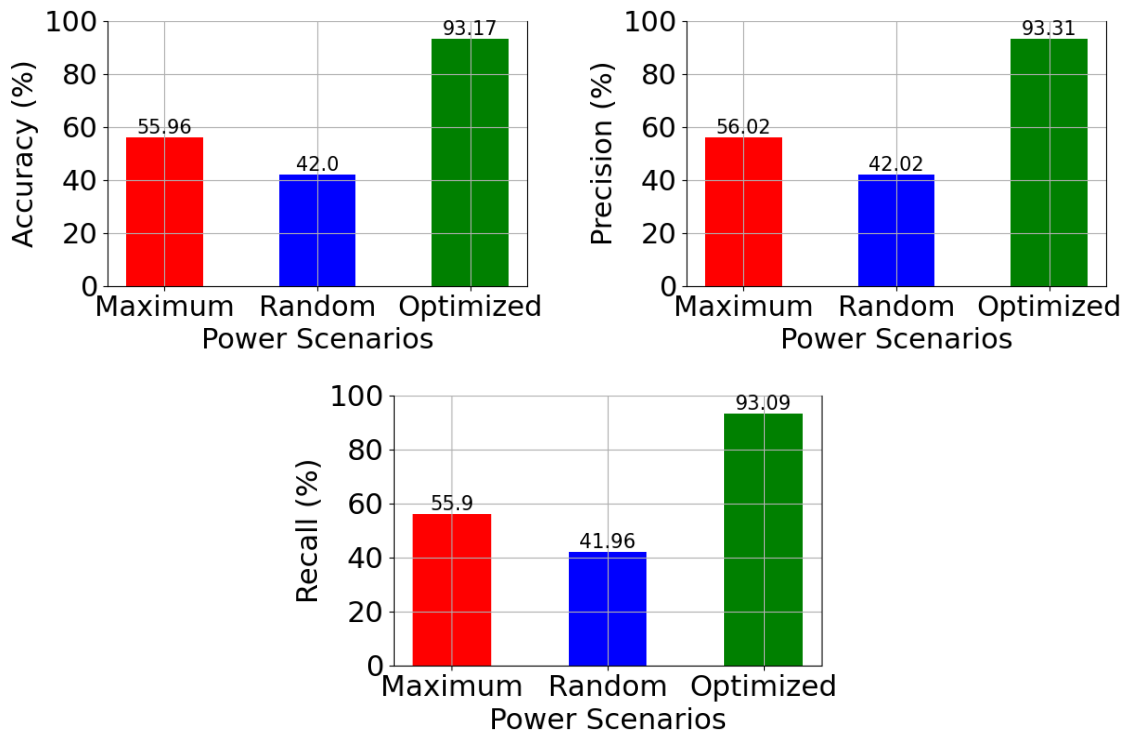


Figure 4.2: FA performance metrics for different power scenarios. Accuracy, Precision and Recall.

The results of FA simulation under 5G network conditions, given in Figure 4.2, show how network factors impact FA performance across three transmission power scenarios. Scenario 1, where maximum transmission power is uniformly applied for all UEs, FA achieves moderate performance results such as accuracy (55.96 %), precision (56.02 %), and recall (55.90 %). While this strategy is simple and straightforward, it has inherent limitations. By applying the same power level to all UEs, it disregards the varying conditions of each UE, such as distance from the gNB or channel conditions. As a result, UEs that are already in good conditions or near the gNB may receive excessive power, causing unnecessary interference and signal congestion. This inefficient energy use and potential signal degradation can degrade network performance, which in turn can negatively affect the performance and reliability of FA, especially in dynamic and dense 5G environments.

Scenario 2, which adopts a random power allocation strategy for UEs, results in even poorer FA performance, with accuracy (42.0 %), precision (42.02 %), and recall (41.96 %). Although this strategy incorporates some randomness, still does not account for network conditions, which may be crucial for the efficiency of FA systems. In FA, the transmission of responses and updates heavily depends on stable network conditions. From the observed results, random power allocation can lead to situations where some UEs are overpowered, causing unnecessary interference, while others may

struggle with inadequate power, leading to poor communication quality. This imbalance in power allocation may degrade the quality of the system response, affecting the overall performance of the FA operation. The observed decrease in accuracy and reliability demonstrates that random power allocation, without considering the specific needs of each UE, can significantly impair the performance of FA.

In contrast, scenario 3, using a channel-aware power allocation algorithm, that adjusts transmission power based on UE distance and channel conditions, produces better performance results, achieving 93.17 % accuracy, 93.31 % precision, and 93.09 % recall. This algorithm adjusts the transmission power dynamically based on each UE's distance from the gNB and the quality of the communication channel conditions. By considering these factors, this adaptive strategy not only ensures that each UE receives the appropriate power but also minimizes energy consumption and interference. This optimization directly contributes to the reduction of wasted energy and congestion, thus improving overall network efficiency. The improved network efficiency directly enhances FA performance, as reliable data transmission is required for accurate updates and responses aggregation. Compared to the first two scenarios, this algorithm minimizes communication delays and packet loss, ensuring that FA updates and responses are transmitted efficiently and without disruption. The improved FA performance compared to the first two scenarios demonstrates the importance of accounting for real-time network conditions in power allocation strategies. These findings show that network-aware power allocation has the potential to keep FA processes reliable. Adjusting power levels based on real-time network conditions can help improve FA performance in 5G networks.

