Security for the Internet of Things in Terrestrial and Non-Terrestrial Networks

Saud Khan

January, 2025

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University

Australian
National
University

School of Engineering
College of Engineering, Computing and Cybernetics
The Australian National University

Dedication

"Hain rawaan us raah par, jis ki koi manzil na ho
Justuju karte hain us ki, jo hamen hasil na ho

Dasht-e-najd-e-yaas mein deewangi ho har taraf
Har taraf mahmil ka shak ho, par kahin mahmil na ho

Wahm ye tujh ko ajab hai, ai jamal-e-kam-numa
Jaise sab kuch ho, magar tu deed ke qabil na ho

Wo khada hai ek bab-e-ilm ki dehleez par
Main ye kehta hoon use, is khauf mein dakhil na ho

Chahta hoon main munir is umr ke anjaam par
Ek aisi zindagi, jo is tarah mushkil na ho"

- Munir Niazi

Declaration

The contents of this thesis, in whole or any part of it, have not been submitted to this or any other university for a degree.

The research work presented in this thesis has been performed jointly with Prof. Salman Durrani (The Australian National University), Assoc. Prof. Xiangyun Zhou (The Australian National University), Dr. Seyit Camtepe (Data61, CSIRO), Dr. Chandra Thapa (Data61, CSIRO), Prof. Sarah J. Johnson (University of Newcastle), and Dr. Basit Shahab (University of Newcastle). The substantial majority of this work was my own.

Saud Khan
School of Engineering,
College of Engineering, Computing and Cybernetics,
The Australian National University,
Canberra, ACT 2601,
Australia

Acknowledgments

Completing this PhD thesis has been a long and challenging journey, and I could not have reached this milestone without the support and encouragement of numerous individuals.

- First and foremost, I would like to express my heartfelt gratitude to my primary supervisor, Prof. Salman Durrani, for his invaluable guidance and unwavering support. Your mentorship has been the cornerstone of this thesis, and I am profoundly grateful for the opportunity to grow as a researcher under your wing. Your support, both professionally and personally, has significantly shaped the person I am today. Working under your supervision has been an honour, and I sincerely hope to make you proud as I venture out of the nest. Thank you for everything.

- I would also like to extend my heartfelt thanks to my supervisory panel: Dr Chandra Thapa, Dr Seyit Camtepe, Assoc Prof. Xiangyun (Sean) Zhou and Assoc Prof. Nan (Jonas) Yang for their support during my PhD. Your guidance and technical contribution during my PhD research have been immensely impactful. The collaborative spirit and intellectual environment you have provided to me have greatly enriched my research experience. I extend my gratitude for your time and energy in providing your input every time I have requested it.

- In the end, I would like to thank myself as well for not giving up.

Abstract

The Internet of Things (IoT) revolution is transforming various sectors by enabling seamless connectivity and data exchange between devices. However, this rapid expansion brings significant security challenges due to IoT devices' constrained resources and diverse nature. This thesis presents a comprehensive security framework addressing these challenges through three interconnected components: detection, identification, and authentication. This integrated approach is crucial for establishing a robust IoT security framework capable of mitigating various threats effectively.

First, we focus on the detection and identification components by investigating the multi-user detection (MUD) problem in uplink grant-free non-orthogonal multiple access (NOMA). This scenario involves identifying the number of active IoT devices and decoding their transmitted data without prior knowledge of device activity levels. The proposed solution leverages an attention-based bidirectional long short-term memory (BiLSTM) network, exploiting the temporal correlation of IoT device transmissions. The BiLSTM network processes the device activation history through forward and reverse-pass LSTMs, while the attention mechanism highlights crucial activation points. This approach forms a hierarchical pathway for detecting active IoT devices and performs blind data detection using complex spreading sequences. The results indicate that this method significantly outperforms existing benchmark schemes, providing superior detection accuracy and flexibility without requiring prior knowledge of device sparsity or channel conditions.

