Trustworthy Aggregation for Federated Learning as a Service Over Wireless Networks

by

Mohamed Ads

A thesis submitted to the
School of Computing
in conformity with the requirements for
the degree of Master of Science

Queen's University
Kingston, Ontario, Canada
June 2024

Dedication

For the sake of Allah

to my parents, and beloved family

to my grandmother may Allah forgive her

Abstract

This thesis investigates the challenges of implementing Federated Learning (FL) in diverse wireless edge networks, focusing on mitigating the impact of device heterogeneity, communication impairments, and trustworthiness concerns. Our research addresses the limitations of traditional FL approaches, which often suffer from slow convergence and reduced accuracy due to varying data quality, transmission rates, and error-prone channels. we propose novel risk-aware and accelerated FL frameworks that leverage device trustworthiness metrics and dynamic aggregation schemes. By classifying clients based on location-dependent performance and trustworthiness profiles, the frameworks prioritize participation from devices with high transmission rates while progressively incorporating data from cell-edge clients to enhance data diversity. In Addition, the frameworks employ debiasing techniques to account for transmission errors. To address the issue of unreliable clients, we explore two distinct validation approaches. The first approach utilizes a dedicated validation dataset to identify and eliminate untrustworthy clients, ensuring data integrity and model accuracy. Given the potential lack of such a dataset, the second approach investigates alternative mechanisms without relying on a separate validation dataset. The effectiveness of these frameworks and validation mechanisms is demonstrated through extensive simulations in a range of wireless settings, including conventional terrestrial

cellular networks and emerging 6G non-terrestrial networks with unmanned aerial vehicles. Results showcase the superior performance of the proposed frameworks compared to conventional FL approaches across these diverse environments, highlighting their adaptability and robustness.

Acknowledgments

With deep gratitude, I first thank Allah, the Most Gracious and Most Merciful, for granting me these achievements. This master's journey would not have been possible without His endless blessings and guidance.

I extend my heartfelt appreciation to Dr. Hesham ElSawy for his unwavering trust and support throughout this entire journy. His endless efforts, guidance, and encouragement have been instrumental in my success. I am also deeply grateful to Prof. Hossam Hassanein for his invaluable help and insightful guidance, which significantly enriched my research experience.

To my beloved family, especially my father and mother, your endless support and belief in me have been my constant source of strength. You made the distance feel insignificant and ensured I never felt alone during this journey. Thank you for everything.

My deepest thanks go to my dear friends, Mahmoud AL Nahhas, Mahmoud Atef, and Aoubaida AlSabbagh, Mohamed Serag and all my fiends from KFUPM. Your endless support and belief in me, especially during moments of self-doubt, have been invaluable. Your friendship has been a true blessing.

To my friend Omar Naserallah, who started this Master's journey with me and stood by my side until its completion, I extend my sincere gratitude. Together, we

faced numerous challenges, but thanks to Allah's grace, we overcame them all. Your companionship and support throughout this journey have been invaluable.

I would also like to express my gratitude to my friends Mahmoud Shawki, Hesham Abbas, Ahmed Abdulshakoor, Ahmed Khaled, Osama Ali, and all other friends of 75-Durham and An Clachan. Thank you for the countless valuable memories, shared laughter, and the delicious meals we cooked together. You made this experience truly special.

Finally, I offer my heartfelt thanks to my companions in the TRL and SWiMS labs, with special recognition to Sherif Azmy. The knowledge gained and the memories shared during our time together have been truly priceless, enriching both my academic and personal journey.

Statement Of Originality

I hereby certify that this MSc thesis is original and that all ideas and inventions attributed to others have been properly referenced.

Contents

List of Tables

List of Figures

List of Abbreviations

AS	Aggregator Server
BS	Base Station
FLaaS	Federated Learning as a Service
LOS	Line of Sight
ML	Machine Learning
NLOS	Non-Line of Sight
PPP	Poisson Point Process
RB	Resource Block
SINR	Signal-to-Interference-Plus-Noise-Ratio
TM	Trustworthiness Metric
UAV	Unmanned Aerial Vehicle

List of Symbols

B	Bandwidth
p_n	Dataset Ratio
μ	Dynamic SINR threshold
$\boldsymbol{g}$	Global Model Weights
T	Global Rounds
r	Horizontal Distance
$\mathcal{L}$	Laplace Transform
γ	Learning Rate
E	Local Epochs
$\boldsymbol{w}$	Local Model Weights
κ	Lower TM threshold
K	Number of Associated Users
ν	TM Scores Mean
Ω	Trust Metric Score
ρ	Upper TM threshould
n	User's Index
U	User's Set
h	Vertical distance

List of Publications

Conference Publications

M. Ads, H. ElSawy and H.S. Hassanein, "Risk-Aware Accelerated Wireless Federated Learning with Heterogeneous Clients," *ICC 2024 - IEEE International Conference on Communications*, Denver, USA, (Accepted)

M. Ads, H. ElSawy, H.M. Abbas, and H.S. Hassanein, "Trustworthy Aggregation for Aerial Federated Learning in Heterogeneous Client Environments," *ICCSPA 2024 - IEEE International Conference on Communications, Signal Processing, and their Applications*, ISTANBUL, Türkiye, (Accepted)

Journal Publications

M. Ads, H. ElSawy and H.S. Hassanein, "RARE-FL: Resilient Accelerated and Risk-Aware Edge Federated Learning in Scarce Data Scenario," in *IEEE Wireless Communications Letters*. (submitted)

Chapter 1

Introduction

1.1 Federated Learning as a Serice

The arrival of 6G networks signifies a transformative shift towards interconnected intelligent ecosystems, revolutionizing wireless communication. Within this landscape, Federated Learning as a Service (FLaaS) [1, 2] emerges as a critical technology at the network edge. FLaaS prioritizes data privacy and minimizes communication overhead by departing from traditional centralized machine learning approaches. Instead of pooling raw data, FLaaS leverages a central aggregator, typically an edge server , to combine locally trained models from edge devices. This iterative process of global model aggregation and local model updates continues until convergence, ensuring efficient utilization of edge device capabilities while maintaining optimal FLaaS performance.

In the age of information (AoI) [3], characterized by explosive data generation from Internet of Things IoT devices [4, 5], edge computing facilitates efficient data processing and decision-making [6]. By bringing computing resources closer to data sources, edge computing reduces latency and alleviates the burden on central servers,

enabling timely data analysis. FLaaS capitalizes on edge computing infrastructure by empowering edge devices to perform local model training and aggregation, minimizing communication latency and upholding data privacy. This decentralized approach enhances the scalability and efficiency of FLaaS systems while empowering edge devices to contribute actively to the collaborative learning process. The benefits of FLaaS extend beyond data privacy and communication efficiency. FLaaS reduces the computational load on resource-constrained edge devices by sharing model updates instead of raw data.. The diverse data from various devices contributes to more robust and generalizable models than to centralized training. Moreover, FLaaS offers scalability and flexibility, adapting to a growing number of devices and heterogeneous data distributions. Despite its advantages, FLaaS faces spesific challenges. Frequent model updates, while reducing overall communication, can still strain network resources. Varying data distributions across devices can lead to convergence issues and impact model performance. Additionally, potential vulnerabilities and attacks targeting FLaaS privacy mechanisms require careful consideration and mitigation.

1.2 Challenges

FLaaS offers a promising approach to collaborative machine learning while preserving data privacy. However, several challenges must be addressed to ensure its effectiveness and reliability [7]. One key challenge is achieving efficient convergence [8] during the learning process. The decentralized nature of FLaaS can lead to slower convergence compared to centralized methods, as model training is distributed across devices with varying computational resources and participation levels. Additionally, data scarcity and non-IID data distributions [9, 10] pose significant hurdles. Traditional

FLaaS algorithms often assume data homogeneity, which can lead to suboptimal performance when faced with diverse user datasets. Researchers are actively exploring solutions such as FedAvg with client-side momentum, asynchronous update rules, adaptive optimization algorithms, federated data augmentation, transfer learning, and personalized FLaaS to address these convergence and data challenges.

Another critical aspect is establishing trust and security within FLaaS systems [11–13]. Protecting against malicious users and ensuring the integrity of the learning process is crucial. Robust user authentication protocols, secure communication channels, and anomaly detection mechanisms are essential to prevent unauthorized access, data breaches, and manipulation of model updates. Techniques like homomorphic encryption [14] and differential privacy [15] can enable secure model aggregation while preserving data confidentiality. Additionally, ongoing research focuses on developing robust methods for detecting and mitigating stealthy attacks, where malicious users attempt to subtly influence the model without detection. FLaaS can reach its full potential as a secure and efficient collaborative learning paradigm by addressing these convergence, data, and trust challenges.

1.3 Contributions

This thesis delves into the challenges FLaaS presents in wireless communication, encompassing terrestrial and non-terrestrial networks. The primary focus is addressing the complexities of non-trusted environments characterized by data scarcity and potentially malicious users.

Firstly, the research investigates the efficacy of accelerated FLaaS techniques in enhancing model convergence and communication efficiency. By exploring methods

such as prioritizing high-data users and implementing efficient model aggregation, the study aims to demonstrate the potential for accelerated FLaaS to expedite the learning process and reduce communication overhead in diverse network settings.

Secondly, the thesis tackles the challenge of non-trusted users in FLaaS scenarios. Novel approaches for ensuring model integrity and privacy preservation are investigated, including the development and evaluation of trustworthiness metrics to mitigate the impact of malicious and risky users on FLaaS performance.

Furthermore, the research explores the role of validation datasets in addressing data scarcity and ensuring model quality across terrestrial and non-terrestrial networks. By analyzing the effectiveness of validation datasets in validating user contributions and detecting malicious behavior, the study provides valuable insights into their importance for enhancing FLaaS performance in real-world wireless communication applications.

Finally, this thesis proposes an alternative approach to address the challenge of non-trusted users in scenarios where validation datasets are absent. This approach aims to enhance the robustness and resilience of FLaaS systems in non-trusted environments, paving the way for more secure and reliable FLaaS implementations in wireless communication.

1.4 Organization of Thesis

This thesis is structured into three main chapters, each building upon the previous to provide a comprehensive exploration of FLaaS and its application in wireless communication, particularly within non-trusted environments. Chapter 2 establishes a

foundational understanding of FLaaS principles, covering essential concepts in machine learning, the implementation of FLaaS over wireless networks, and key security considerations for FLaaS systems. Building upon this foundation, Chapter 3 delves into utilizing validation datasets to address challenges posed by non-trusted users in FLaaS environments. This chapter explores the effectiveness of validation datasets in ensuring model quality, validating user contributions, and detecting malicious behavior. Chapter 4 extends the investigation to scenarios where validation datasets may be unavailable, proposing and evaluating alternative approaches for handling non-trusted users and enhancing the robustness of FLaaS systems in such challenging environments. Each chapter addresses specific research questions related to the challenges of FLaaS in wireless communication, ultimately contributing to developing more secure and reliable FLaaS implementations for diverse network settings. Finally, Chapter 5 concludes and summarizes the thesis, giving insights and future work recommendations.

Chapter 2

Background

2.1 Federated Learning Preliminaries

2.1.1 Machine Learning (ML)

Before delving into Federated Learning (FL), it is crucial to establish a solid understanding of the fundamental principles and techniques in ML [16] and deep learning (DL) [17]. Machine learning forms the basis of FL systems, providing the framework for training models on distributed data sources. In ML, supervised learning [18] involves learning from labeled data, where algorithms are trained to map input features to corresponding output labels. On the other hand, unsupervised learning [19] deals with uncovering hidden patterns and structures in unlabeled data. Reinforcement learning [20] focuses on learning optimal decision-making strategies through interaction with an environment. Deep learning techniques, particularly neural networks, are integral to FL because they can learn complex patterns and representations from data. Convolutional neural networks (CNNs) [21] are specialized neural networks for processing grid-like data, such as images, leveraging convolutional layers to extract spatial hierarchies of features. Recurrent neural networks (RNNs) [22] are designed

to handle sequential data by incorporating feedback loops, allowing them to capture temporal dependencies in sequences. Additionally, understanding optimization algorithms like stochastic gradient descent (SGD) [23] and its variants, regularization techniques such as L1 and L2 regularization, and model evaluation metrics like accuracy, precision, recall, and F1 score is essential for effectively training, optimizing, and evaluating FL models. By mastering these foundational concepts in ML and DL, researchers and practitioners can navigate the landscape of FL with confidence, harnessing its potential for distributed and collaborative model training.

2.1.2 Federated Learning

FL is a decentralized machine learning approach that trains a global model using datasets distributed across various geographical locations. Instead of centralizing data, FL harnesses the computational power and data availability of numerous devices referred to as clients. Each client trains a local model on its own dataset and transmits the model parameters to a Base Station (BS) or Aggregator Server (AS). The AS plays a crucial role in this framework by aggregating these local models using a specific aggregation algorithm to form an updated global model. This process is iterative, aiming to optimize the following objectives:

$$\min_{w} F(\boldsymbol{w}) = \sum_{n=1}^{U} p_n F_n(\boldsymbol{w}), \tag{2.1}$$

Here, $p_n = \frac{D_n}{\sum_{n=1}^{U} D_n}$ represents the proportion of the total dataset contributed by user n, where D_n is the size of the dataset held by user n. The term $F_n(\boldsymbol{w})$ denotes the average loss function for user n, reflecting the performance of the model $\boldsymbol{w}$ on user n's data. This loss function aggregates the errors made across all samples in user n's

dataset, guiding the optimization to improve the model's accuracy and utility across diverse data distributions and is represented by:

$$F_n(\boldsymbol{w}) = \frac{1}{D_n} \sum_{s \in D_n} f_n(\boldsymbol{w}, s), \tag{2.2}$$

where $f_n(\boldsymbol{w}, s)$ is the loss function at weight $\boldsymbol{w}$ on sample s. The iterative process of updating and aggregating local models continues until the global model achieves satisfactory accuracy or meets other predefined criteria. This method leverages the unique strengths of distributed data while adhering to privacy constraints, as raw data never leaves its original location.

The distribution of datasets across clients can significantly influence the performance and complexity of the learning process. There are two primary types of dataset distributions:

- **Independent and Identically Distributed (IID)**: In this type of distribution, each client's local dataset is assumed to represent the overall dataset fairly. This means that the statistical properties of the data, as well as the proportion of each class, are consistent across all clients. Each dataset is like a random sample from the same population, ensuring no single dataset is biased toward specific characteristics.

- **Non-Independent and Identically Distributed (Non-IID)** [24] [25]: This distribution is more complex and realistic, where the data across different clients vary significantly. The variations can include differences in the class distribution (some classes might be overrepresented or underrepresented in certain datasets),

and some classes might even be exclusive to specific datasets. This heterogeneity can introduce challenges in learning a global model that performs well across all clients because the model must generalize well across diverse, unevenly distributed data sources.

Understanding these distinctions is crucial for designing effective FL algorithms, as the data distribution type can impact both the convergence behavior of the learning algorithm and the final model's performance.

One of the foundational aggregation algorithms in FL is Federated Averaging (FedAvg) [26]. This algorithm has garnered significant attention due to its simplicity, effectiveness, and ease of implementation. As shown in Fig. 2.1, FedAvg aims to aggregate the received model weights from participating clients using a weighted average, where each client's contribution is proportional to its dataset size. Mathematically, as expressed in the following equation:

$$\boldsymbol{g}_t = \sum_{n=1}^{U} p_n \boldsymbol{w}_{n,t}, \tag{2.3}$$

at each iteration, the aggregated global model $\boldsymbol{g_t}$ is computed as the sum of weighted model updates from individual clients, represented as $\sum_{n=1}^{U} p_n \boldsymbol{w}_{n,t}$.