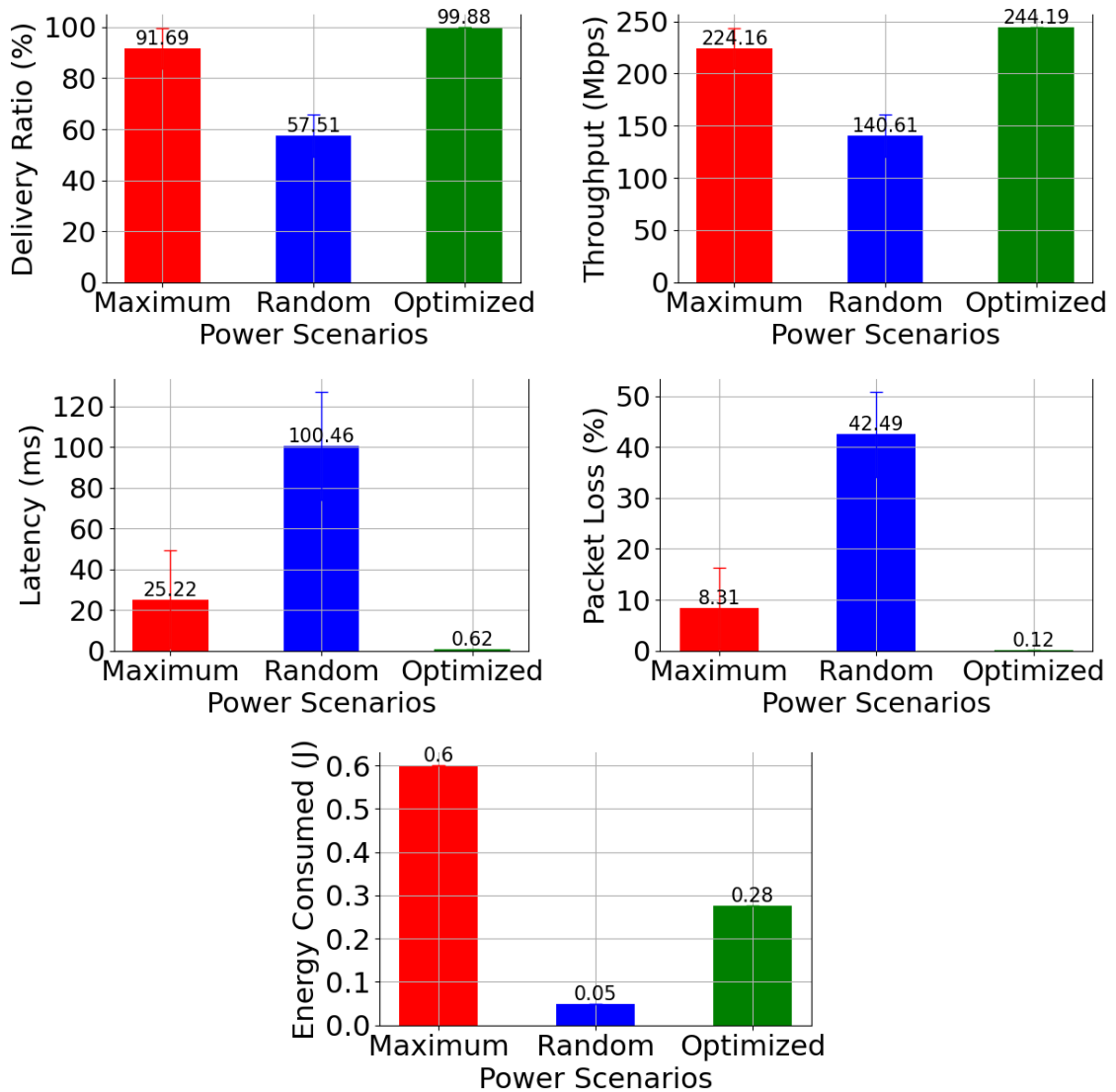


Figure 4.3: Network performance metrics for different power scenarios. Delivery ratio, Throughput, Latency, Packet loss and Energy consumed.

The simulation results of FA have revealed that network performance significantly influences the quality of the final aggregated answer and the overall FA performance. As given in Figure 4.3, in scenario 1, moderate network performance was observed, with a delivery ratio of 91.69 %, throughput of 224.16 Mbps, and latency of 25.22 ms. Although this configuration presents a relatively high delivery ratio, the lack of adaptation to real-time network conditions likely introduced some inefficiencies. Maximum transmission power allocation for UEs without considering network conditions might have led to interference and inefficiencies, resulting in 8.31 % packet loss, which likely reduced the number of successful responses sent to the central server, and consequently, affecting the reliability of the FA aggregation process. It is possible that retransmissions due to packet loss have contributed to this moderately high latency, further impacting the efficiency of FA. The inability to dynamically adjust transmission power suggests that this approach,

while straightforward, fails to optimize resource utilization in complex 5G environments.

Scenario 2 further degraded network performance. The delivery ratio dropped to 57.51 %, throughput decreased to 140.61 Mbps, and latency increased to 100.46 ms. The observed inconsistency caused by the random power allocation for UEs without considering network conditions might have resulted in severe packet loss (42.49 %), leading to a dramatic reduction in the number of responses reaching the server, hence, degrading the aggregated results. Again, the observed high latency could be attributed to frequent retransmissions caused by high packet loss, which further contributed to the increased latency.