Then, we investigate the IoT authentication component by introducing an innovative physical-layer authentication scheme tailored for terrestrial IoT devices with limited computational capabilities. The proposed scheme utilises the inherent properties of the IoT devices' transmission model for seed generation and continuous authentication. The scheme eliminates the need for repeated key generation and verification by verifying access time slots and spreading sequences. This approach reduces computational overhead and enhances security by concealing seed information from potential attackers. The results demonstrate a near threefold reduction in the misdetection rate of illegitimate

devices and a false alarm rate close to zero, even with varying numbers of active devices and signal-to-noise ratios. The scheme boasts at least half the computational cost of benchmark methods, underscoring its practicality for real-world IoT deployments.

Finally, we address the unique security challenges associated with IoT authentication in non-terrestrial Low Earth Orbit (LEO) satellite-based IoT networks. Traditional terrestrial authentication methods, such as Authentication and Key Management for Applications (AKMA), are inadequate for LEO networks due to their dynamic environment and frequent handovers. This work proposes a modified AKMA framework incorporating local key refreshment for decentralised and continuous authentication. This modification reduces the need for repeated authentication attempts with satellites, mitigating the risks of man-in-the-middle and spoofing attacks. The framework's performance is evaluated in the presence of an illegitimate Unmanned Aerial Vehicle (UAV), showing improved authentication rates for legitimate devices and reduced misdetection rates for illegitimate devices compared to existing shared key and physical channel-based authentication schemes. The modified AKMA framework demonstrates its applicability and effectiveness in enhancing security for LEO satellite-based IoT networks.

In summary, this thesis presents a holistic IoT security framework that effectively addresses the critical detection, identification, and authentication components. Each component offers significant advancements in its respective domain, and their integration forms a comprehensive framework to safeguard IoT devices against a broad spectrum of security threats. This work not only contributes valuable insights into IoT security but also provides practical solutions that can be implemented to ensure the secure operation of IoT networks in various environments. The results underscore the importance of a multi-faceted approach to IoT security, paving the way for future research and development in this vital field.

Contents

List of Publications

The research work of my Ph.D. candidature has been published, accepted, or submitted for publication as journal papers or conference proceedings in [1–9]. For ease of reference, [1–9] are denoted by [J1-J6] and [C1-C3], respectively. Note that this thesis is based on J1-J3 and C1. The other publications are not used in this thesis. They are listed as follows:

Journal Papers

J1. **S. Khan**, S. Durrani, M. B. Shahab, S. J. Johnson, S. Camtepe, "Joint User and Data Detection in Grant-Free NOMA With Attention-Based BiLSTM Network," *IEEE Open J. Commun.*, vol. 4, pp. 1499 – 1515, Jul. 2023 (Chapter 1)

J2. **S. Khan**, C. Thapa, S. Durrani, S. Camtepe, "Access-Based Lightweight Physical-Layer Authentication for the Internet of Things Devices," *IEEE Internet Things J.*, vol. 11, no. 7, pp.11312 – 11326, Nov. 2023 (Chapter 2)

J3. **S. Khan**, S. Durrani, C. Thapa, S. Camtepe, "Modified AKMA for Decentralized Authentication in LEO Satellite-Based IoT Networks," *accepted in IEEE Internet Things J.*, Jan. 2025 (Chapter 3)

J4. S. Ahmad, **S. Khan**, K. S. Khan, F. Naeem, M. Tariq, "Resource Allocation for IRS-Assisted Networks: A Deep Reinforcement Learning Approach," *IEEE Comms. Stand. Mag.*, vol. 7, no. 3, pp. 48 – 55, Sep. 2023

J5. F. Naeem, G. Kaddoum, **S. Khan**, K. S. Khan, N. Adam, "IRS-Empowered 6G Networks: Deployment Strategies, Performance Optimization, and Future Research Directions," *IEEE Access*, vol. 10, pp. 118676 – 118696, Nov. 2022

J6. S. J. Siddiqi, F. Naeem, **S. Khan**, K. S. Khan, M. Tariq, "Towards AI-enabled traffic management in multipath TCP: A survey," *Comput. Commun.*, vol. 181, no. 1, pp. 412-427, Nov. 2021

Conference Proceedings

C1. **S. Khan**, C. Thapa, S. Durrani, S. Camtepe, "Beyond Key-Based Authentication: A Novel Continuous Authentication Paradigm for IoTs," *in Proc. IEEE GLOBECOM Wkshps*, Kuala Lumpur, Malaysia, Dec. 2023 (Chapter 2)