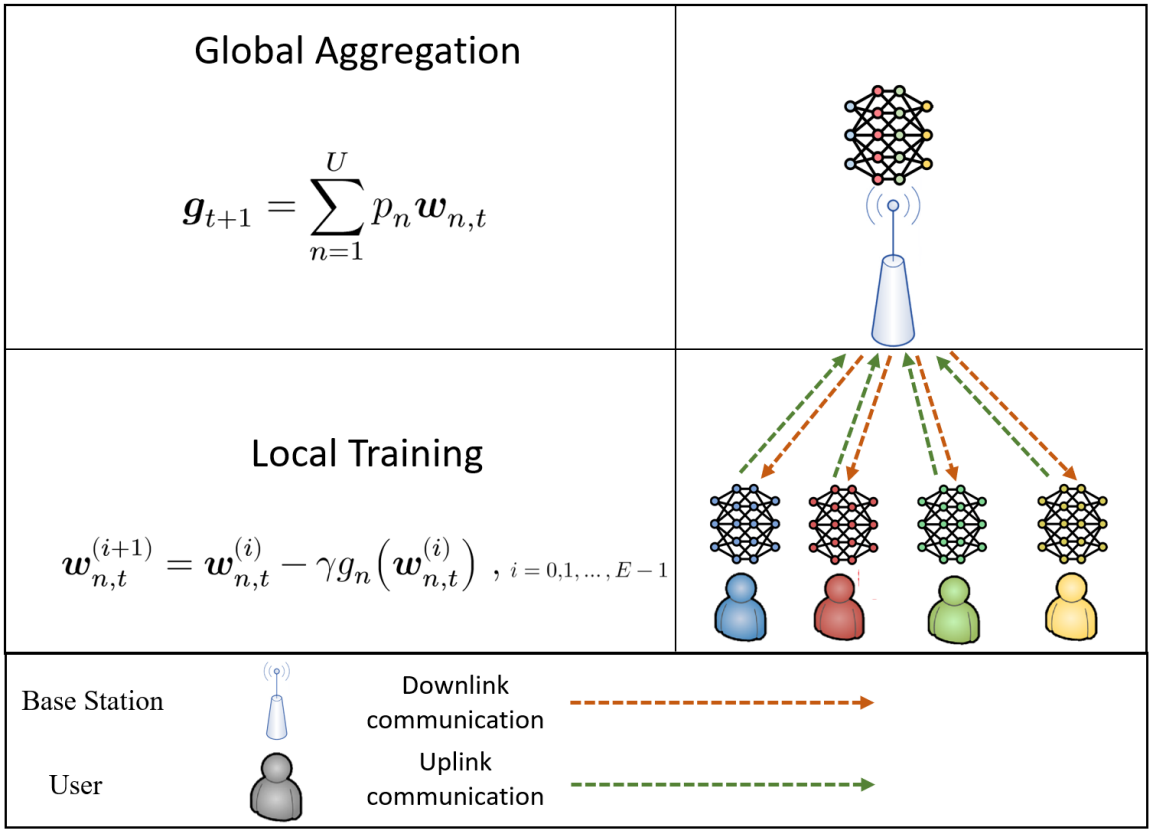

Figure 2.1: FedAvg Architecture

2.1.3 FedAvg

The process of FL unfolds through a series of key steps:

1. **Global Weight Initialization**: This crucial initial phase sets the foundation for FL. The system broadcasts the initial global model g_0 to all participating devices. This global model is tailored to the specific task and is designed considering factors such as the dataset and learning objectives.

2. **Local Edge Training**: Following global weight initialization, each device embarks on its training using the received global model g_0. Through local edge

training, devices refine their model weights $\boldsymbol{w}_{n,t}$ using iterative optimization techniques such as Stochastic Gradient Descent (SGD) as in [27]. The update equation,

$$\boldsymbol{w}_{n,t}^{(i+1)} \leftarrow \boldsymbol{w}_{n,t}^{(i)} - \gamma g_n(\boldsymbol{w}_{n,t}^{(i)}), \tag{2.4}$$

guides this process over a predetermined number of epochs i ranging from 0 to $E-1$. Here, γ denotes the learning rate essential for fine-tuning the model parameters.

3. **Local Model Uploading**: With local training complete, all devices transmit their refined local weights back to the AS. This step prepares the gathered model updates for the subsequent aggregation process.

4. **Aggregation and rebroadcasting the weights**: The pivotal role of the AS comes to the forefront in this phase. Leveraging the received model updates from participating devices, the AS aggregates them to generate the new global model g_{t+1}. Once aggregated, the AS redistributes the updated global model to all users, ensuring that the latest models are disseminated across the federated network.

FedAvg's key advantages are its simplicity, efficiency, and scalability, making it a popular choice for FL tasks. FedAvg operates under the assumption of IID data, where each client's local dataset mirrors the overall dataset's statistical distribution. In such scenarios, FedAvg excels, leveraging the homogeneous data to efficiently aggregate model updates and achieve impressive performance across various domains, including image classification, natural language processing, and healthcare. However, in non-IID settings where data distributions vary significantly among clients,

FedAvg's performance may degrade due to mismatched data characteristics. Despite this limitation, FedAvg continues to be a popular choice for FL tasks, driving ongoing research efforts to develop more sophisticated aggregation algorithms that can adapt to diverse data distributions and enhance performance in real-world scenarios.

2.2 Federated Learning Over over Wireless Communication

2.2.1 Wireless Impairments and Stochastic Geometry

Wireless communication revolutionizes connectivity by offering enhanced mobility, broader coverage, and higher data rates than wired alternatives. This technology enables seamless communication across diverse environments, empowering users with unprecedented flexibility and convenience. However, the advantages of wireless communication are accompanied by inherent challenges. These challenges stem from the dynamic and unpredictable nature of the wireless medium, which exposes communication links to various impairments. Three primary factors contributing to these impairments are fading, interference, and path loss.

Fading

Fading refers to the fluctuations in signal strength experienced by wireless signals as they propagate through the environment. These fluctuations can be attributed to various factors, including multipath propagation and obstacles encountered along the signal path. Multipath propagation occurs when wireless signals travel multiple paths from the transmitter to the receiver, resulting in signal reflections, diffraction, and scattering. As a result, different copies of the transmitted signal arrive at the receiver with different delays and phases, leading to constructive or destructive interference.

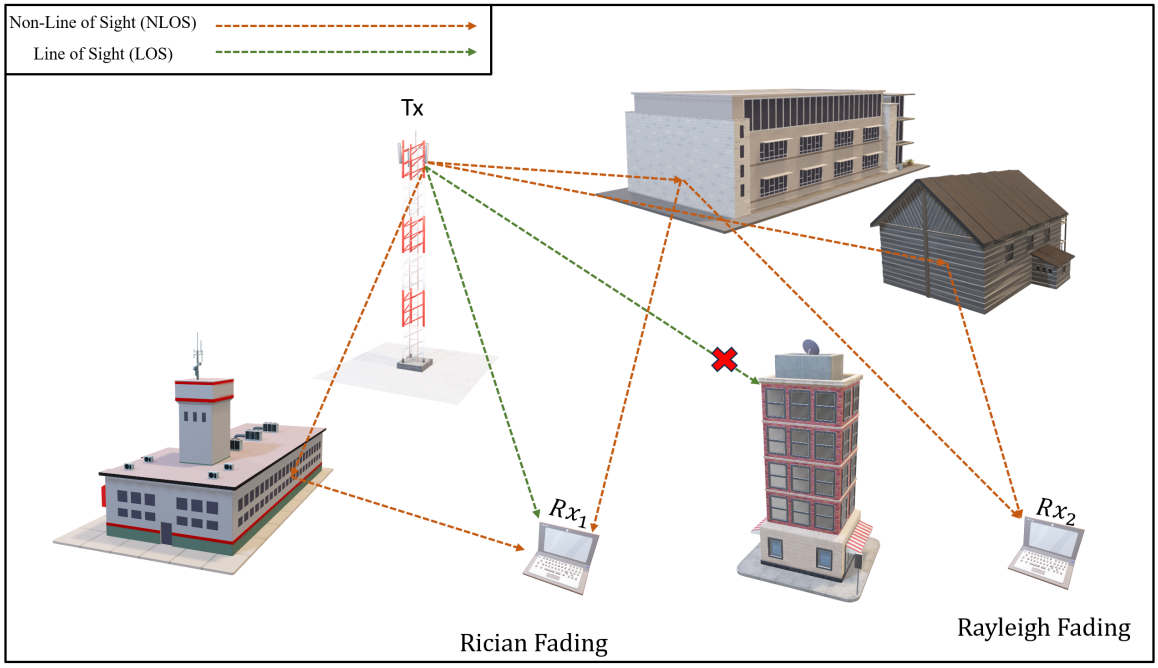

Figure 2.2: Fadings Type

Additionally, obstacles in the signal path, such as buildings, trees, and terrain features, can attenuate or block wireless signals. Two primary scenarios shown in Fig. 2.2 describe how fading occurs in wireless networks:

1. **Line of Sight (LOS) Network**

   - In a LOS network, communication paths are direct and unobstructed between transmitters and receivers. This direct link ensures minimal signal attenuation and interference, providing clear and reliable communication.
   - LOS scenarios are particularly advantageous for wireless systems operating in open spaces or with elevated antennas, where obstacles like buildings or terrain are less likely to obstruct the signal path.
   - The fading model commonly used in LOS scenarios is Rician Fading, which

combines a strong line-of-sight signal component with scattered signals. This model suits environments with clear line-of-sight paths and significant reflected signals, such as urban areas with buildings causing multipath propagation.

2. **Non-Line of Sight (NLOS) Network**

   - In an NLOS network, communication paths involve reflections from surrounding objects or surfaces, leading to signal scattering and multipath propagation. These reflected links can result in signal attenuation, delay spread, and increased interference.
   - NLOS scenarios are common in environments with obstructed line-of-sight paths, such as dense urban areas, indoor spaces, or regions with foliage and terrain variations.
   - The fading model typically used to model NLOS scenarios is Rayleigh Fading, which represents the random amplitude fluctuations of signals due to multipath propagation. A lack of dominant line-of-sight signals characterizes rayleigh fading and is suitable for modeling complex fading patterns in NLOS environments.

3. **Nakagami-m Fading**

   - Nakagami-m fading is a mathematical model proposed in [28] to capture both the effects of LOS and NLOS environments. It incorporates characteristics from both Rician and Rayleigh fading models, making it versatile for scenarios where there are varying degrees of direct line-of-sight and multipath interference.

- This model is particularly useful in wireless communication systems where the signal experiences a combination of clear line-of-sight paths and scattered reflections, such as in urban environments with buildings and foliage.
- Nakagami fading parameters can be adjusted to represent different fading conditions, providing flexibility in simulating realistic wireless channel behaviors.

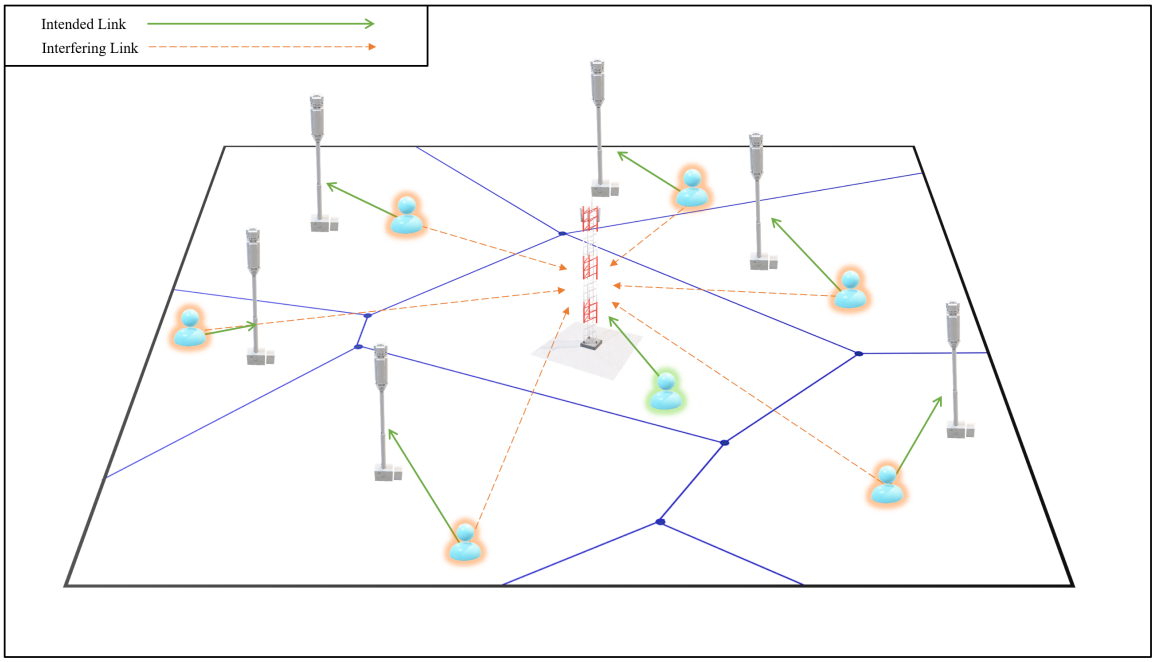

Figure 2.3: Interference Links Resulting from Surrounding Users

Interference

Interference arises from the simultaneous transmission of signals by multiple users or devices over the same resource (time, frequency, resource block, phase), as shown in Fig.2.3, leading to signal degradation and reduced reliability. In wireless communication systems, managing interference is critical for maintaining desired quality of

service and system performance.

Stochastic Geometry in Interference Analysis Stochastic geometry is fundamental in analyzing and mitigating interference in wireless networks. Here is how stochastic geometry contributes to interference analysis:

- **Modeling Spatial Distributions**: Stochastic geometry models the spatial distribution of wireless nodes, such as BSs, users, and interference sources, in a probabilistic manner. This modeling approach accounts for the randomness in node placements, which is crucial for realistic interference modeling [29].

- **Analyzing Interference Patterns**: Using stochastic geometric models, researchers can analyze interference patterns and quantify interference levels across different spatial regions within the network. This analysis includes node densities, transmit power levels, antenna patterns, and propagation environments.

- **Coverage and Capacity Assessment**: Stochastic geometry enables the assessment of coverage probabilities and capacity limits considering interference effects. It helps determin the coverage probability for a given Signal-to-Interference-Plus-Noise-Ratio (SINR) threshold and evaluate achievable data rates under interference-limited conditions.

- **Optimization and Design**: Utilizing stochastic geometry-based interference models, network planners can optimize resource allocation, design efficient interference mitigation techniques, and enhance overall network performance. These models assist in making informed decisions regarding deployment strategies, power control schemes, and interference management policies.

Stochastic geometry provides a rigorous mathematical framework for understanding, analyzing, and managing interference in wireless communication systems. Its application in interference analysis is instrumental in optimizing network design, improving spectral efficiency, and ensuring reliable communication in complex wireless environments.