In contrast, scenario 3 outperformed the other scenarios, with a delivery ratio of 99.88 %, throughput of 244.19 Mbps, and latency of 0.62 ms. This optimized allocation minimized packet loss to 0.12 %, ensuring that more intermediate responses successfully reached the server, which is fundamental for improving the final FA aggregated result. The algorithm also reduced energy consumption to 0.28 J, which highlights the efficiency of the approach. These results suggest that channel-aware optimization improves network performance, leads to more reliable data transmission and, consequently, more accurate and timely results in FA.

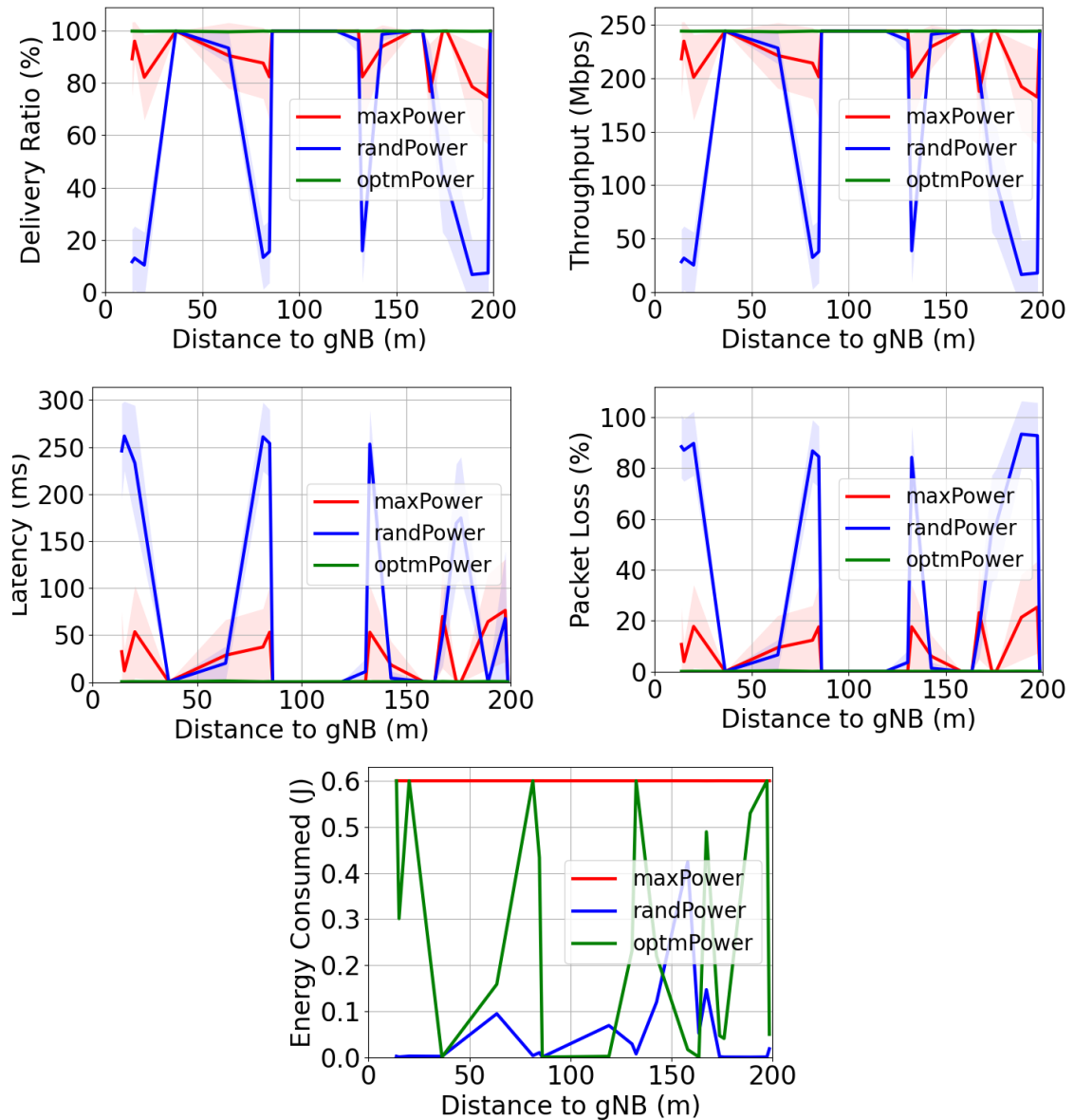


Figure 4.4: Network performance metrics in terms of distance. Delivery ratio, Throughput, Latency, Packet loss and Energy consumed.

Based on the results shown in Figure 4.4, within the 200 meters simulation coverage radius, scenario 3 demonstrates minimal variations in network performance, with fluctuations likely staying below 1. This may indicate that the adaptive power allocation approach is effective in maintaining stability by adjusting transmission power based on real-time conditions such as distance, channel quality, and signal strength. By adapting to these factors, it helps maintain a stable communication link and reduce inefficiencies caused by overpowering or underpowering nodes.

In contrast, scenario 1 and scenario 2 show larger performance variations. Scenario 1 experiences moderate fluctuations, which suggests some level of stability but still lacks the responsiveness seen in scenario 3. Meanwhile, scenario 2 shows even

greater instability, likely because it lacks any dynamic adjustment, leading to inconsistent performance. The larger variations in scenario 1 and scenario 2 can indicate that the static or random power allocation used in these scenarios does not properly account for real-time network conditions, resulting in less consistent performance.

This comparison highlights the potential of the channel-aware approach in scenario 3, as it provides more stable and reliable performance by adjusting transmission power in response to varying conditions. The results suggest that incorporating real-time power adaptation, as demonstrated in scenario 3, can improve the consistency and reliability of network performance, which is critical for better FA operation.