C2. S. Idrees, X. Jia, **S. Khan**, S. Durrani, X. Zhou, "Deep Learning Based Passive Beamforming for IRS-Assisted Monostatic Backscatter Systems," *in Proc. IEEE ICASSP*, Singapore, Singapore, May 2022

C3. **S. Khan**, S. Durrani, X. Zhou, "Transfer Learning Based Detection for Intelligent Reflecting Surface Aided Communications," *in Proc. IEEE PIMRC*, Helsinki, Finland, Sep. 2021

List of Acronyms

3GPP	3rd Generation Partnership Project
ADMM	Alternative-Direction-Method-of-Multipliers
AI	Artificial Intelligence
AKMA	Authentication and Key Management for Applications
AP	Access Point
AUD	Active User Detection
BER	Bit Error Rate
BHT	Binary Hypothesis Testing
BiLSTM	Bidirectional Long Short-Term Memory
BSASP	Block Sparsity Adaptive Subspace Pursuit
CFR	Channel Frequency Response
CIR	Channel Impulse Response
CS	Compressed Sensing
CSI	Channel State Information
D-AUD	Deep Active User Detection
DFS	Doppler Frequency Shift
DH	Diffie-Hellman
DNN	Deep Neural Network
IORLS	Iterative Order Recursive Least Square
IoT	Internet of Things
IRSA	Irregular Repetition Slotted ALOHA
LEO	Low Earth Orbit

LS-OMP	Least Squares Orthogonal Matching Pursuit
LSTM	Long Short-Term Memory
LTE	Long-Term Evolution
ML	Machine Learning
MMSE	Minimum Mean Square Error
MMV	Multiple Measurement Vector
MUD	Multi-User Detection
MUSA	Multi-User Shared Access
NFSR	Non-Linear Feedback Shift Register
NOMA	Non-Orthogonal Multiple Access
NN	Neural Network
OF	Overloading Factor
PIAASP	Prior-Information Aided Adaptive Subspace Pursuit
QPSK	Quadrature Phase Shift Keying
RSSI	Received Signal Strength Indicator
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The rapid advancement and integration of the Internet of Things (IoT) into modern society are transforming countless sectors, including healthcare, agriculture, manufacturing, and urban development [10,11]. The seamless connectivity and intelligent data exchange between devices that IoT enables are not just conveniences but are becoming essential components of digital infrastructure [12]. This infrastructure is foundational to developing smart cities, autonomous vehicles, and advanced manufacturing systems [13,14], and it is contributing to the IoT market's unprecedented growth rate globally. As of 2024, there are an estimated 18 billion IoT devices worldwide, and this number is expected to reach 22 billion by 2026 [15]. The IoT sector's market value is projected to exceed $1 trillion globally by 2026, reflecting its expanding role in various industries [16]. In Australia, the IoT market is also experiencing significant growth, with an estimated value of $9 billion by 2025 [17].

IoT is fundamentally reshaping the landscape of digital communication by embedding sensors, software, and other technologies into everyday objects, enabling them to collect and exchange data. However, while beneficial, this proliferation of interconnected devices introduces significant security challenges that must be addressed to safeguard sensitive information and maintain the integrity of these systems [18]. The unique characteristics of IoT devices, such as their resource constraints, diverse operational environments, and massive scale of deployments, complicate traditional security measures. Unlike conventional computing devices, many IoT devices are designed to be small, inexpensive, and energy-efficient, often lacking the computational power and memory required to support robust security protocols [19]. This makes them more vulnerable to attacks and harder

to secure using traditional methods.

Additionally, the diverse range of IoT applications means that these devices are deployed in various environments, each with its own specific security requirements and challenges. For example, an IoT device used in a smart home environment may face different threats compared to one used in a remote sensing setting. This diversity necessitates a flexible and adaptable approach to IoT security that can address the specific needs of different applications and deployment scenarios. Accordingly, the concept of IoT in terms of deployment extends to two primary domains: cellular networks and satellite networks.

1.1.1 IoT in Terrestrial Networks

Cellular IoT refers to connecting IoT devices using cellular networks such as 5G and the emerging 6G technology. This integration provides wide coverage, mobility, and support for a large number of devices, making it ideal for applications like smart cities, fleet management, and remote monitoring [20, 21]. Cellular networks offer robust infrastructure but also face unique security challenges, including ensuring the confidentiality and integrity of the data transmitted over these widely accessible networks.