Path Loss

Path loss refers to the attenuation or weakening of a wireless signal as it propagates through the environment from the transmitter to the receiver. As the signal travels over a distance, it gradually loses power due to free-space spreading, absorption, and scattering. Additionally, penetrating obstacles like buildings, trees, and terrain further attenuate the signal, resulting in reduced signal strength at the receiver compared to the transmitted power. Understanding path loss is crucial in wireless communication systems as it directly impacts signal coverage, range, and quality, influencing the design and deployment of wireless networks. Combining these factors, the SINR equation becomes crucial for assessing wireless communication quality. The SINR denotes the proportion of the signal power of interest relative to the added power of noise and interference at the receiver.

$$\text{SINR} = \frac{P h_0^2 r^{-\eta}}{N_0 + I} \tag{2.5}$$

In the equation (2.5), P denotes the original transmitted power, h_0^2 represents the fading power, $r^{-\eta}$ corresponds to the path exponent capturing the signal attenuation over distance r with η as the path loss exponent, N_0 denotes the inherent noise power in the receiver, and I denotes the interference power from other sources.

A higher SINR indicates better signal quality and improved communication performance, highlighting the importance of optimizing transmit power, antenna design, and interference management strategies to maximize SINR and enhance overall system reliability and efficiency.

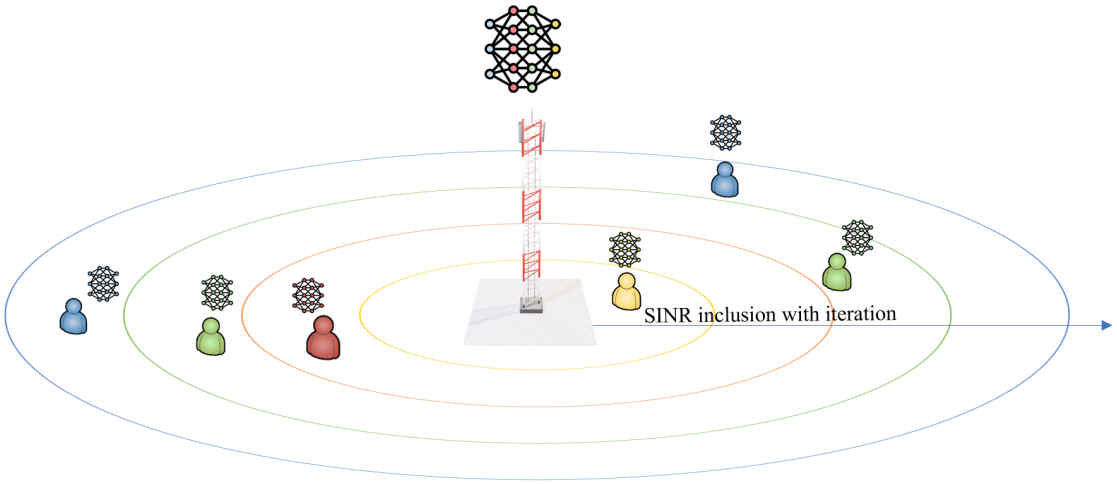

Figure 2.4: Interference Links resulted from Surrounding users

2.2.2 Accelerated FL

The concept of accelerated FL was initially introduced in the literature [8]. As shown in Fig. 2.4, the approach aims to expedite the initial training rounds by leveraging a higher SINR, which translates to higher data rates. The strategy involves starting with a higher SINR threshold in the early training rounds to accelerate the convergence process. As the training progresses, the system gradually relaxes the SINR threshold, allowing for the inclusion of more users in the FL process. This adaptive approach optimizes the trade-off between data rate and system capacity, enabling

faster convergence without a noticeable degradation in performance. The effectiveness of this approach was demonstrated in the referenced paper, showcasing significant enhancements in convergence speed while maintaining satisfactory performance levels.

2.2.3 Unbiased FL

The presence of wireless impairments, as discussed in Section 2.2.1, introduces variability in the signals received by the AS from different users. This variability can lead to situations where the AS may not reliably decode signals from users with a lower Signal-to-Interference-and-Noise Ratio (SINR). Consequently, users with lower SINR may fail to participate in certain FL training rounds. The repeated exclusion of low SINR users from multiple training rounds can introduce bias into the FL model, favoring users with higher SINR who consistently participate. To address this biasing effect, techniques for unbiased FL are being explored in [8,30,31]. The authors achieve this by weighting the contribution of user x by the reciprocal of the probability of successful participation of that user as follows:

$$\boldsymbol{g_t} \leftarrow \boldsymbol{g_{t-1}} + \sum_{n=1}^{U} \frac{p_n}{S_{n,t}} (\boldsymbol{w}_{n,t} - \boldsymbol{g_{t-1}}), \tag{2.6}$$

where the term $(\boldsymbol{w}_{n,t} - \boldsymbol{g_{t-1}})$ represents the actual contribution from the devices. Thus this contribution is weighted by the factor $S_{n,t}$ is given by

$$\begin{aligned} S_{n,t}(\zeta_t, r_n) &= \mathbb{P}(\mathrm{SINR}_n^{(t)} > \zeta_t) \\ &= \mathbb{P}\left(\frac{P h_0^2 r^{-\eta}}{N_0 + I_{agg}} > \zeta_t\right) \end{aligned} \tag{2.7}$$

These techniques aim to ensure fair and equitable participation of all users in the FL process regardless of their SINR levels. They are crucial in maintaining model fairness and representativeness in heterogeneous wireless environments.

2.3 Security in FL

2.3.1 Byzantine Attacks

FL faces various security challenges, including attacks that aim to compromise the learning process's integrity, confidentiality, or availability. These attacks can disrupt the collaborative nature of FL, leading to inaccurate models, data leakage, or system downtime. Examples of attacks in FL include the following:

- **Model Poison Attack**: In a model poison attack, adversaries aim to compromise the integrity of the global model by injecting malicious modifications during the training process. This can involve substituting portions of the original model with a compromised version, introducing backdoors, or alter the model's parameters to induce specific misclassifications. Adversaries may create fake users [32] or exploit existing genuine users [33] to distribute the poisoned models across the FL network. The consequences of a successful model poison attack can range from subtle biases in model predictions to outright malfunctioning, posing significant risks to the integrity and reliability of FL systems.

- **Data Poison Attack**: Data poison attacks involve manipulating training data to undermine the learning process and compromise the integrity of the resulting model. Attackers may modify the dataset by adding, deleting, or altering data samples, injecting adversarial examples [34], or introducing biases to skew the

model's learning dynamics. These manipulations can lead to the propagation of erroneous patterns and biases throughout the FL network, ultimately resulting in compromised model performance and reliability.

2.3.2 Attacks Mitigation

Mitigating attacks in FL is crucial to safeguard the integrity and reliability of the entire learning process. While the FedAvg algorithm, which employs a simple mean aggregation rule to combine local model updates, is effective under non-adversarial settings, it can be vulnerable to manipulation in adversarial environments. Adversaries can exploit compromised worker devices to manipulate the mean of the local models arbitrarily, leading to potentially compromised global model parameters [1, 35, 36].

To develop a resilient approach for dealing with attacks in FL, two main approaches are commonly followed:

1. **Validation Dataset-Driven**

   In validation dataset-driven approaches, researchers focus on designing aggregation algorithms that mitigate the influence of malicious users while making use of an assumption about the existence of a small validation dataset. This dataset should ideally cover the statistical distribution of local datasets across all users, enabling the identification and exclusion of malicious updates [37,38].

2. **Local Statistical Model-Driven**: Conversely, local statistical model-driven approaches do not rely on assumptions about the availability of a small validation dataset. Instead, they develop aggregation functions that operate solely based on the statistical properties of the local models received from the clients.

Several advanced aggregation algorithms have been proposed to address this challenge, including:

(a) **Krum**: Krum (Krumming) is an aggregation algorithm designed to mitigate the impact of Byzantine attacks in FL. It selects the local updates closest to the median of all updates regarding euclidean distance, effectively filtering out potential outliers introduced by malicious participants [35].

(b) **Bulyan**: Bulyan is another aggregation algorithm developed to enhance robustness against Byzantine attacks. It works by selecting a subset of local updates that are the most similar to each other in terms of cosine similarity, effectively reducing the influence of malicious updates while maintaining computational efficiency [39].

(c) **Trimmed-Mean**: The trimmed-mean aggregation algorithm involves discarding a certain percentage of the highest and lowest local updates before computing the mean of the remaining updates. This approach helps mitigate the impact of outliers introduced by malicious participants while preserving the contributions of genuine updates [40].

2.3.3 Trustworthy Environment

In the context of FL, establishing a trustworthy environment is paramount to ensure the integrity and effectiveness of collaborative model training. Researchers have dedicated efforts to develop mechanisms for evaluating the trustworthiness of participating clients, providing valuable insights and recommendations regarding user behavior within FL systems.

To initiate this process, authors in [41, 42] have proposed a sophisticated trust

mechanism tailored specifically to address the challenges posed by "cold start" users—those lacking historical interaction or contribution within the FL ecosystem. This trust mechanism aims to derive TM scores, often called reputations or trustworthiness, for such users, enabling the system to make informed decisions about their reliability and contribution potential. By effectively addressing the unique characteristics and uncertainties associated with cold start users, this approach enhances the overall robustness and fairness of the FL environment.

Building upon these foundations, researchers in [43] have extended the concept of trustworthiness assessment to serve as a pivotal tool for detecting and mitigating malicious behavior within FL systems. Leveraging the TM scores derived from the mechanism mentioned above, they propose a comprehensive approach for identifying potentially malicious users. By establishing a predefined threshold for trustworthiness metric, users falling below this threshold are deemed suspicious or malicious entities and subsequently excluded from collaborative model training.

Chapter 3

Risk-Aware Accelerated Wireless Federated Learning with Heterogeneous Users

3.1 Contribution

This chapter presents a risk-aware FL aggregation model that accounts for wireless communication factors (e.g., interference and fading) and the trustworthiness of each device. The main contributions of this chapter are as follows:

- Development of a Comprehensive Risk-Aware Framework: This framework integrates considerations of wireless channel conditions and device security profiles into the Federated Learning as a Service (FLaaS) paradigm. This holistic approach ensures the model training is robust against communication errors and malicious contributions from untrustworthy users.

- Investigation of Non-Trustworthy User Impact: The chapter delves into the detrimental effects of non-trustworthy users on accelerated wireless FL, particularly in scenarios with scarce data sources. This analysis highlights the potential risks associated with incorporating data from unreliable users and

motivates the need for effective mitigation strategies.

- Utilization of Validation Datasets for Mitigation: This work proposes the utilization of validation datasets as a mechanism to mitigate the contributions from non-trustworthy users during the model fine-tuning stage.

3.2 System Model

Our model considers a single-tier cellular network where the Aggregator Server (AS) can be either traditional terrestrial Base Station (BS) or Unmanned Aerial Vehicle (UAV) acting as aerial ASs. The ASs locations are modeled according to a Poisson Point Process (PPP) to capture the randomness of their deployment. Users are assumed to be uniformly distributed within the network area. Each user associates with the AS that offers the strongest received signal typically the geographically closest AS. This association is illustrated in Fig.3.1 for a terrestrial BS scenario and Fig.3.2 for a UAV-based scenario. Without loss of generality, we consider a typical AS (either terrestrial BS or UAV) located at an arbitrary origin.

3.2.1 FLaaS Model

The FLaaS is provided by terrestrial BSs or UAVs acting as AS for proximate edge devices. Let U be the number of devices that possess related data for the FLaaS transaction per AS, then each AS is tasked with the execution of a distributed optimization problem, formally expressed as:

$$\min_{\boldsymbol{g}_t} f(\boldsymbol{g}_t) = \sum_{n=1}^{U} a_n F_n(\boldsymbol{g}_t), \tag{3.1}$$

where a_n is a weighting factor for the local model aggregation that accounts for the per-device parameters such as the amount of possessed data and the Trustworthiness Metric (TM) score. An FLaaS transaction involves several global iterations $t \in \{0, T\}$, where the loss function $f_n(\cdot)$ is assessed at the global model parameter $\boldsymbol{g}_t$ for each device n. Hence, the optimization in (3.1) seeks the most suitable vector $\boldsymbol{g}_t$ that minimizes the average loss across all devices, thereby facilitating a comprehensive and effective learning model that is reflective of the aggregate insights derived from the distributed network. The selection of T can be determined either by a predefined value or dynamically. In the dynamic approach, the learning process continues until the difference between consecutive global model weights magnitude, $|\boldsymbol{g}_t|$ and $|\boldsymbol{g}_{t+1}|$, falls below a specified threshold δ. This criterion ensures that the training stops when the model's improvement becomes sufficiently small, indicating a successful convergence.

For each global iteration of the global model in (3.1), every device updates its local model parameters $e \in \{0, E-1\}$ times through the execution of the following update rule:

$$\boldsymbol{w}_{n,t}^{(e+1)} = \boldsymbol{w}_{n,t}^{(e)} - \gamma_n \nabla F_x(\boldsymbol{w}_{n,t}^{(e)}), \tag{3.2}$$

where $\boldsymbol{w}_{n,t}^{(e)}$ represents the local model of device n after e epochs during the t^{th} global round, γ represents the learning rate, and $\nabla F_n(\cdot)$ is the gradient of the objective function to the local model parameters. Each device transmits its final parameter vector $\boldsymbol{w}_{n,t}$ to the AS.It is worth noting that in the context of Non-IID datasets, increasing the value of E (the number of local epochs) can negatively affect the learning process. When E is large, users may overfit their local models to their specific datasets. This overfitting leads to a divergence from the global model, as each user's

model becomes more specialized to their local data and less representative of the overall data distribution. Consequently, the aggregation of these highly specialized local models can degrade the performance and generalization ability of the global model.

Therefore, careful selection of E is crucial in federated learning with Non-IID data to balance the trade-off between local training efficiency and global model generalization.

3.2.2 Network Model

The considered networks adopt a universal frequency reuse scheme for a set of orthogonal channels such that each user n occupies one Resource Block (RB). Following this approach and having the number of RB greater than or equal to the number of devices in each cell would allow the users to send their data simultaneously without interfering with each other. However, interference is generated from devices occupying the same RB at different ASs. To simplify the analysis, we assume perfect downlink communication. In addition, we adopted constant power transmission P for all users in U. We assume that the noise inherent in the AS, denoted by N_0, follows a Gaussian distribution with variance σ^2. The system model is adaptable to both terrestrial and non-terrestrial network environments, with specific characteristics and considerations for each scenario detailed in the subsequent sections.

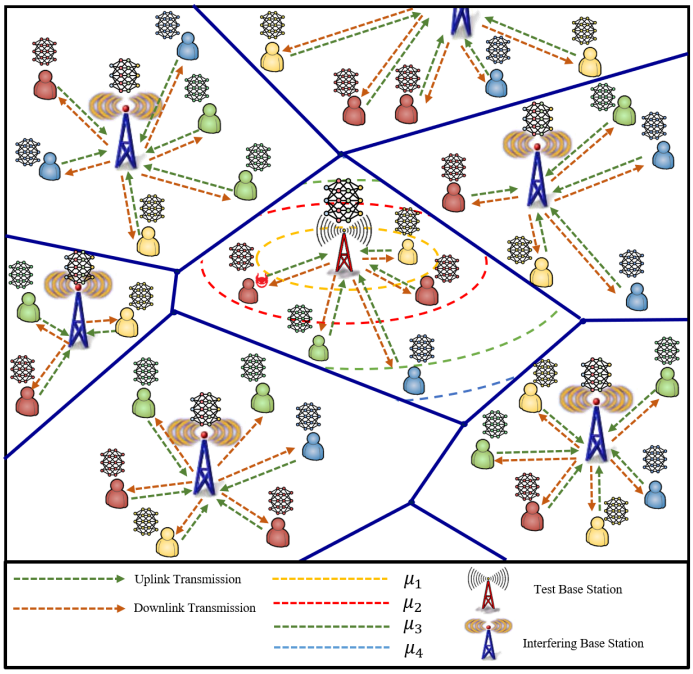

Figure 3.1: System model showing a terrestrial BS implementing dynamic SINR-based aggregations with four levels.