4.2.2 Statistical Analysis

For statistical analysis of the impact of the three established transmission power allocation scenarios (maximum power, random power, and optimized power (channel-aware algorithm)) on the network and FA performance, a Shapiro-Wilk normality test was first performed on the data from the experiment results to check if the underlying distributions of the network performance metrics (delivery ratio, throughput, latency, packet loss and energy consumed) and the FA performance metrics (accuracy, precision, and recall) followed a normal distribution. Based on the results of the normality test, it was determined that the data for the FA metrics followed a normal distribution, while the network data did not follow a normal distribution. The p-values from the normality tests for FA data were sufficiently high (above 0.05), indicating that the assumption of normality was not violated. However, the p-values from the normality tests for network data were sufficiently low (below 0.05), leading to violation of the normality assumption. Given that the FA data passed this test, the next step was to conduct parametric ANOVA tests and Tukey's Honest Significant Difference (HSD) tests for post-hoc analysis. The Tukey's test was chosen due to the normality of the data and its ability to compare the means of multiple groups while controlling for Type I error. Since the network data failed the normality test, their non-parametric counterparts, the Kruskal-Wallis tests and the Dunn tests, were performed for post-hoc analysis. All statistical analyses were conducted using Python modules, including scipy, scikit-posthocs and statsmodels. For each pairwise comparison between the transmission power scenarios, the following hypotheses were tested:

- Null Hypothesis (H_0): The optimized power strategy (Channel-aware power allocation algorithm) does not significantly improve the network and FA performance compared to the maximum power and random power allocation scenarios. Mathematically can be expressed as $H_0 : \mu_{optm} = \mu_{max}$ and $\mu_{optm} = \mu_{rand}$.

- Alternative Hypothesis (H_1): The optimized power strategy (Channel-aware power allocation algorithm) significantly improves the network and FA performance compared to the Maximum power or random power allocation scenarios. Mathematically can be expressed as $H_1 : \mu_{optm} \neq \mu_{max}$ OR $\mu_{optm} \neq \mu_{rand}$.

Where:

- μ_{optm} → represents the mean of network or FA performance metrics under optimized power allocation scenario.
- μ_{max} → represents the mean of network or FA performance metrics under maximum power allocation scenario.
- μ_{rand} → represents the mean of network or FA performance metrics under random power allocation scenario.

In parametric tests, the the mean difference between each pair of scenarios was computed, along with p-values to evaluate the null hypothesis. The p-value was compared to the significance level (0.05). If the p-value was below 0.05, the null hypothesis (no difference between means) was rejected, indicating that there was a statistically significant difference between the two scenarios. The parametric test also provided confidence intervals (lower and upper) for mean differences to show the range within which the true difference in means is likely to lie. In non-parametric test, besides the p-value, the test statistic based on the rank sums of the data, was computed to provide information about the magnitude of the difference between the scenarios being compared, relative to the variability within the scenarios. The test statistic was applied to determine whether there are significant differences between the scenario pairs by considering the magnitude of those differences relative to the expected variation within the scenarios. The following discussion presents a detailed statistical analysis of the results obtained in the experiment.

Table 4.3: Statistical Significance for Accuracy

| Comparison Pairs | MeanDiff. | p-value Adj. | Lower | Upper | Reject |
|-----------------------|-----------|--------------|---------|---------|--------|
| Maximum and Optimized | 0.3722 | 0.000 | 0.3715 | 0.3728 | True |
| Maximum and Random | -0.1395 | 0.000 | -0.1402 | -0.1389 | True |
| Optimized and Random | -0.5117 | 0.000 | -0.5124 | -0.5111 | True |

Table 4.4: Statistical Significance for Precision

| Comparison Pairs | MeanDiff. | p-value Adj. | Lower | Upper | Reject |
|-----------------------|-----------|--------------|---------|---------|--------|
| Maximum and Optimized | 0.3729 | 0.000 | 0.3724 | 0.3733 | True |
| Maximum and Random | -0.14 | 0.000 | -0.1404 | -0.1396 | True |
| Optimized and Random | -0.5129 | 0.000 | -0.5133 | -0.5124 | True |

Table 4.5: Statistical Significance for Recall

| Comparison Pairs | MeanDiff. | p-value Adj. | Lower | Upper | Reject |
|-----------------------|-----------|--------------|---------|---------|--------|
| Maximum and Optimized | 0.3719 | 0.000 | 0.3712 | 0.3725 | True |
| Maximum and Random | -0.1394 | 0.000 | -0.1401 | -0.1388 | True |
| Optimized and Random | -0.5113 | 0.000 | -0.5119 | -0.5106 | True |

The Tables 4.3, 4.4 and 4.5 show the statistical significance of differences in accuracy, precision and recall across the three power allocation scenarios. In the context of the post-hoc analysis using Tukey's HSD test, the results indicate significant differences between all pairs of power allocation scenarios, with the "Reject" column marked as "True" for each pairwise comparison. This suggests that the FA performance outcomes (accuracy, precision, recall) for each of the three power allocation strategies (maximum power, random power, and optimized power) are significantly different from each other.

Taking the optimized power allocation scenario (channel-aware power allocation algorithm) as reference, we compared its performance with the other two scenarios. Statistical analysis showed that the significance value between the optimized and maximum power allocation scenarios is $p\text{-value} = 0.000$, indicating a significant difference on all metrics (accuracy, precision, recall). This suggests that the channel-aware power allocation algorithm significantly outperforms the maximum power allocation. Note that The maximum power allocation applies the same maximum transmission power to all UEs regardless of their location or channel conditions, which can lead to inefficient energy usage and excessive interference for some UEs, especially those closer to the gNB. By contrast, the optimized power strategy adjusts transmission power dynamically based on each UE's distance from the gNB and the quality of the communication channel, ensuring more efficient power usage and reducing interference. This improvement could be reflected in the better performance results achieved in the optimized scenario.