The advent of 6G technology is particularly significant for IoT as it promises to deliver higher data speeds, reduced latency, and the ability to connect a vast number of devices simultaneously [22]. These enhancements are crucial for real-time applications such as autonomous driving and smart grid management, where timely data processing and response are critical. However, the widespread deployment of 6G also introduces new vulnerabilities and potential attack vectors that must be addressed to protect IoT ecosystems.

1.1.2 IoT in Non-Terrestrial Networks

In contrast, IoT in non-terrestrial networks utilise satellites to extend IoT connectivity to remote and hard-to-reach areas where terrestrial networks are unavailable. Satellites can provide global coverage, making them crucial for applications like environmental monitoring, disaster management, and global asset tracking [23, 24]. However, the use of satellite networks for IoT also introduces distinct security concerns, such as the vulnerability of satellite communications to interception and jamming.

Satellites are essential for ensuring connectivity in remote regions, including oceans, deserts, and polar areas, where terrestrial infrastructure is impractical or impossible to deploy. This capability is vital for applications like tracking wildlife, monitoring climate

change, and managing natural resources. Nevertheless, the reliance on satellites also necessitates robust encryption and authentication mechanisms to prevent unauthorised access and ensure the integrity of the data transmitted [25, 26].

1.2 Elements of IoT Security

This section explores the IoT security framework that systematically addresses the critical aspects of IoT security. Addressing these elements is crucial for developing robust security mechanisms that can effectively mitigate the unique challenges posed by the diverse and resource-constrained nature of IoT deployments [19].

1.2.1 Detection

The first element of an end-to-end IoT security framework is the detection of devices at the access point (AP). This step is crucial for ensuring that all IoT devices can be identified without bias. Given the resource-constrained nature of IoT devices, they typically remain in a low-power state to conserve energy, waking only to transmit data before returning to sleep. This sporadic activity requires the AP to efficiently detect which IoT devices are active and transmitting data.

Detection must be impartial, not relying on specific device types, and resource-efficient to accommodate the limited capabilities of IoT devices. The burden of detection lies primarily with the AP, which must perform the heavy lifting in detecting active devices without exhausting the devices' resources. An impartial and efficient detection process is the foundational step in securing IoT networks, ensuring that all devices are accurately detected, regardless of their heterogeneity or resource limitations.

1.2.2 Identification

Following detection, the next critical element is the identification of the detected IoT devices. Identification involves recognising the specific devices that have been detected by the AP, even when no prior information about these devices is available. This step is essential because it bridges the gap between merely detecting data transmissions and knowing the origin of these transmissions.

Identification must be performed without prejudice and should not impose significant resource demands on the IoT devices. The AP must be capable of accurately determining which device each data transmission belongs to, ensuring that the network can appropriately address and manage each device. This capability is vital for maintaining a secure

and organised IoT environment, as it allows for the proper allocation of resources and the implementation of targeted security measures.

1.2.3 Authentication

The final element in the IoT security framework is the authentication of the detected and identified devices. Authentication verifies the identity of each IoT device, ensuring that it is authorised to access network resources and transmit data securely. This step is crucial for preventing unauthorised access and protecting sensitive information within the IoT network.

Authentication processes must be lightweight and continuous, similar to detection and identification, to accommodate the diverse and resource-constrained nature of IoT devices. The AP must handle the bulk of the authentication process, minimising the resource usage on the IoT devices. Robust and efficient authentication mechanisms are essential for maintaining the integrity and security of the IoT network, ensuring that all devices are legitimate and can be trusted to operate within the network.

1.3 IoT Security Challenges

This section explores the primary security challenges associated with IoT devices, focusing on their resource constraints, scalability issues, the diversity and complexity of the devices and networks, and the emergent security threats they face. Addressing these challenges is crucial to developing robust security frameworks for IoT ecosystems.