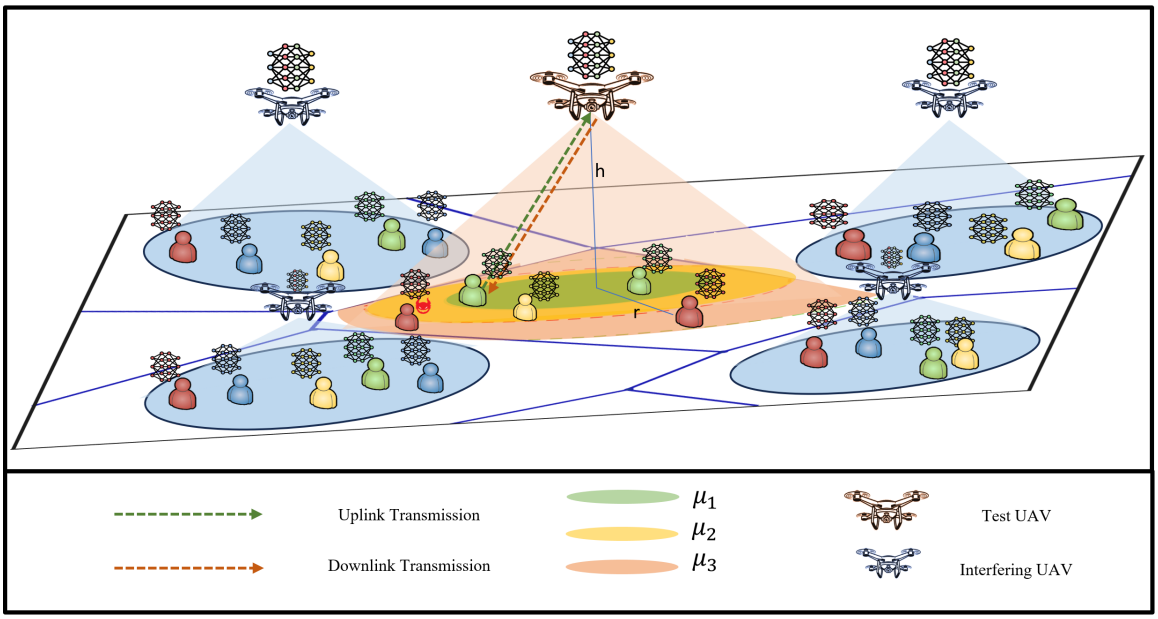

Figure 3.2: A dynamic SINR-based aggregation implemented on a UAV with three levels

Terrestrial Network Model

In the terrestrial network model, small-scale fading effects are represented using a Rayleigh distribution for the channel coefficient, denoted by h_0. The large-scale path loss is modeled using a power-law relationship where the signal strength diminishes proportionally to $r^{-\eta}$ with increasing propagation distance r. The path-loss exponent, η, characterizes the signal attenuation rate in the terrestrial environment.

Aerial Network Model

We employ the estimated probability of Line of Sight (LOS) as detailed in [44] to model the impact of obstruction in the surroundings.. This probability characterizes the likelihood of establishing a direct line of sight between user n and its associated UAV and is expressed as follows:

$$P_{\mathrm{NLOS}}(r_n) = \frac{1}{1 + a \exp\left(-b\left[\frac{180}{\pi}\arctan\left(\frac{h_U}{r_n}\right) - a\right]\right)}, \tag{3.3}$$

in this context, the environmental parameters a and b are constants, and their values are specified in [44]. The variable h_U denotes the height of the UAV and r_n represents the horizontal distance between user n and the center ground of the UAV. The likelihood of a Non-Line of Sight (NLOS) scenario is expressed as $P_{\mathrm{N}}(r_n) = 1 - P_{\mathrm{L}}(r_n)$.

Additionally, for NLOS and NLOS conditions, the fading characteristics are modeled using the Nakagami-m distribution, where the power gains are represented as independent and identically distributed (IID) gamma variables with shape parameters m_{L} and m_{N}, respectively. Path loss in the aerial environment is also modeled using a distance-dependent power-law relationship with distinct path loss exponents

χ_L and χ_N for NLOS and NLOS conditions, respectively.

We assume that both UAVs and devices utilize directional antennas to enhance the desired signal and reduce interference. The antenna patterns are simplified using a discretized sectored gain model [45]. We denote the main lobe and side lobe by G_{uM} and G_{um}, respectively, for the UAV. Following the same notation, the main and side lobes for the user's devices are G_{nM} and G_{nm} respectively. Additionally, we assume that the link between the UAV and its associated user's device is perfectly aligned with $G_0 = G_{uM}G_{nM}$. However, the interfering links have uniformly distributed beam directions with gains G and probability $\mathbb{P}(G)$ as shown below:

Table 3.1: Beam Gain with Gain Probabilities

	0	1	2	3
G_i	$G_{uM}G_{nM}$	$G_{uM}G_{nm}$	$G_{um}G_{nM}$	$G_{um}G_{nm}$
$\mathbb{P}$	$(\frac{\theta_u}{2\pi})(\frac{\theta_d}{2\pi})$	$(\frac{\theta_u}{2\pi})(1-\frac{\theta_d}{2\pi})$	$(1-\frac{\theta_u}{2\pi})(\frac{\theta_d}{2\pi})$	$(1-\frac{\theta_u}{2\pi})(1-\frac{\theta_d}{2\pi})$

3.2.3 Trustworthiness Model

The users are ranked with a TM score obtained from their profiles, type of devices, and participation history in prior FL processes. The TM score follows the dual-faceted approach that encompasses two key factors:

- **Quality:** This metric provides insights into the data quality as well as the computing capabilities and accuracy of the devices.
- **Security:** This metric provides insights into the user's authenticity and security profile (e.g., security software and updates).

The synthesis of the quality and security metrics leads to a unified measure referred to as the users' trustworthiness metric [46] [47]. This comprehensive metric provides an

assessment of users that is crucial for the FL model integrity. To account for random user profiles in large-scale networks, the TM score $\Omega_n \sim Beta(\alpha, \beta)$ is modeled using the beta distribution with parameters α and β. The choice of the beta distribution is driven by its adaptability and flexibility as a continuous distribution that varies from 0 to 1, allowing for nuanced representation and adaptation to the dynamic nature of the TM score.

Users that have $\Omega_n \geq \rho$ are categorized as fully trusted users $U_F \subset U$, where ρ is a high threshold that distinguishes reliable devices with up-to-date security patches. On the contrary, users with a TM score of $\Omega_n \leq \kappa$ are considered malicious users $U_M \subset U$, where κ is a low threshold that distinguishes faulty devices and adversary users. Users with a TM score of $\kappa < \Omega_n < \rho$ are risky users $U_R \subset U$ that are reporting the following manipulated weights:

$$\boldsymbol{w}'_{n,t} = \boldsymbol{w}_{n,t} * \left(1 + \frac{(1 - \Omega_n)}{10}\right), \tag{3.4}$$

which can be due to low-quality data, imprecise computation at the device, or a sort of covert (i.e., hard to detect) model poisoning attack [48]. These users hold valuable data from which the model would benefit if used properly. In all cases, we assume that the magnitude of the model deviation is inversely proportional to the TM score of the user Ω_n. This modeling approach enables us to incorporate the impact of users with varying TM scores, effectively capturing the impact of risky users. By employing this equation, we simulate the manipulative actions of risky users and account for model imperfections from users with moderate TM scores.

It is worth noting that each user in U falls in only of the trustworthiness categories such that $U_F \cup U_R \cup U_M = U$ and $U_F \cap U_R \cap U_M = \phi$. While U_F are fully trusted, they

do not possess sufficient data for FL training. Hence, we propose utilizing the set of users U_R in the initial training stages. Then, U_F will be restricted from fine-tuning the model during the later stages of the learning process.

3.3 Aggregation Algorithm

3.3.1 Success Uploading Model

To expedite the convergence speed of FL, the model selectively prioritizes high data rate users during the initial rounds while temporarily excluding others. As the training progresses, the system gradually incorporates an increasing number of users with lower data rates to ensure broader participation over time. Adhering to this approach, users with high data rates play a pivotal role in rapidly constructing a robust model, significantly reducing the overall training time. Therefore, the transmission rate in the t^{th} round is determined as $\log(1+\mu_t)$, where the transmission of the n^{th} user is deemed successful only if $\text{SINR}_c^{(t)} > \mu_t$. A visual representation of this dynamic SINR training process is depicted in Fig.3.1. To mitigate the bias resulting from the approach above, we need to amplify the users with low SINR (i.e., located far from the AS) by multiplying by a location-dependent weighted factor that is given by the following lemmas:

Lemma 1. *The location-dependent weighted factor $S_{n,t}(\mu_t, r_n)$ for a randomly selected user n in a terrestrial wireless environment is given by,*

$$S_{n,t}(\mu_t, r_n) = \mathbb{P}(\text{SINR}_n^{(t)} > \mu_t) = \mathcal{L}\left(\frac{\mu_t}{Pr_n^{-\eta}}\right) \exp\left(-\frac{\mu_t N_0}{Pr_n^{-\eta}}\right), \tag{3.5}$$

Proof. see Appendix A section .1. □

Lemma 2. *The location-dependent weighted factor $S_{n,t}(\mu_t, r_n)$ for a randomly selected user n in an aerial wireless environment is given by,*

$$S_{n,t}(\mu_t, r_n) = \mathbb{P}(\mathrm{SINR}_n^{(t)} > \mu_t) = \sum_{n=1}^{m_L} (-1)^{n+1} \binom{m_L}{n} \exp\left(-N_0 \zeta_L\right) \mathcal{L}_{\mathcal{L}}\left(\zeta_L\right) \mathcal{L}_{\mathcal{N}}\left(\zeta_L\right) P_L(r_n) + \sum_{n=1}^{m_N} (-1)^{n+1} \binom{m_L}{n} \exp\left(-N_0 \zeta_N\right) \mathcal{L}_{\mathcal{L}}\left(\zeta_N\right) \mathcal{L}_{\mathcal{N}}\left(\zeta_N\right) P_N(r_n). \tag{3.6}$$

where $\zeta_n = \frac{g_n (d_n^2 + h^2)^{\chi_n/2} \mu_t}{P G_0}$

Proof. see Appendix A section .2. □

As a result, as shown in Lemmas [1 and 2], for a randomly selected user, the model will be weighted by the factor $\frac{1}{S_{n,t}(\mu_t, r_n)}$. The weighting factor consists of the probability that the device's SINR exceeds the threshold μ_t. It operates as an amplification mechanism to the contribution of each user to the global model, such that devices with lower success probability have higher contributions. Here $\mathcal{L}(s)$ represents the Laplace transform (LT) of the interference in the uplink, which can be obtained via systematic stochastic geometry analysis as [29]:

For the terrestrial environment:

$$\mathcal{L}(s) = \exp\left\{-2\pi\lambda \int_0^{\infty} \left(\frac{\left(1 - \exp\left(-\pi\lambda r^2\right)\right)}{1 + \frac{r^{\eta}}{sP}}\right) r dr\right\}. \tag{3.7}$$

For the aerial environment:

$$\mathcal{L}_x(s) = \exp\left\{-2\pi\lambda \sum_{q=1}^{4} \mathbb{P}_q \int_0^{\infty} (1 - \exp\{-\pi\lambda z^2\}) \left(1 - \left(1 + \frac{sPG_q(z^2+h^2)^{-\chi_x/2}}{m_x}\right)^{-m_x}\right) z \quad dz\right\}. \tag{3.8}$$

3.3.2 Dynamic SINR Thresholds and Trustworthiness Integration: Algorithm Details

Algorithm 1 outlines the enhanced dynamic SINR approach with an incorporation of trustworthiness considerations. It is often assumed that a limited dataset is available, typically derived from publicly accessible datasets or provided by the service provider. This practice aims to ensure similarity in data distribution to the private data held by users, as demonstrated in prior works such as [1, 37, 38]. In the initial rounds, the system accommodates the set of users U' that contains fully trusted and risky users $U_F \cup U_R$. This inclusion of users is particularly effective in the context of non-IID datasets, where diverse devices with distinct datasets contribute positively to system enhancement. Upon receiving the local weights by the AS, the algorithm proceeds by categorizing these users into a newly subset denoted as U". This subset only contains users who have met the SINR requirements.

Algorithm 1: Risk-Aware Accelerated Wireless Federated Learning

Data: $\mu_t, \boldsymbol{w}_0, E, T, \gamma, \mu$

Result: $\boldsymbol{g}_t$

Initialization;

$\boldsymbol{g}_0 \leftarrow$ initial value depends on the learning task;

$\boldsymbol{Q} \leftarrow$ array of size T;

$U' \leftarrow U_F \cup U_R$;

for $t \leftarrow 0$ **to** $T-1$

do

Broadcast μ_t for all clients;

for n ***in*** U' *In Parallel*

do

$\boldsymbol{w}_{\boldsymbol{n,t}}^{\boldsymbol{(0)}} \leftarrow \boldsymbol{g_t}$

for $i \leftarrow 1$ **to** E **do**

$\boldsymbol{w}_{\boldsymbol{n,t}}^{\boldsymbol{(i)}} \leftarrow \boldsymbol{w}_{\boldsymbol{n,t}}^{\boldsymbol{(i-1)}} - \gamma \boldsymbol{g_n}(\boldsymbol{w}_{\boldsymbol{n,t}}^{\boldsymbol{(i-1)}})$

end

$\boldsymbol{w}'_{n,t} \leftarrow \boldsymbol{w}_{\boldsymbol{n,t}}^{\boldsymbol{(E)}} \cdot \left(1 + \frac{(1-\Omega_n)}{10}\right)$;

transmit $\boldsymbol{w}_{n,t}$ **if** $\mathrm{SINR}_n^{(t)} > \mu_t$

then

Add user n to U'';

end

$n \leftarrow n+1$

end

$\boldsymbol{g}_{t+1} \leftarrow \boldsymbol{g}_t + \frac{1}{U''} \sum_{n=1}^{U''} \frac{1}{S_{n,t}} (\boldsymbol{w}'_{n,t} - \boldsymbol{g}_t)$

$\boldsymbol{Q}[t] \leftarrow$ evaluate the accuracy at $(\boldsymbol{g_{t+1}})$;

transmit $\boldsymbol{g}_{t+1}$ to all users ;

if $\boldsymbol{Q}[t]$ *is smaller than the* μ *preceding elements* **then**

$U' \leftarrow U_F$;

end

end

return Result;

To achieve robust and reliable model aggregation in federated learning, it is crucial to consider both the quality of wireless communication and the trustworthiness of participating devices. The following equation embodies this principle by incorporating both wireless and trust-related factors into the aggregation process:

$$\boldsymbol{g_{t+1}} \leftarrow \boldsymbol{g_t} + \frac{1}{U''} \sum_{n=1}^{U''} \frac{\mathbb{1}\{\mathrm{SINR}_n^{(t)} > \mu_t\}}{S_{n,t}(\mu_t, r_n)} (\boldsymbol{w}'_{n,t} - \boldsymbol{g_t}), \tag{3.9}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function which takes the value 1 when $\{\cdot\}$ is true and zero otherwise. As the communication rounds progress, the updating formula (3.9), which accounts for the cumulative contributions of eligible users, leads to an accumulation of noise arising from the assumptions in the adversary model. This cumulative effect reaches a critical point, detrimentally impacting the overall model. At this point, including U_R no longer contributes positively to the system; on the contrary, it degrades the performance. Consequently, the system dynamically transitions to exclusively considering U_F to optimize and fine-tune overall performance. The dynamic transition can be achieved by utilizing a trust window μ. The system only considers authenticated users if the global accuracy evaluated at the AS decreases within the trusted window μ. It is worth noting that this aggregation function simplifies to FedAvg when all users are in close proximity to the base station and are fully trusted. Under these conditions, the standard FedAvg algorithm can be employed without any additional modifications. The convergence analysis of FedAvg over wireless networks is detailed in [31]. Building on this foundation, our proposed algorithm incorporates a trustworthiness metric to enhance the performance and robustness of federated learning in more complex and adversarial environments. By introducing this metric,

we dynamically adjust the participation of users based on their trust scores. This adjustment is particularly crucial when dealing with non-IID data distributions and potential adversarial attacks.