Furthermore, the significance value between optimized and random power allocation scenarios is $p\text{-value} = 0.000$, which also indicates a significant difference (accuracy, precision, recall). The random power allocation strategy introduces variability in power distribution, but still does not account for the specific needs of UEs based on their network conditions. This can lead to a poor power distribution where some UEs might receive excessive power, causing interference, while others may experience inadequate power, leading to poor communication quality. The channel-aware power allocation algorithm, on the other hand, adjusts power according to real-time network conditions, leading to better overall performance. This approach has the ability to reduce unnecessary energy consumption and signal congestion, thus improving network efficiency and the reliability of the FA process.

Table 4.6: Statistical Significance for Delivery Ratio

| Comparison Pairs | p-value Adj. | Test Statistic | Reject |
|-----------------------|--------------|----------------|--------|
| Maximum and Random | 0.123718 | 12.236742 | False |
| Maximum and Optimized | 0.449844 | 12.236742 | False |
| Optimized and Random | 0.001500 | 12.236742 | True |

Table 4.7: Statistical Significance for Throughput

| Comparison Pairs | p-value Adj. | Test Statistic | Reject |
|-----------------------|--------------|----------------|--------|
| Maximum and Random | 0.123718 | 12.236742 | False |
| Maximum and Optimized | 0.449844 | 12.236742 | False |
| Optimized and Random | 0.001500 | 12.236742 | True |

Table 4.8: Statistical Significance for Latency

| Comparison Pairs | p-value Adj. | Test Statistic | Reject |
|-----------------------|--------------|----------------|--------|
| Maximum and Random | 0.161172 | 8.073909 | False |
| Maximum and Optimized | 1.000000 | 8.073909 | False |
| Optimized and Random | 0.016749 | 8.073909 | True |

Table 4.9: Statistical Significance for Packet Loss

| Comparison Pairs | p-value Adj. | Test Statistic | Reject |
|-----------------------|--------------|----------------|--------|
| Maximum and Random | 0.127876 | 12.135333 | False |
| Maximum and Optimized | 0.449896 | 12.135333 | False |
| Optimized and Random | 0.001579 | 12.135333 | True |

Table 4.10: Statistical Significance for Energy Consumed

| Comparison Pairs | p-value Adj. | Test Statistic | Reject |
|-----------------------|------------------|----------------|--------|
| Maximum and Random | $4.129562e - 04$ | 36.527027 | True |
| Maximum and Optimized | $7.225302e - 09$ | 36.527027 | True |
| Optimized and Random | $9.343858e - 02$ | 36.527027 | False |

The results of the statistical significance analysis for the network performance metrics across the three power allocation scenarios are presented in Tables 4.6, 4.7, 4.8, 4.9 and 4.10 . From the observed results, for delivery ratio, throughput, latency, and

packet loss, the reject column indicates a false value when comparing maximum power allocation to either random or optimized (channel-aware algorithm) power allocation scenarios. This means that no significant difference was found between maximum and random or maximum and optimized power allocation in these metrics. However, when comparing the optimized power allocation to the random power allocation, the reject column shows true, indicating that the optimized power allocation significantly improves the network performance for these metrics. This outcome reveals that the channel-aware power allocation algorithm, which adjusts transmission power dynamically based on network conditions, can be more efficient in minimizing interference and ensuring better power distribution compared to the random power allocation scenario, leading to improved performance in terms of delivery ratio, throughput, latency, and packet loss.

On the other hand, for energy consumption, the reject column shows true when comparing both maximum and random, as well as maximum and optimized power allocation, but false when comparing optimized and random power allocation. This indicates that significant differences in energy consumption are observed when comparing maximum and random power allocation, and maximum and optimized power allocation, but not between optimized and random power allocation. The maximum power allocation scenario, which applies a maximum fixed power level to all UEs regardless of their specific needs, leads to higher energy consumption, especially for UEs that may not require as much power. In contrast, both random and optimized power allocation strategies introduce variability in power distribution, but the optimized scenario aims to minimize unnecessary energy usage by dynamically adjusting the transmission power. However, the difference in energy consumption between optimized and random power allocation is not statistically significant, suggesting that while the optimized power allocation is more efficient than the maximum power allocation, the energy usage between the channel-aware power algorithm and random power allocation scenario does not differ significantly. This indicates that, in terms of energy consumption, random allocation can be considered to be almost as efficient as the optimized scenario, although the channel-aware power allocation algorithm significantly outperforms the random scenario in other network performance metrics.

Conclusion and Future Research Works

5.1 Conclusion

In this study, we simulated the FA-5GLENA framework, integrating Federated Analytics (FA) with the 5G-LENA simulator in NS-3. With the goal to analyze the impact of different network conditions on FA performance, the framework applied two algorithms: one responsible for the integration (FA-5GLENA integrated algorithm), another for optimizing transmission power allocation (channel-aware dynamic power allocation algorithm). The simulation involved multiple User Equipment (UEs-clients) communicating with FA server through a gNB under varying transmission power levels. The FA process used the MNIST dataset, where network conditions influenced client selection for training and classification. The FA server aggregated results using two algorithms: one based on Bayesian probability for prediction and another using a weighted average for performance metrics.

We tested three power allocation scenarios: uniform maximum power, random power, and channel-aware dynamic power allocation. The results showed that channel-aware power control outperformed the other scenarios in terms of network stability, reliability, and efficiency. Dynamically adjusting power based on network conditions like UE distance, channel quality, and signal strength led to more reliable communication and efficient data transmission to the FA server. This improved the accuracy and reliability of the aggregated result in FA.

On the other hand, static or random power allocation caused higher packet loss, latency, and inefficient energy use, thereby reducing FA performance. The channel-aware algorithm minimized these issues, ensuring that more responses reached the server with lower energy consumption. This study demonstrates the importance of intelligent power control in FA applications within 5G networks. The results suggest that transmission power should not be fixed but adjusted based on real-time network conditions. This research lays the foundation for future work in optimizing FA frameworks for large-scale 5G environments, and development of new algorithms.

The statistical analysis of FA performance metrics using Tukey's HSD test indicated significant differences in accuracy, precision, and recall across the three power allocation scenarios. The optimized power allocation scenario showed significantly higher performance metrics compared to both the maximum and random power allocation scenarios. These findings demonstrate that channel-aware power allocation algorithm directly impacts FA performance, with the optimized algorithm achieving statistically superior results.

The statistical analysis of network performance metrics using Dunn's test across the three power allocation scenarios indicated varying degrees of significance. No significant differences were found between maximum power allocation and either random or optimized power allocation for delivery ratio, throughput, latency, and packet loss. However, the optimized power allocation scenario showed a statistically significant improvement over the random power allocation in these metrics. Regarding energy consumption, significant differences were observed when comparing maximum power allocation to both random and optimized scenarios, but not between optimized and random power allocation. This indicates that while both scenarios have reduced energy consumption compared to the maximum power approach, their efficiency in minimizing energy usage does not differ significantly.

5.2 Future Research Works

While this research provides a strong foundation for integrating network-aware power allocation into FA, several areas may require further exploration to refine and optimize this approach:

- **Integration of Additional Network Metrics:** Future studies could incorporate more network parameters, such as UE mobility, network congestion, interference levels, retransmissions, RTT, link throughput, and other modulation schemes, to refine the power allocation algorithm and enhance adaptability in dynamic environments.
- **Scalability and Robustness Analysis:** Evaluating the algorithm under different conditions, such as varying UE densities, changing interference patterns, and dynamic network loads, could provide deeper insights into its scalability and robustness in large-scale 5G deployments.
- **Energy-Efficient Strategies:** While this study demonstrated a reduction in power consumption, further optimization of energy-efficient transmission strategies is necessary. Exploring low-power transmission techniques, adaptive duty cycles, or more advanced AI-driven power control, such as network data analytics function (NWDAF) and other state-of-the-art (SOTA) algorithms could minimize the environmental and operational costs of FA in 5G networks.

- **Resource Allocation Management:** Given the variety of schedulers available in the 5G-LENA simulator, such as Round Robin (RR), Quality of Service (QoS) Scheduling, Maximum Rate (MR), and others, exploring and implementing these schedulers could help better understand their impact on network performance and further optimize resource allocation in 5G simulations.
- **Real-World Validation:** Extending the study to field experiments with varying UE positions, traffic loads, and real-world interference will help validate the proposed approach under practical conditions. This would ensure that the algorithm is effective in real-time, large-scale, and dynamically changing network environments.

References

- [38.901 2017]38.901, G. T. *Study on channel model for frequencies from 0.5 to 100 GHz*. [S.I.]: 3GPP Sophia Antipolis, France, 2017.
- [Anatel - Resolução N^o 747 2021]ANATEL - Resolução N^o 747. 2021. Disponível em: <<https://informacoes.anatel.gov.br/legislacao/resolucoes/2021/1593-resolucao-747>>.
- [Beksultanova, Vatyukova e Yalmaeva 2020]BEKSULTANOVA, A.; VATYUKOVA, O. Y.; YALMAEVA, M. Application of digital technologies in the educational process. In: ATLANTIS PRESS. *2nd International Scientific and Practical Conference on Digital Economy (ISCDE 2020)*. [S.I.], 2020. p. 82–87.
- [Ceballos et al. 2021]CEBALLOS, H. Z. et al. *Wireless Network Simulation: A Guide Using Ad Hoc Networks and the Ns-3 Simulator*. Apress, 2021. ISBN 978-1-4842-6848-3 978-1-4842-6849-0. Disponível em: <<https://link.springer.com/10.1007/978-1-4842-6849-0>>.
- [Chen et al. 2021]CHEN, D. et al. *Digital twin for federated analytics using a Bayesian approach*. [S.I.]: IEEE, 2021. 16301–16312 p.
- [Chen et al. 2021]CHEN, M. et al. A Joint Learning and Communications Framework for Federated Learning Over Wireless Networks. *IEEE Transactions on Wireless Communications*, v. 20, n. 1, p. 269–283, 2021.
- [Comer 2011]COMER, D. *Computer Networks and Internets*. Pearson Education, 2011. ISBN 978-0-13-300210-2. Disponível em: <<https://books.google.com.br/books?id=ZSorAAAAQBAJ>>.
- [CTTC-LENA 2024]CTTC-LENA. *5G-LENA Module*. jul 2024. Accessed on December 26, 2024. Disponível em: <<https://5g-lena.cttc.es/>>.
- [Damsgaard et al. 2023]DAMSGAARD, H. J. et al. Approximate computing in B5G and 6G wireless systems: A survey and future outlook. v. 233, p. 109872, 2023. ISSN 13891286. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1389128623003171>>.