3.4 Numerical Results

We consider the density of the AS and the UAV λ to be $50/km^2$, and the simulation area is 3000 x 3000 km^2. Each AS has 30 RBs; hence, it can serve up to 30 devices, each transmitting at the same power P of 10 dBm. . Regarding the ML model, we use a 2-layer convolutional neural network followed by two fully connected layers while having the optimizer set to be a momentum stochastic gradient descent (SGD) with one epoch, a momentum of 0.5, and a learning rate γ of 0.01. We trained the model on the MINSIT digits classification dataset [49].

3.4.1 Terrestrial Network

The path loss exponent η is set to 4 for the urban environment. Concerning the modeling parameter of trustworthiness, the coefficients α and β are picked differently to achieve different trustworthiness means. For a trustworthiness mean of 0.95, we used 11 and 1, respectively. For a trustworthiness mean of 0.85, we used 5 and 1, respectively. Finally, for a trustworthiness mean of 0.75, we used 3 and 1, respectively. We set ρ and κ to be 0.9 and 0.3, respectively. The dynamic threshold associated with the SINR is established from 10 dB to 0 dB with a 0.25 step size.

3.4.2 Aerial Network

The UAV is positioned at an elevation h of 45 m. The path loss exponent for NLOS χ_L is set at 2.5, while NLOS χ_N is configured to 4. Furthermore, the main lobe beamwidth (θ) for both the UAV and users' devices is designated as 40°. In addition, a main gain lobe of 3.162 is assumed for both the UAV and the devices, accompanied by a side gain lobe of 1. We chose the parameters of the trustworthiness metric to get a mean of $\nu = 70$ and $\nu = 83$. To achieve this, we use $\alpha = 5, \beta = 1$ and $\alpha = 5, \beta = 2$, respectively. We set κ and ρ to be 0.9 and 0.3, respectively. The dynamic threshold associated with the SINR ranges from 10 dB to 1 dB with a step size of 0.25.

We discern the following three distinct cases for comparison for the numerical results.

- **Conservative:** only fully trusted users U_F are considered.
- **Risk-Agnostic**: all users in U_F and U_R are considered untill the end of the simulation.
- **Risk-Aware:** all users in U_F and U_R are considered until global accuracy starts degrading and then switching to U_F only for fine-tuning.

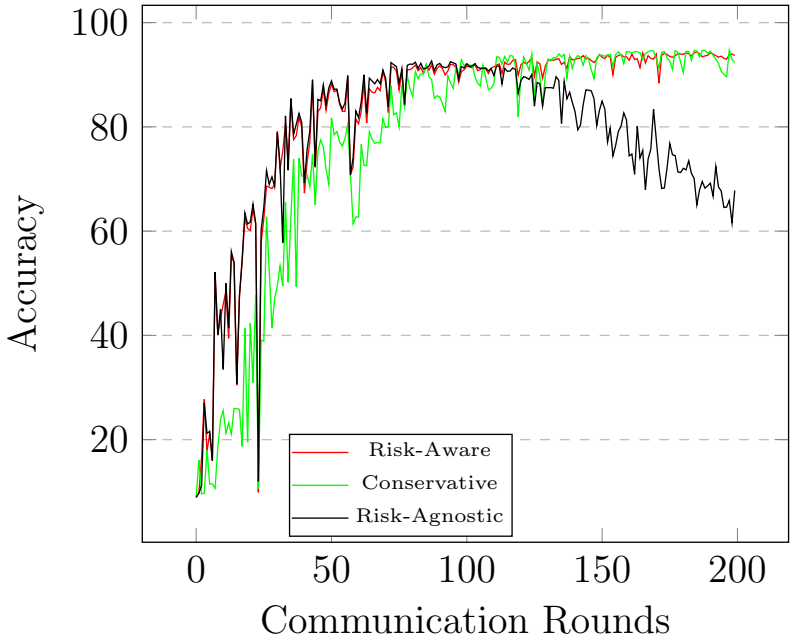

Figure 3.3: Accuracy over communication rounds for mean 0.95

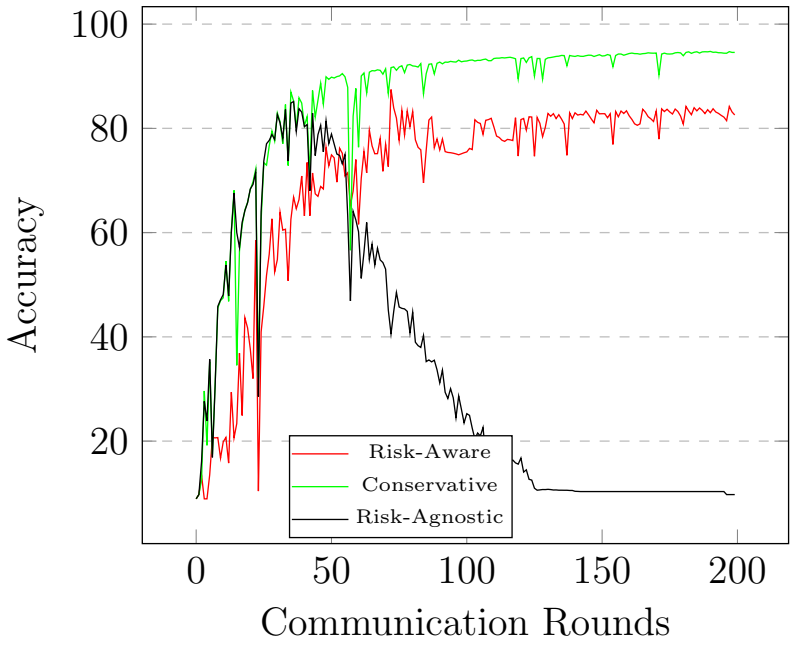

Figure 3.4: Accuracy over communication rounds for mean 0.85

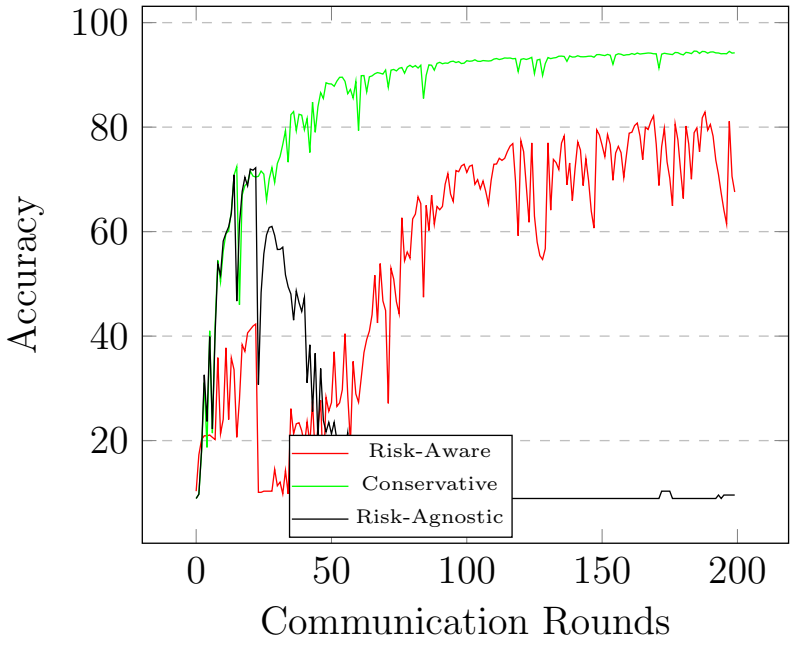

Figure 3.5: Accuracy over communication rounds for mean 0.75

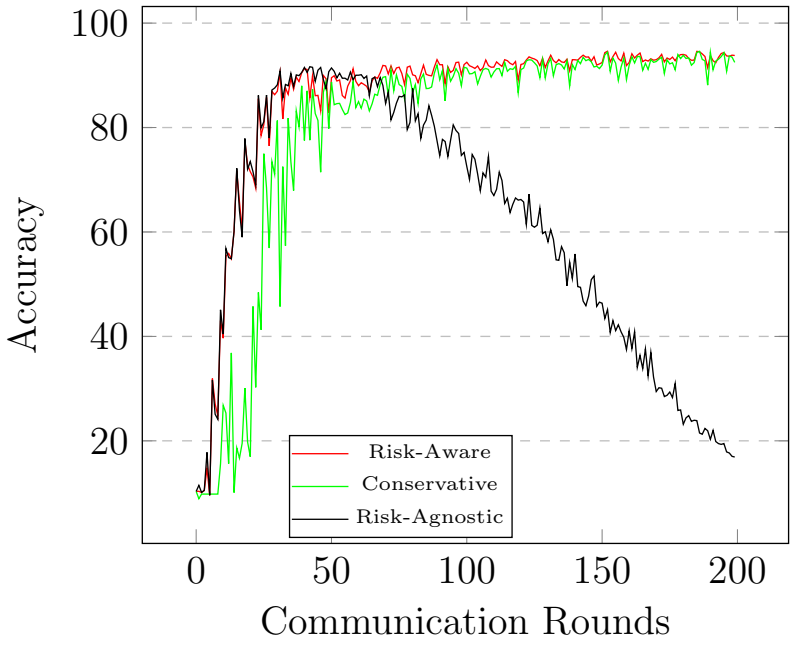

Figure 3.6: Accuracy over communication rounds for mean 0.90

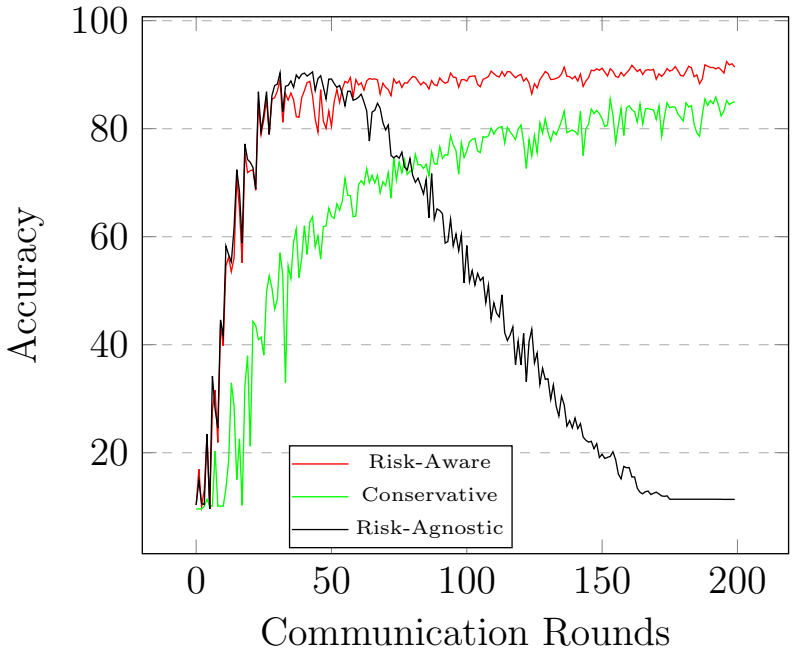

Figure 3.7: Accuracy over communication rounds for mean 0.80

From the numerical results, several insights emerge. First, adopting a conservative approach is insufficient regarding the convergence rate and the maximum accuracy achieved. This effect is more noticeable in networks with lower trustworthiness, as shown in Figs. 3.3, 3.4, 3.5, 3.6 and 3.7. Second, there is a specific point that further consideration of risky users U_R will negatively affect the overall system due to the cumulative added noise for each user. Third, the risk-aware approach, which involves fully trusted and risky users in the early rounds and fine-tuning the model using only the authenticated users, results in the highest accuracy and a high convergence rate. The gap between the two approaches widens as the trustworthiness mean gets lower. Another crucial observation lies in the paradoxical outcome that, despite neglecting a portion of the users, there is an increased accuracy. Remarkably, the model does not exhibit bias towards the retained users. This intriguing phenomenon is attributed to

the fact that, when certain users are removed from the training process, the remaining users may still possess data corresponding to the same classes as those omitted. This intricacy ensures that removing specific users does not compromise the model's accuracy. Moreover, the amount of data required to fine-tune an existing model is less than the required to construct it.

Chapter 4

RARE-FL: Resilient Accelerated and Risk-Aware Edge Federated Learning in Scarce Data Scenario

4.1 Contribution

Current FL schemes are inadequate for scenarios with scarce data and/or stealthy edge devices where a validation dataset is absent. This letter covers such a gap by proposing a resilient, accelerated, and risk-aware edge FL (RARE-FL) that utilizes Trustworthiness Metric (TM) scores to provide fast and trustworthy Federated Learning as a Service (FLaaS) without relying on a validation dataset. Inspired by transfer learning, we formulate a novel RARE-FL aggregation function that progressively reduces the influence of edge devices in the learning process in which the magnitude of the reduction is determined by the device TM score compared to the average TM scores across the network. The proposed RARE-FL follows the descending Signal-to-Interference-Plus-Noise-Ratio (SINR) inclusion of edge devices with a per-device decaying function to ensure accelerated convergence with fair contribution from devices. The numerical results in a 6G non-terrestrial network underscore higher accuracy and faster convergence of the proposed RARE-FL compared to state-of-the-art models.

For both FLaaS and network model, we adopt the same models in sections 3.2.1, 3.2.2 and 3.2.3.

4.2 RARE-FL ALGORITHM

This section details the FL model in (3.1) for the proposed RARE-FL scheme, where we define the per-device weighting factor a_n and utilize the following aggregation function:

$$\boldsymbol{g_{t+1}} = \boldsymbol{g_t} + \sum_{n=1}^{K} \underbrace{p_n \xi_n}_{a_n} W_{n,t} \times (\boldsymbol{w'}_{n,t}(\Omega_n) - \boldsymbol{g_t}), \tag{4.1}$$

where $p_n = {}^{D_n}/_{D}$, D_n is the amount of data possessed at device x, $D = \sum_{n=1}^{K} D_n$ is the total amount of available data, $\boldsymbol{w'}_{n,t}(\Omega_n)$ is the reported local model given by (3.4), $w_{n,t}$ is a wireless communication factor and ξ_n is a trustworthy factor detailed in the sequel.