- [Elkordy et al. 2023]ELKORDY, A. R. et al. Federated Analytics: A Survey. v. 12, n. 1, 2023. ISSN 2048-7703. Disponível em: <<http://www.nowpublishers.com/article/Details/SIP-2022-0063>>.
- [Energy Model - Nsnam 2024]ENERGY Model - Nsnam. 2024. Disponível em: <https://www.nsnam.org/wiki/Energy_model/Project_Goals>.
- [ETSI 2022]ETSI. Multi-Access Edge Computing (MEC) Framework and Reference Architecture. *ETSI Standard GS MEC 003*, 2022.
- [Fachola et al. 2023]FACHOLA, C. et al. Federated Learning for Data Analytics in Education. v. 8, n. 2, p. 43, 2023. ISSN 2306-5729. Disponível em: <<https://www.mdpi.com/2306-5729/8/2/43>>.
- [Fernández et al. 2012]FERNÁNDEZ, J. et al. Adjustments of Log-Distance Path Loss Model for Digital Television in Lima. In: *Anais de XXX Simpósio Brasileiro de Telecomunicações*. Sociedade Brasileira de Telecomunicações, 2012. Disponível em: <<http://biblioteca.sbrt.org.br/articles/349>>.
- [Ghafar 2024]GHAFAR, Z. Digital Technologies in Health Care: A Comprehensive Review of Current Status and Future Perspectives. v. 2, n. 3, p. 353–370, 2024. ISSN 3025-7670. Disponível em: <<https://journal.multitechpublisher.com/index.php/ijasr/article/view/1558>>.
- [Han et al. 2022]HAN, Q. et al. PCFed: Privacy-Enhanced and Communication-Efficient Federated Learning for Industrial IoTs. v. 18, n. 9, p. 6181–6191, 2022. ISSN 1551-3203, 1941-0050. Disponível em: <<https://ieeexplore.ieee.org/document/9740428/>>.
- [Hapanchak e Costa 2022]HAPANCHAK, V. S.; COSTA, A. D. Emulation of Multipath Solutions in Heterogeneous Wireless Networks Over Ns-3 Platform. In: JIANG, D.; SONG, H. (Ed.). *Simulation Tools and Techniques*. Springer International Publishing, 2022. v. 424, p. 3–18. ISBN 978-3-030-97123-6 978-3-030-97124-3. Disponível em: <https://link.springer.com/10.1007/978-3-030-97124-3_1>.
- [Imoize et al. 2021]IMOIZE, A. L. et al. 6G Enabled Smart Infrastructure for Sustainable Society: Opportunities, Challenges, and Research Roadmap. v. 21, n. 5, p. 1709, 2021. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/21/5/1709>>.
- [Kadir et al. 2021]KADIR, E. A. et al. B5G and 6G: Next Generation Wireless Communications Technologies, Demand and Challenges. In: *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*. IEEE, 2021. p. 1–6. ISBN 978-1-66541-224-7. Disponível em: <<https://ieeexplore.ieee.org/document/9493470/>>.

- [Kairouz et al. 2021]KAIROUZ, P. et al. *Advances and Open Problems in Federated Learning*. 2021. Disponível em: <<http://arxiv.org/abs/1912.04977>>.
- [Khaled, Ajib e Mcheick 2022]KHALED, Z. E.; AJIB, W.; MCHEICK, H. Log Distance Path Loss Model: Application and Improvement for Sub 5 GHz Rural Fixed Wireless Networks. v. 10, p. 52020–52029, 2022. ISSN 2169-3536. Disponível em: <<https://ieeexplore.ieee.org/document/9755972/>>.
- [Koutlia et al. 2022]KOUTLIA, K. et al. Calibration of the 5g-lena system level simulator in 3gpp reference scenarios. *Simulation Modelling Practice and Theory*, Elsevier, v. 119, p. 102580, 2022.
- [Kurose e Ross 2021]KUROSE, J.; ROSS, K. *Computer Networking: A Top-Down Approach, Global Edition*. Pearson Education, 2021. ISBN 978-1-292-40551-3. Disponível em: <<https://books.google.com.br/books?id=374zEAAAQBAJ>>.
- [Kurt et al. 2021]KURT, G. et al. *A Vision and Framework for the High Altitude Platform Station (HAPS) Networks of the Future*. 2021. Disponível em: <<http://arxiv.org/abs/2007.15088>>.
- [Larrañaga et al. 2023]LARRAÑAGA, A. et al. An open-source implementation and validation of 5g nr configured grant for urllc in ns-3 5g lena: A scheduling case study in industry 4.0 scenarios. *Journal of Network and Computer Applications*, Elsevier, v. 215, p. 103638, 2023.
- [Li 2016]LI, H. *Communications for control in cyber physical systems: theory, design and applications in smart grids*. [S.l.]: Morgan Kaufmann, 2016.
- [Lin et al. 2020]LIN, Y. et al. *Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training*. 2020. Disponível em: <<http://arxiv.org/abs/1712.01887>>.
- [Luo, Li e Wang 2018]LUO, X.; LI, H.; WANG, X. Wireless Information and Energy Transfer in MIMO Communication With Interference Channels. v. 6, p. 65845–65861, 2018. ISSN 2169-3536. Disponível em: <<https://ieeexplore.ieee.org/document/8502928/>>.
- [Luo et al. 2018]LUO, Y. et al. A Novel Mobile and Hierarchical Data Transmission Architecture for Smart Factories. v. 14, n. 8, p. 3534–3546, 2018. ISSN 1551-3203, 1941-0050. Disponível em: <<https://ieeexplore.ieee.org/document/8333734/>>.
- [Maurseth 2018]MAURSETH, P. B. The effect of the Internet on economic growth: Counter-evidence from cross-country panel data. v. 172, p. 74–77, 2018. ISSN 01651765. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0165176518303537>>.

- [Miragem 2019]MIRAGEM, B. A lei geral de proteção de dados (lei 13.709/2018) e o direito do consumidor. *Revista dos Tribunais*, v. 1009, 2019.
- [nnsnam]nnsnam. *Ns-3*. Disponível em: <<https://www.nnsnam.org/>>.
- [Oliveira-Jr et al. 2020]OLIVEIRA-JR, A. et al. IoT Sensing Platform as a Driver for Digital Farming in Rural Africa. v. 20, n. 12, p. 3511, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/12/3511>>.
- [Oyadomari, Costa e Ribeiro 2023]OYADOMARI, W.; COSTA, R. S.; RIBEIRO, M. M. Personal data protection: Privacy and reliability in the digital environment. 2023.
- [Parra-Ullauri et al. 2024]PARRA-ULLAURI, J. M. et al. Federated Analytics for 6G Networks: Applications, Challenges, and Opportunities. v. 38, n. 2, p. 9–17, 2024. ISSN 0890-8044, 1558-156X. Disponível em: <<https://ieeexplore.ieee.org/document/10401940/>>.
- [Qadir et al. 2023]QADIR, Z. et al. Towards 6G Internet of Things: Recent advances, use cases, and open challenges. v. 9, n. 3, p. 296–312, 2023. ISSN 24059595. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405959522000959>>.
- [Ramachandran e Tsokos 2009]RAMACHANDRAN, K. M.; TSOKOS, C. P. *Mathematical Statistics with Applications*. [S.l.]: Academic Press, 2009. ISBN 978-0-12-374848-5.
- [Salahdine, Han e Zhang 2023]SALAHDINE, F.; HAN, T.; ZHANG, N. 5G, 6G, and Beyond: Recent advances and future challenges. v. 78, n. 9, p. 525–549, 2023. ISSN 1958-9395. Disponível em: <<https://doi.org/10.1007/s12243-022-00938-3>>.
- [Santos 2021]SANTOS, C. O. d. Lei geral de proteção de dados pessoais (lgpd), lei nº 13.709/2018: direito à privacidade aplicada às redes sociais. Pontifícia Universidade Católica de Goiás, 2021.
- [Shi et al. 2022]SHI, S. et al. Federated Anomaly Analytics for Local Model Poisoning Attack. v. 40, n. 2, p. 596–610, 2022. ISSN 0733-8716, 1558-0008. Disponível em: <<https://ieeexplore.ieee.org/document/9566475/>>.
- [Tarokh et al. 2009]TAROKH, V. et al. *New directions in wireless communications research*. [S.l.]: Springer, 2009.
- [Toka et al. 2023]TOKA, L. et al. 5G on the roads: Latency-optimized Federated Analytics in the Vehicular Edge. *IEEE Access*, IEEE, 2023.
- [Wang et al. 2022]WANG, D. et al. Federated Analytics: Opportunities and Challenges. v. 36, n. 1, p. 151–158, 2022. ISSN 0890-8044, 1558-156X. Disponível em: <<https://ieeexplore.ieee.org/document/9617564/>>.

- [Wang et al. 2024]WANG, Z. et al. A survey on federated analytics: Taxonomy, enabling techniques, applications and open issues. *arXiv preprint arXiv:2404.12666*, 2024.
- [Yan et al. 2023]YAN, F. et al. Insights into security and privacy issues in smart healthcare systems based on medical images. v. 78, p. 103621, 2023. ISSN 22142126. Disponible em: <<https://linkinghub.elsevier.com/retrieve/pii/S2214212623002053>>.
- [You et al. 2021]YOU, X. et al. Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. v. 64, n. 1, p. 110301, 2021. ISSN 1674-733X, 1869-1919. Disponible em: <<http://link.springer.com/10.1007/s11432-020-2955-6>>.
- [Yue, Kontar e Gómez 2024]YUE, X.; KONTAR, R. A.; GÓMEZ, A. M. E. Federated Data Analytics: A Study on Linear Models. v. 56, n. 1, p. 16–28, 2024. ISSN 2472-5854, 2472-5862. Disponible em: <<http://arxiv.org/abs/2206.07786>>.
- [Zhao et al. 2022]ZHAO, L. et al. Communication-Efficient Semihierarchical Federated Analytics in IoT Networks. v. 9, n. 14, p. 12614–12627, 2022. ISSN 2327-4662, 2372-2541. Disponible em: <<https://ieeexplore.ieee.org/document/9659816/>>.
- [Zhao et al. 2024]ZHAO, L. et al. A Decentralized Communication-Efficient Federated Analytics Framework for Connected Vehicles. p. 1–6, 2024. ISSN 1939-9359. Disponible em: <<https://ieeexplore.ieee.org/abstract/document/10478150>>.
- [Zhu et al. 2020]ZHU, F. et al. Parallel Transportation Systems: Toward IoT-Enabled Smart Urban Traffic Control and Management. v. 21, n. 10, p. 4063–4071, 2020. ISSN 1524-9050, 1558-0016. Disponible em: <<https://ieeexplore.ieee.org/document/8824092/>>.