4.2.1 Wireless Communications Factor

The aggregation function (4.1) is performed at the UAV, and the local model update (i.e., in (3.2) and (3.4)) is performed at the edge device. Hence, at each iteration t, each device must transmit its own local update to the UAV for aggregation. A descending transmission rate is employed in the learning phases to accelerate the FL convergence. That is, at the t^{th} round, the local model is transmitted at a rate of:

$$\mathcal{R}_t = B \log_2(1 + \mu_t), \tag{4.2}$$

such that $\mu_0 > \mu_1 > \cdots > \mu_{T-1}$. Due to path loss and fading, the uplink transmission of the local model is successfully received at the UAV with probability $\mathbb{P}\{\text{SINR}_{x,t}^{(z)} >$

$\mu_t\}$, where the instantaneous (i.e., at the t^{th} round) uplink SINR is expressed as:

$$\text{SINR}_{x,t}^{(z)} = \frac{PG_0H_{z,t}\left(\sqrt{h^2+r_n^2}\right)^{-\chi_z}}{I_L + I_N + N_0}, \tag{4.3}$$

where $z \in \{N, L\}$ indicates a LOS or NLOS transmission, χ_z is the path loss exponent, P is the uplink transmission power, $H_{z,t}$ is the instantaneous fading power gain, N_0 is the background noise, I_{L_t} and I_{N_t} are the instantaneous interference originating from LOS and NLOS communication.

Starting with high transmission rates shortens the duration of initial rounds at the cost of limiting the aggregation to high SINR devices. Proceeding with the learning phases, the transmission rate requirements are relaxed, and more devices are included in the learning process to fine-tune the FL model. To ensure fair contribution from all devices, the wireless communication factor in (4.1) is expressed as:

$$W_{n,t} = \frac{\mathbb{1}(\text{SINR}_{x,t} > \mu_t)}{\mathbb{P}\{\text{SINR}_{n,t} > \mu_t\}}, \tag{4.4}$$

where the numerator ensures the inclusion of only successfully transmitted models, and the denominator amplifies the contribution of devices with lower transmission success probability. Following well-celebrated stochastic geometry analysis for uplink UAV networks [50], the distance-dependent average success probability can be

obtained as:

$$\mathbb{P}\{\text{SINR}_{n,t} > \mu_t\} = \sum_{k=1}^{z_N} (-1)^{k+1} \binom{z_N}{k} \exp\left(-N_0 \zeta_N\right) \mathcal{L}_L\left(\zeta_N\right) \mathcal{L}_N\left(\zeta_N\right) P_N(r_n) + \sum_{k=1}^{z_L} (-1)^{k+1} \binom{z_L}{k} \exp\left(-N_0 \zeta_L\right) \mathcal{L}_L\left(\zeta_L\right) \mathcal{L}_N\left(\zeta_L\right) P_L(r_i), \quad (4.5)$$

where $\mathcal{L}_L(\cdot)$ and $\mathcal{L}_N(\cdot)$ are the Laplace transforms (LT) of the probability density functions of the LOS and NLOS interfering links, respectively, and $\zeta_z = \frac{g_z (r_n^2 + h^2)^{\chi_z/2} \mu_t}{P G_0}$ for $z \in \{N, L\}$. Following the stochastic geometry analysis for Nakagami-m fading environments [29], the LTs in (4.5) can be expressed as:

$$\mathcal{L}_z(s) = \exp\left\{ -2\pi\lambda \sum_{q=0}^{3} \mathbb{P}_q \int_0^{\infty} (1 - \exp\{-\pi\lambda q^2\}) \left(1 - \left[1 + \frac{s P G_q (q^2 + h^2)^{-\chi_z/2}}{m_z}\right]^{-m_z}\right) q \, dq \right\}. \quad (4.6)$$

4.2.2 Trustworthy Factor

The factor $W_{n,t}$ is meant to reduce the impact of the wireless communication impairments, to ensure accelerated and unbiased FL model aggregation. To further account for the TM scores, we implement the following trustworthy factor:

$$\xi_n = \mathbb{1}(\Omega_n > \kappa) \exp\{-(1 - \Omega_n)(1 - \nu) \mathbb{P}\{\text{SINR}_{n,t} > \mu_t\} t\}, \quad (4.7)$$

where κ is the minimum acceptable TM score and $\nu = \sum_{n=1}^{k} \frac{\Omega_n}{K}$ is the average TM score across the contributing edge devices.

The trustworthy factor in (4.7) is inspired by the concept of transfer-learning for scarce data scenarios [51, 52]. The inspiration stems from the utilization of transfer

learning to pre-trained models (i.e., on close but not similar scenarios) that are then fine-tuned by the scarce data of the intended scenario. Such knowledge transfer has been shown to outperform overfitted models solely trained via the scarce data of the intended scenario. We can project the same criterion to our use case by considering that transfer learning is initialized with distorted data (i.e., the data of the pre-trained model) and is fine-tuned by accurate data (i.e., data of the intended scenario). On the same line of thought, devices with lower TM scores are included at initial training rounds to accelerate convergence and avoid overfitting while devices with higher TM scores are used for fine-tuning. Based on the TM score, the exponentially decaying factor $\exp\{1-(1-\Omega_n)(1-\nu)\mathbb{P}\{\mathrm{SINR_{n,t}} > \mu_t\}t\}$ is crafted to gradually diminish the influence of distorted models in a proportion that is relative to their TM score, which reflects the magnitude of model distortion. Such progressive elimination exploits all the available data while keeping devices with higher TM scores (i.e., more accurate data) for more iterations that fine-tune the model and improve the overall accuracy. Nevertheless, (4.7) utilizes the indicator function $\mathbb{1}(\Omega_n > \kappa)$ to completely exclude devices with intolerably low TM scores (i.e., highly distorted local models) from the aggregation function to avoid steep FL accuracy divination.

Unlike transfer learning, the proposed RARE-FL has no hard boundary between the pre-training and fine-tuning phases. In contrast, the RARE-FL presents a per-device soft transition between the pre-training and fine-tuning phases via the trustworthy factor in (4.7). In particular, the four components that determine the per-iteration contribution of each device in the FLaaS transaction are detailed below

1. **Per-device TM Factor:** The per-device factor $(1-\Omega_x)$ reflects the reported local model's distortion magnitude. Hence, devices with lower TM scores are

eliminated faster to ensure that devices with more accurate local models have greater influence on the fine-tuned FL model. Note that fully trusted $\Omega_x = 1$ devices are never eliminated from (4.7).

2. **Universal TM Factor:** A lower ν implies, on average, more distorted local models among the edge devices. Hence, the universal factor (1-ν) ensures a faster transition from pre-training to fine-tuning for scenarios with higher distortions in local models (i.e., lower ν), and vice versa.

3. **Per-device Participation Factor:** The factor ($\mathbb{P}\{\text{SINR}_{n,t} > \mu_t\} \times t$) accounts for the distant-dependent probability of success for the local model transmissions. Hence, accounting for the successful participation of each device in the global rounds rather than the absolute iteration index.

4. **The Indicator Function** The factor $\mathbb{1}(\Omega_x > \rho)$ excludes devices with highly distorted local models to avoid steep accuracy divination.

1. **Per-device TM Factor:** The per-device factor $(1-\Omega_n)$ reflects the magnitude of distortion of the reported model. Hence, devices with lower TM scores are eliminated faster to ensure that devices with more accurate local models have greater influence on the fine-tuned FL model. Note that fully trusted $\Omega_n = 1$ devices are never eliminated from (4.7).

2. **Universal TM Factor:** A lower ν implies, on average, more distorted local models among the edge devices. Hence, the universal factor (1-ν) ensures a faster transition from pre-training to fine-tuning for scenarios with higher distortions in local models (i.e., lower ν), and vice versa.

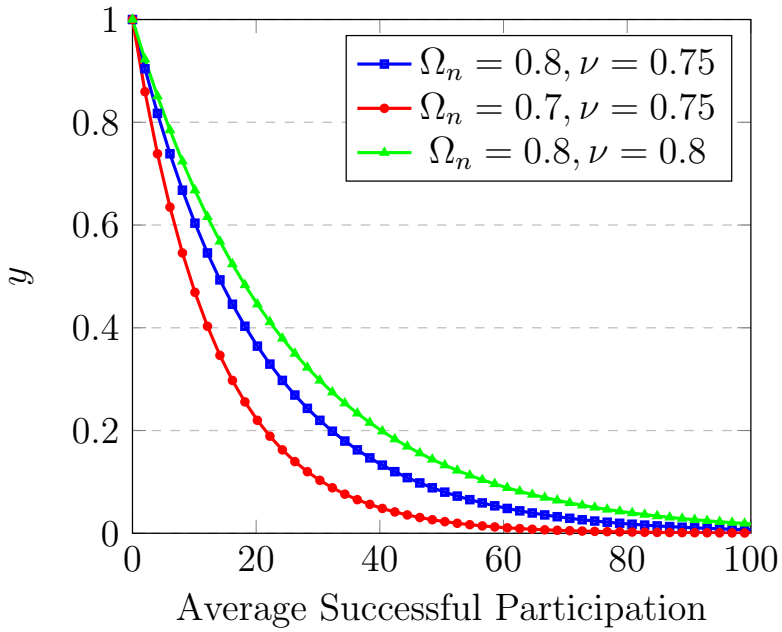

Figure 4.1: Illustration of decaying function's behavior for different values of ν and Ω_n.

3. **Per-device Participation Factor:** The factor $(\mathbb{P}\{\text{SINR}_{n,t} > \mu_t\} \times t)$ accounts for the distant-dependent probability of success for the local model transmissions. Hence, accounting for the successful participation of each device in the global rounds rather than the absolute iteration index.

4. **The Indicator Function** The factor $\mathbb{1}(\Omega_n > \kappa)$ excludes devices with highly distorted local models to avoid steep accuracy divination.

A pictorial illustration for the trustworthy factor in (4.7) is shown in Fig. 4.1 for different values of Ω_n and ν. The figure shows the diminishing contribution of the devices as a function of successful contribution in the model aggregation. As shown in the figure, the trustworthy factor in (4.7) enforces faster decay for devices with lower Ω_n and/or for scenarios with lower ν to ensure sufficient rounds of fine-tuning

with accurate data.

4.2.3 The RARE-FL Algorithm

The RARE-FL algorithm, as detailed in Algorithm 1, orchestrates dynamic and robust FLaaS transactions. The algorithm starts with an initial global model $\boldsymbol{g}_0$ sent from a UAV to all K devices to commence local training using (3.2). Then, devices transmit their local models back to the UAV, as in (3.4). At the UAV, devices are included in the aggregation to update the global model according to the weights defined in (4.1). Hence, mitigating potential deviation due to wireless channel impairments and local model distortions guarantees resilient, trustworthy, and accurate FLaaS transactions.

It is worth mentioning that, by virtue of the aggregation function in (4.1) with the trustworthy factor defined in (4.7), **Algorithm 2** is self-sufficient. Different from the trustworthy FL in chapter 3 the UAV for the proposed RARE-FL does not need to possess a validation dataset to either validate the reported local models or to track the convergence of the global model.

Algorithm 2: Aerial Wireless Federated Safeguard

Data: $T, \mu_t, E, \gamma_n, \Omega_n$

Result: $\boldsymbol{g}_t$

Initialization $t \leftarrow 0$;

$\boldsymbol{g}_0 \leftarrow$ Initial Model;

while $t < T$ **do**

Broadcast μ_t for all clients;

$n \leftarrow 0$;

while $n \leq K-1$ ***for each client***

do

$\boldsymbol{w}_{n,t}^{(0)} \leftarrow \boldsymbol{g}_t$;

$e \leftarrow 0$;

while $e \leq E-1$ **do**

$\boldsymbol{w}_{n,t}^{(e+1)} \leftarrow \boldsymbol{w}_{n,t}^{(e)} - \gamma_n \nabla F_n(\boldsymbol{w}_{n,t}^{(e)})$;

$e \leftarrow e+1$

end

$\boldsymbol{w'}_{n,t} \leftarrow \boldsymbol{w}_{n,t}^{(E)} \cdot \left(1 + \frac{(1-\Omega_n)}{10}\right)$;

transmit $\boldsymbol{w'}_{n,t}$;

$n \leftarrow n+1$;

end

At the BS:

$\boldsymbol{g}_{t+1} \leftarrow \boldsymbol{g}_t + \sum_{n=1}^{K} p_n \xi_n W_{n,t} \times (\boldsymbol{w'}_{n,t} - \boldsymbol{g_t})$;

transmit $\boldsymbol{g}_{t+1}$ to all users ;

$t \leftarrow t+1$

end

return Result;

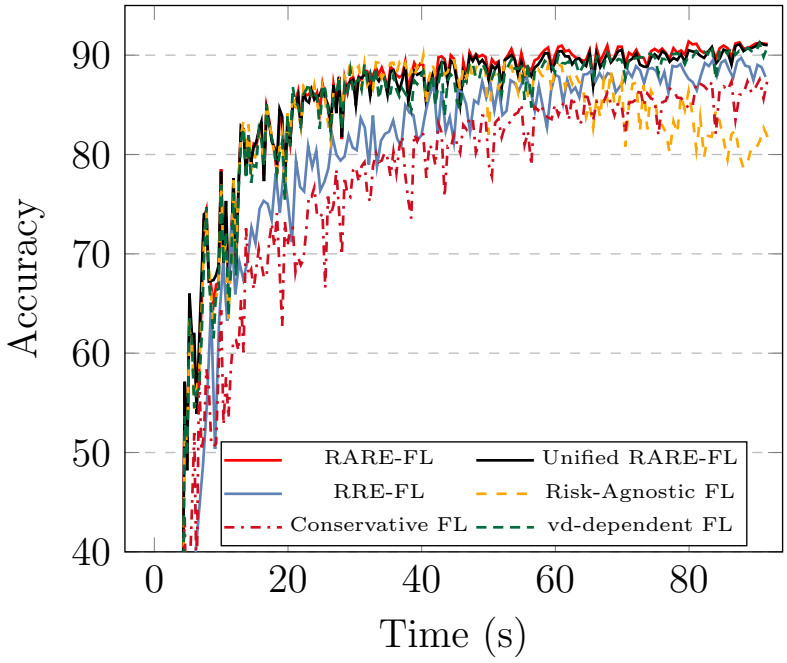

Figure 4.2: Accuracy over time for mean $\nu = 0.85$

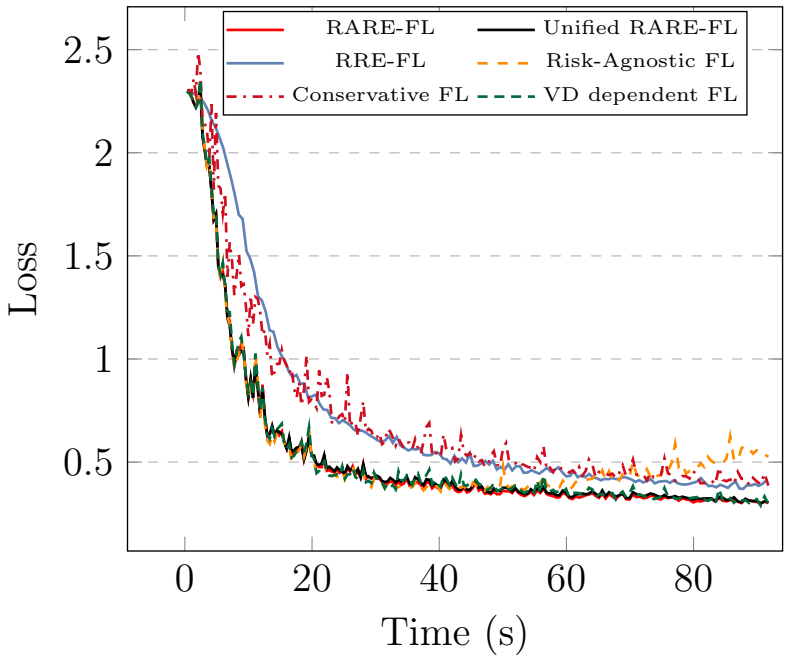

Figure 4.3: Loss over time for mean $\nu = 0.85$

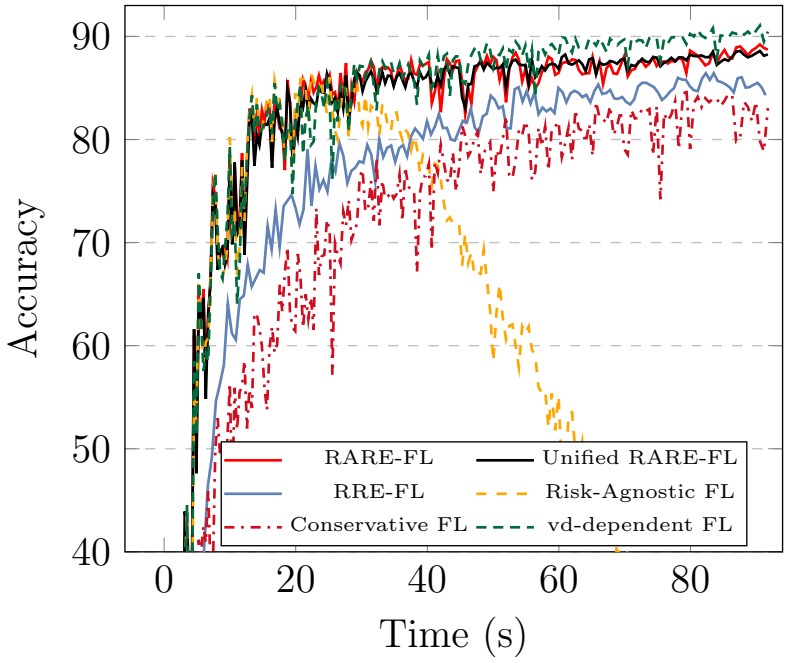

Figure 4.4: Accuracy over time for mean $\nu = 0.7$

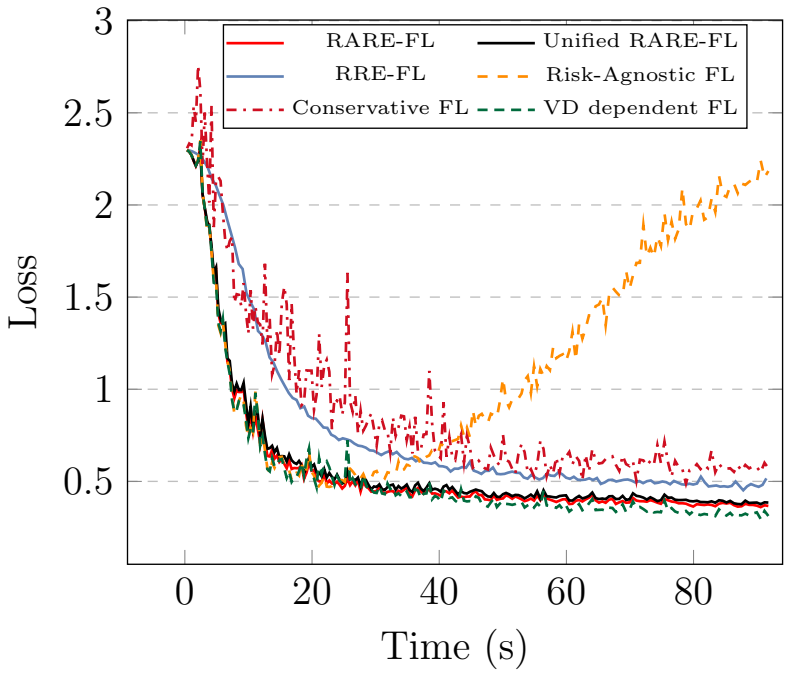

Figure 4.5: Loss over time for mean $\nu = 0.7$

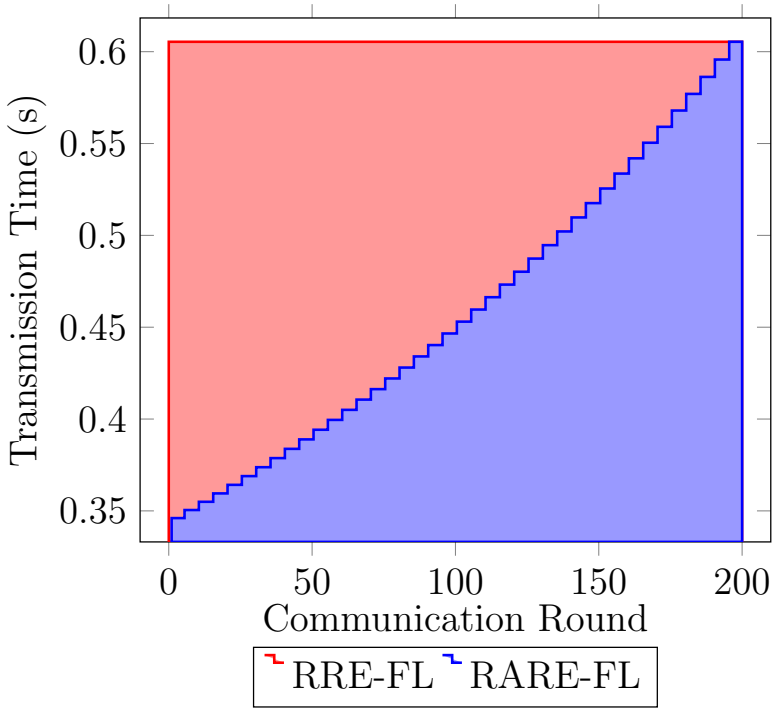

Figure 4.6: Transmission time of RARE-FL vs RRE-FL

4.3 Numerical Results

The FLaaS is implemented on a non-terrestrial network in which the locations of the UAVs are generated via a 2-D Poisson Point Process with intensity (λ) 50/km^2 located at a height (h) of 45m on an area of 3000×3000 km^2. The UAVs are equipped with RBs, each allowing them to serve up to $K = 30$ concurrent devices. The devices operate over $B = 1$ MHz channels with an uplink transmission power of 10 dBm.

The path loss exponent for signal propagation is set to be $\chi_N = 4$ for NLoS and $\chi_L = 2.5$ for LoS links. The transmission of both UAVs and devices is characterized by a main lobe beamwidth (θ) of 40 degrees with a main lobe gain of 5 dBi and a side lobe gain of 0 dBi. For accelerated FLaaS, a descending SINR threshold τ_t is employed from 5 dB to 1 dB with a step size of -0.1 dB. For the trustworthiness landscape, we consider

K_t fully trusted devices (i.e., with unity TM score) and $K_r = K - K_t$ risky devices with TM scores randomly drawn from a beta distribution with shape and scale parameters α and β. Two scenarios with different levels of scarcity of trustworthy devices are considered. Trustworthiness mean $\nu = 0.85$ by setting $K_t = K_r = 15$, $\alpha = 10$ and $\beta = 2$ and trustworthiness mean $\nu = 0.7$ by setting $K_t = 10$, $K_r = 20$, $\alpha = 10$ and $\beta = 3.75$. We also introduce a threshold $\kappa = 0.3$ to detect and eliminate devices with intolerable local deviations. Our learning architecture leverages the MNIST dataset [49], a standard benchmark in machine learning, to train a Convolutional Neural Network (CNN). The training consists of one epoch per round for each participant with the a local model size of 90 kB. We benchmark the proposed RARE-FL against the following approaches:

- **Risk-Agnostic FL**: an aggressive strategy employing an aggregation function oblivious to the TM scores by setting $\kappa_x = 1$. Hence, equally combining accurate and inaccurate local models.

- **Conservative FL**: a strategy employing a binary inclusion/exclusion rule to consider only fully trusted devices in the aggregation function by setting $\kappa_x = \mathbb{1}\{\gamma_x = 1\}$. Hence, completely avoiding inaccurate local models.

- **Validation-dataset (VD) dependent FL**: a strategy introduced in chapter 3 that utilizes a validation dataset to track of the global model convergence and exclude risky devices upon model accuracy degradation.

- **Two variants of RARE-FL:**

    1. **Unified RARE-FL**: This strategy eliminates the per-device success probability from (4.7) and utilizes the following trustworthy factor $\kappa_x = \mathbb{1}(\gamma_x >$

$\rho) \exp\{-(1-\gamma_x)(1-\mu)t\}$ that is a function of the absolute iteration index t. Hence, it emphasizes the importance of fair device participation across the global rounds.

2. **RRE-FL**: This is a resilient risk-aware edge (RRE) FL that operates at the minimum SINR-level allowing cell-edge devices to participate in all global rounds. Hence, explicitly emphasising the impact of the descending SINR allocation for accelerated convergence.

Figs. 4.2, 4.3, 4.4 and 4.5 demonstrate how RARE-FL and its variants outperform the conservative and risk-agnostic schemes. On one hand, the conservative scheme achieves low accuracy and high loss due to the scarcity of data. On the other hand, the risk-agnostic scheme shows good initial accuracy and loss followed by a steep degradation in both due to distorted local models' accumulated effect. By virtue of the proposed aggregation function in (4.7), the proposed RARE-FL benefits from all available data to avoid overfitting while keeping only accurate models for fine-tuning to achieve high accuracy and low loss. The VD-dependent scheme shows slightly higher accuracy and lower loss than the proposed RARE-FL, which is no surprise given the assumption of the available validation dataset at the UAV. A validation dataset enables continuous monitoring of the global model accuracy and optimal switching between initial training and fine-tuning phases. Nonetheless, the availability of the validation dataset can be questionable in many scenarios. when comparing Figs. 4.2 and 4.4, it is clear that the accuracy gain of RARE-FL is higher as the scarcity of trustworthy data increases (i.e., lower ν).

We compare the RARE-FL to its variants in Figs. 4.2, 4.3, 4.4 and 4.5 to justify

its composition. The RRE-FL shows slower convergence when compared to RARE-FL, which quantifies the impact of the dynamic SINR allocations. Interestingly, the RRE-FL also shows lower final accuracy and higher final loss, which can be attributed to the larger number of risky devices involved per global round. The Unified RARE-FL shows slightly lower accuracy and higher loss than the RARE-FL becuase farther devices contribute less to the overall learning. Due to the dynamic SINR allocation, farther devices are included at later iterations (i.e., high t) of the RRE-FL, which entails a lower contribution to the aggregation due to the exponentially decaying trustworthy factor κ_t in the iteration round t.

Fig. 4.6 further emphasises the impact of descending SINR allocation on the FL convergence speed. In particular, the figure shows the time required to transmit the local model at rate $B\log_2(1+\tau_t)$ in the uplink to the UAV. For the RARE-FL, the descending $\tau_0 < \tau_1 < \cdots < \tau_t < \cdots < \tau_{T-1}$ implies shorter transmission times for the initial rounds as shown in the blue staircase curve of Fig. 4.6. On the other hand, the RRE-FL utilizes the minimum τ_{T-1} for all rounds for inclusive devices contribution across all aggregation rounds, which shows a constant transmission time in Fig. 4.6. By virtue of the dynamic transmission time for the local models, the RARE-FL shows the faster convergence, high accuracy and lower loss as reported in Figs. 4.2, 4.3, 4.4 and 4.5.

Chapter 5

Conclusions

5.1 Summary and Conclusion

This thesis investigates methods for enhancing the robustness of Federated Learning as a Service (FLaaS), specifically tackling the challenges arising from client heterogeneity in terms of communication quality and trustworthiness. The research focuses on two key areas that build upon each other to create a comprehensive solution.

1. **Risk-Aware Accelerated Wireless Federated Learning with Heterogeneous Users (Chapter 3)**: in this chapter, we explored the domain of FLaaS, a widely recognized distributed machine learning paradigm, we conducted a thorough analysis uncovering the inherent heterogeneity among clients. This heterogeneity manifested in variations related to security, computational resources, and closeness to the base station. To address the challenges posed by this diversity, we developed a dynamic system that strategically operates at a high data rate during the initial stages of communication rounds, thereby effectively speeding up the learning process. As the learning advances, the system adjusts its parameters, aiming to seamlessly accommodate a broader user base.

Our experimental findings underscored the significance of a well defined proposed approach to being risk-aware. Contrary to the conservative approach of disregarding users with moderate trust scores, we discovered that such exclusion did not necessarily yield enhancements to the system. Likewise, adopting a Risk-Agnostic approach, which involves incorporating all fully trusted and risky, users results in significant model degradation in the long run. The pivotal contribution of our work lies in introducing a new algorithm that considers the combined impact of trustworthiness and wireless network impairments. This innovative algorithm combines elements from both conservative and Risk-Agnostic approaches, demonstrating a substantial enhancement in various network trust environments.

2. **Resilient Accelerated & Risk-Aware Edge Federated Learning (RARE-FL) in Scarce Data Scenario (Chapter 4)**: Building upon the foundations laid in Chapter 3, this chapter introduces a resilient, accelerated, risk-aware edge FL (RARE-FL). In the FLaaS ecosystem, edge devices are assigned a trustworthy metric (TM). We proposed a novel aggregation function that scales the contributions of edge devices according to their data amount, wireless channel quality, and TM score. The wireless channel quality factor is designed to reduce the impact of wireless communication impairments, ensuring an accelerated and unbiased global model. Inspired by transfer learning, the trustworthy factor is crafted to diminish the contribution of the devices to the global learning phases in proportion to their TM scores. Hence, all data for initial training should be exploited to accelerate convergence and avoid overfitting while fine-tuning with trustworthy data to improve the overall FL accuracy. The proposed RARE-FL

is validated using the MNIST dataset over a 6G wireless network where unmanned aerial vehicles (UAV) provide FLaaS to terrestrial edge devices. The numerical results demonstrate the superior performance of the proposed RARE-FL compared to conventional risk-agnostic and conservative FL schemes. The RARE-FL shows more significant performance gain in scenarios with a higher scarcity and lower loss of trustworthy devices.

5.2 Recommendations and Future Work

This thesis has laid the foundation for secure and reliable FLaaS through the development of novel TM systems and attack mitigation techniques. Building upon this work, several avenues for future research and development emerge.

5.2.1 Dynamic Trustworthiness Metric Updates

A crucial aspect for future exploration is the development of mechanisms to update device TM scores dynamically. This would enable the system to adapt to evolving user behavior and contribution patterns. Potential approaches include:

1. **Contribution-Based Updates:** Similar to the approach in [53], TM scores could be adjusted based on the measured contribution of each user during the learning process. This would incentivize positive participation and allow users to improve their reputation over time.

2. **Performance-Based Updates:** TM scores could be linked to the performance of the models trained using the data provided by each user. Devices consistently contributing to accurate and robust models would see their TM scores increase, reflecting their trustworthiness.

5.2.2 Enhancing FLaaS Architecture Resilience

Further research is needed to enhance the resilience and robustness of the FLaaS architecture against a broader range of potential threats. This could involve implementing Multi-layered Defense Strategies, a tiered defense system with multiple protection levels could provide comprehensive security. The first layer could employ statistical methods, as discussed in Section 2.3.2, part 2, to detect and filter out significant deviations caused by malicious users. Subsequent layers could utilize the TM-based approaches developed in Chapters 3 and 4 to identify and mitigate more subtle attacks from imperfect or stealthy users.

5.2.3 Transfer Learning with Dynamic Transition

Chapter 4 introduced a dynamic transition approach within the FLaaS framework. Initially, the system leverages data from all available users to expedite the training process. As the learning progresses and trust is established, the system transitions to exclusively utilizing data from trusted users for fine-tuning the model, ensuring greater security and reliability.

This concept of dynamic transition holds significant potential beyond the specific application presented in this thesis. It can be extended to various domains within Transfer Learning (TL) where data are scarce for the target task. Instead of relying on a binary decision of either including or excluding data from a source domain, a more nuanced approach can be adopted.

Bibliography

[1] C. Mazzocca, N. Romandini, M. Mendula, R. Montanari, and P. Bellavista, "TruFLaaS: Trustworthy federated learning as a service," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21 266–21 281, 2023.

[2] N. Kourtellis, K. Katevas, and D. Perino, "FLaaS: Federated learning as a service," in *Proceedings of the 1st Workshop on Distributed Machine Learning*, ser. DistributedML'20. New York, NY, USA: Association for Computing Machinery, 2020, p. 7–13. [Online]. Available: https://doi.org/10.1145/3426745.3431337

[3] M. Akbari, A. Syed, W. S. Kennedy, and M. Erol-Kantarci, "Constrained federated learning for aoi-limited sfc in uav-aided mec for smart agriculture," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 1, pp. 277–295, 2023.

[4] Y. Nabil, H. ElSawy, S. Al-Dharrab, H. Mostafa, and H. Attia, "Data aggregation in regular large-scale IoT networks: Granularity, reliability, and delay tradeoffs," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 767–17 784, 2022.

[5] A. Ramadan, N. Zorba, and H. S. Hassanein, "Uplink cluster-based radio resource scheduling for hetnet mmtc scenarios," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 1–6.

[6] O. Naserallah, S. B. Azmy, N. Zorba, and H. S. Hassanein, "Impact of users' mobility on the quality of edge sensing systems," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 4196–4201.

[7] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6g: Applications, challenges, and opportunities," 2021.

[8] H. Xia, Y. Li, C. Liu, and Y. Zhu, "Stochastic client scheduling with dynamic sinr thresholds for fast federated learning," in *2022 IEEE/CIC International Conference on Communications in China (ICCC)*, 2022, pp. 632–637.

[9] T. Xia, A. Ghosh, and C. Mascolo, "FLea: Improving federated learning on scarce and label-skewed data via privacy-preserving feature augmentation," 2023.

[10] S. A. Khowaja, K. Dev, S. M. Anwar, and M. G. Linguraru, "SelfFed: Self-supervised federated learning for data heterogeneity and label scarcity in iomt," 2023.

[11] A. Tariq, M. A. Serhani, F. Sallabi, T. Qayyum, E. S. Barka, and K. A. Shuaib, "Trustworthy federated learning: A survey," 2023.

[12] K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, and J. Qadir, "Explainable, trustworthy, and ethical machine learning for healthcare: A survey," *Computers in Biology and Medicine*, vol. 149, p. 106043, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522007569

[13] S. Tyagi, I. S. Rajput, and R. Pandey, "Federated learning: Applications, security hazards and defense measures," in *2023 International Conference on Device*

Intelligence, Computing and Communication Technologies, (DICCT), 2023, pp. 477–482.

[14] M. Hao, H. Li, G. Xu, S. Liu, and H. Yang, "Towards efficient and privacy-preserving federated deep learning," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[15] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. Vincent Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *Trans. Info. For. Sec.*, vol. 15, p. 3454–3469, jan 2020. [Online]. Available: https://doi.org/10.1109/TIFS.2020.2988575

[16] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[18] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," in *Machine learning techniques for multimedia: case studies on organization and retrieval*. Springer, 2008, pp. 21–49.

[19] H. B. Barlow, "Unsupervised learning," *Neural computation*, vol. 1, no. 3, pp. 295–311, 1989.

[20] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[22] S. Grossberg, "Recurrent neural networks," *Scholarpedia*, vol. 8, no. 2, p. 1888, 2013.

[23] S. Ruder, "An overview of gradient descent optimization algorithms," 2017.

[24] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," 2018. [Online]. Available: https://arxiv.org/abs/1806.00582

[25] L. Gao, H. Fu, L. Li, Y. Chen, M. Xu, and C.-Z. Xu, "FedDC: Federated learning with non-iid data via local drift decoupling and correction," 2022.

[26] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, A. Singh and J. Zhu, Eds., vol. 54. PMLR, 20–22 Apr 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[27] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017. [Online]. Available: http://arxiv.org/abs/1602.05629

[28] M. NAKAGAMI, "The m-distribution—a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. HOFFMAN, Ed. Pergamon, 1960, pp. 3–36. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780080093062500054

[29] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, 2013.

[30] R. Zhagypar, N. Kouzayha, H. ElSawy, H. Dahrouj, and T. Y. Al-Naffouri, "Characterization of the global bias problem in aerial federated learning," *IEEE Wireless Communications Letters*, vol. 12, no. 8, pp. 1339–1343, 2023.

[31] M. Salehi and E. Hossain, "Federated learning in unreliable and resource-constrained cellular wireless networks," *IEEE Transactions on Communications*, vol. 69, no. 8, pp. 5136–5151, 2021.

[32] X. Cao and N. Z. Gong, "Mpaf: Model poisoning attacks to federated learning based on fake clients," 2022.

[33] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, S. Chiappa and R. Calandra, Eds., vol. 108. PMLR, 26–28 Aug 2020, pp. 2938–2948. [Online]. Available: https://proceedings.mlr.press/v108/bagdasaryan20a.html

[34] V. Tolpegin, S. Truex, M. E. Gursoy, and L. Liu, "Data poisoning attacks against federated learning systems," in *Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part I 25*. Springer, 2020, pp. 480–501.

[35] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: byzantine tolerant gradient descent," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 118–128.

[36] D. Yin, Y. Chen, K. Ramchandran, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," 2021.

[37] L. Yi, X. Shi, W. Wang, G. Wang, and X. Liu, "FedRRA: Reputation-aware robust federated learning against poisoning attacks," in *2023 International Joint Conference on Neural Networks (IJCNN)*, 2023, pp. 1–8.

[38] X. Cao, M. Fang, J. Liu, and N. Z. Gong, "FLTrust: Byzantine-robust federated learning via trust bootstrapping," in *28th Annual Network and Distributed System Security Symposium, NDSS 2021, virtually, February 21-25, 2021*. The Internet Society, 2021. [Online]. Available: https://www.ndss-symposium.org/ndss-paper/fltrust-byzantine-robust-federated-learning-via-trust-bootstrapping/

[39] E. M. El Mhamdi, R. Guerraoui, and S. Rouault, "The hidden vulnerability of distributed learning in Byzantium," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning

Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 3521–3530. [Online]. Available: https://proceedings.mlr.press/v80/mhamdi18a.html

[40] D. Yin, Y. Chen, R. Kannan, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 5650–5659. [Online]. Available: https://proceedings.mlr.press/v80/yin18a.html

[41] O. A. Wahab, G. Rjoub, J. Bentahar, and R. Cohen, "Federated against the cold: A trust-based federated learning approach to counter the cold start problem in recommendation systems," *Information Sciences*, vol. 601, pp. 189–206, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025522003632

[42] X. Lin, J. Wu, C. Zhou, S. Pan, Y. Cao, and B. Wang, "Task-adaptive neural process for user cold-start recommendation," in *Proceedings of the Web Conference 2021*, ser. WWW '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 1306–1316. [Online]. Available: https://doi.org/10.1145/3442381.3449908

[43] Z. Yin, K. Li, and H. Bi, "Trusted multi-domain ddos detection based on federated learning," *Sensors*, vol. 22, no. 20, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/20/7753

[44] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal lap altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[45] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 403–430, 2017.

[46] A. Gholami, N. Torkzaban, and J. S. Baras, "Trusted decentralized federated learning," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, 2022, pp. 1–6.

[47] R. Kumar and R. Goyal, "Performance based risk driven trust (prtrust): On modeling of secured service sharing in peer-to-peer federated cloud," *Comput. Commun.*, vol. 183, no. C, p. 136–160, feb 2022. [Online]. Available: https://doi.org/10.1016/j.comcom.2021.11.013

[48] N. Tezuka, H. Ochiai, Y. Sun, and H. Esaki, "Resilience of wireless ad hoc federated learning against model poisoning attacks," in *2022 IEEE 4th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA)*. Los Alamitos, CA, USA: IEEE Computer Society, dec 2022, pp. 168–177. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/TPS-ISA56441.2022.00030

[49] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[50] Y. Nabil, H. ElSawy, S. Al-Dharrab, H. Attia, and H. Mostafa, "Ultra-reliable device-centric uplink communications in airborne networks: A spatiotemporal analysis," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 9484–9499, 2023.

[51] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "Safe and accelerated deep reinforcement learning-based O-RAN slicing: A hybrid transfer learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 310–325, 2024.

[52] T. Balachandran, T. Abreu, M. Naloufi, S. Souihi, F. Lucas, and A. Janne, "IoT and transfer learning based urban river quality prediction," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 257–262.

[53] G. Wang, C. X. Dang, and Z. Zhou, "Measure contribution of participants in federated learning," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 2597–2604.

[54] H. Alzer, "On some inequalities for the incomplete gamma function," *Math. Comput.*, vol. 66, no. 218, p. 771–778, apr 1997. [Online]. Available: https://doi.org/10.1090/S0025-5718-97-00814-4

Appendix

Proofs

This appendix details the proof of lemmas 1 and 2

.1 Lemma 1 Proof

Lemma 1 can be proved as follows:

$$\begin{aligned}
\frac{1}{S_{n,t}(\zeta_t, r_n)} &= \frac{1}{\mathbb{P}(\mathrm{SINR}_n^{(t)} > \zeta_t)} \\
\frac{1}{S_{n,t}(\zeta_t, r_n)} &= \frac{1}{\mathbb{P}\left(\frac{Ph_0^2 r^{-\eta}}{N_0 + I_{agg}} > \zeta_t\right)} \\
&\overset{(a)}{=} \frac{1}{\mathbb{E}_{Iagg}\left[\exp\left(\frac{-\zeta_t(N_0+I_{agg})}{Pr^{-\eta}}\right)\right]} \\
&\overset{(b)}{=} \frac{1}{\exp\left(\frac{-\zeta_t N_0}{Pr^{-\eta}}\right) \cdot \mathcal{L}\left(\frac{\zeta_t}{Pr^{-\eta}}\right)} \\
&= \frac{\exp\left(\frac{\zeta_t N_0}{Pr^{-\eta}}\right)}{\mathcal{L}\left(\frac{\zeta_t}{Pr^{-\eta}}\right)}
\end{aligned}$$

where (a) is achieved by the cumulative distribution function (CDF) of the exponential distribution of the fading and (b) is obtained by substituting the expectation on the aggregated interference. Note that $\mathcal{L}(s)$ is the Laplace transform of the probability

density function (PDF) of the aggregated interference.

.2 Lemma 2 Proof

Lemma 2 can be proved as follows:

The success probability for a LoS client is given by

$$(S_{i,t}(\mu_t, r_i)|i = L) = \mathbb{P}(\text{SINR}_i^{(t)} > \mu_t)$$
$$(S_{i,t}(\mu_t, r_i)|i = L) = \mathbb{P}\left(\frac{PG_0h_0^2(r_i^2 + h^2)^{-\chi_L/2}}{N_0 + I_{Los} + I_{NLos}} > \mu_t\right)$$

since h_0^2 follows a gamma distribution, using Alzer's inequality in [54] gives us an approximate of

$$\approx \sum_{n=1}^{m_L}(-1)^{n+1}\binom{m_L}{n}\cdot$$
$$\mathbb{E}_{Iagg}\left[\exp\left(\frac{-S_L n\mu_t\left(N_0 + I_L + I_N\right)}{PG_0(r_i^2 + h^2)^{-\chi_L/2}}\right)\right],$$

where $S_L = m_L(m_L!)^{\frac{-1}{m_L}}$

$$= \sum_{n=1}^{m_L}\left((-1)^{n+1}\binom{m_L}{n}\exp\left(\frac{-\mu_t N_0 S_L n(r_i^2 + h^2)^{\chi_L/2}}{PG_0}\right)\cdot\right.$$
$$\mathcal{L}_{\mathcal{L}}\left(\frac{S_L n(r_i^2 + h^2)^{\chi_L/2}\mu_t}{PG_0}\right)\cdot\mathcal{L}_{\mathcal{N}}$$
$$\left.\left(\frac{S_L n(r_i^2 + h^2)^{\chi_L/2}\mu_t}{PG_0}\right)\right).$$

Following the same steps for a NLOS client, we replace m_L with m_N and S_L with S_N

have

$$(S_{i,t}(\mu_t, r_i)|i = N) = \frac{\sum_{n=1}^{m_i}(-1)^{n+1}\binom{m_L}{n}}{\exp\left(\frac{\mu_t N_0 S_N n(r_i^2+h^2)^{\chi_L/2}}{PG_0}\right)} \cdot \mathcal{L}_L\left(\frac{S_N(r_i^2+h^2)^{\chi_N/2}\mu_t}{PG_0}\right) \cdot \mathcal{L}_N\left(\frac{S_N(r_i^2+h^2)^{\chi_N/2}\mu_t}{PG_0}\right)$$

Now, considering the law of total probability,

$$\begin{aligned}(S_{i,t}) &= (S_{i,t}|i = L)P_{Los}(r_i) + (S_{i,t}|i = N)P_N(r_i) \\ &= \frac{\sum_{n=1}^{m_L}(-1)^{n+1}\binom{m_L}{n}}{\exp\left(\frac{\mu_t N_0 S_L n(r_i^2+h^2)^{\chi_L/2}}{PG_0}\right)}\mathcal{L}_L\left(\frac{S_L(r_i^2+h^2)^{\chi_L/2}\mu_t}{PG_0}\right) \cdot \\ &\mathcal{L}_N\left(\frac{S_L(r_i^2+h^2)^{\chi_L/2}\mu_t}{PG_0}\right)P_L(r_i) + \\ &\frac{\sum_{n=1}^{m_i}(-1)^{n+1}\binom{m_L}{n}}{\exp\left(\frac{\mu_t N_0 S_N n(r_i^2+h^2)^{\chi_L/2}}{PG_0}\right)} \cdot \mathcal{L}_L\left(\frac{S_N(r_i^2+h^2)^{\chi_N/2}\mu_t}{PG_0}\right) \cdot \\ &\mathcal{L}_N\left(\frac{S_N(r_i^2+h^2)^{\chi_N/2}\mu_t}{PG_0}\right)P_N(r_i),\end{aligned}$$

which proves Lemma (2).