1.3.1 Resource Constraints and Scalability Issues

IoT devices are typically designed to be low-cost and energy-efficient, often resulting in limited computational power and minimal storage capacity. These constraints pose significant challenges in implementing robust security protocols. Traditional security mechanisms, such as encryption and multi-factor authentication, require substantial computational resources and memory, making them impractical for resource-constrained IoT devices [19]. The limitations of IoT devices necessitate the development of lightweight security solutions that can operate within their constraints. For example, physical layer authentication, a lightweight alternative to conventional key-based methods, can be more suitable for IoT devices due to their limited capabilities. This approach leverages the unique properties of the IoT devices' transmission model for authentication, reducing the need for computationally intensive key generation and management processes. However,

even this approach faces limitations such as poor robustness under channel fluctuations and reconciliation overhead, necessitating further research into more efficient and scalable solutions [27, 28].

Scalability is another critical issue in IoT security. The sheer number of devices in an IoT system can be overwhelming, making deploying and maintaining security solutions challenging. The dynamic nature of IoT environments, with devices constantly joining and leaving the network, further complicates scalability. Solutions must be designed to handle this dynamic and vast scale efficiently [29]. For instance, connected devices, such as traffic sensors, streetlights, surveillance cameras, and public transportation systems, can reach millions in a smart city deployment. Ensuring that each device is securely connected and communicating with the central network without introducing significant latency or vulnerability is a monumental task. Traditional security solutions that work well for smaller, static networks may not scale effectively to such a large and dynamic environment [30, 31].

Moreover, the resource constraints of IoT devices mean that any security solution must be lightweight and efficient, both in terms of computational overhead and energy consumption. This requires innovative approaches that can provide robust security without overburdening the devices. Techniques such as lightweight cryptography, secure bootstrapping, and decentralised security protocols are all areas of active research to address these challenges [32, 33].

1.3.2 Diversity of Devices and Complexity of Networks

The diversity of IoT devices adds another layer of complexity to securing IoT ecosystems. These devices range from simple sensors and actuators to more complex entities like smart meters and industrial control systems. Each type of device has different capabilities and security needs, requiring flexible security frameworks that can adapt to these varying requirements [34]. For instance, a simple temperature sensor in a smart home may only need basic authentication and data encryption to ensure that the data it transmits is secure and that the device cannot be easily hijacked. In contrast, a smart meter in an industrial setting may require more sophisticated security measures, such as tamper detection, secure firmware updates, and advanced anomaly detection to prevent unauthorised access and ensure the data's integrity [35].

Lightweight authentication schemes that leverage device-specific attributes, such as access time slots and spreading sequences, can provide continuous and robust authentication for a wide range of IoT devices without imposing significant computational overhead.

These schemes are particularly useful in environments with high device heterogeneity, such as smart cities or industrial IoT, where devices must operate under different conditions and threat models [36]. Network complexity further exacerbates these security challenges. IoT networks often involving multiple interconnected devices form complex topologies. Each connection point can introduce new vulnerabilities, making it crucial to effectively design security protocols to manage and protect these intricate networks. For example, in Low Earth Orbit (LEO) satellite-based IoT networks, the dynamic nature of satellite communication, with frequent handovers and variable latency, necessitates decentralised and continuous authentication frameworks that can operate efficiently under such conditions [37].

1.3.3 Emergent Security Threats

As IoT systems become more prevalent, they attract increasingly sophisticated cyber-attacks. These threats exploit the unique vulnerabilities of IoT networks, such as standard communication protocols and widespread device distribution, making security a moving target [38, 39]. Emergent threats include man-in-the-middle attacks, spoofing, and eavesdropping. In LEO satellite-based IoT networks, for instance, adversaries can exploit the high mobility and frequent handoffs to launch attacks that traditional terrestrial network security protocols cannot counter effectively. This highlights the need for innovative security solutions tailored to the specific characteristics of IoT networks [40]. Similarly, man-in-the-middle attacks involve an attacker intercepting and potentially altering the communication between two devices. In an IoT context, this could mean intercepting the data from a sensor (such as a UAV) and injecting false information, leading to potentially dangerous consequences [41].

Spoofing attacks involve an attacker pretending to be a legitimate device or user to gain unauthorised access to a network or system. In IoT networks, this could involve an attacker masquerading as a legitimate device to gain access to sensitive data or control other devices. This is particularly concerning in scenarios where IoT devices control physical systems, such as in smart homes or industrial automation [42]. Similarly, eavesdropping attacks involve an attacker passively monitoring the communication between devices to gain access to sensitive information. In IoT networks, this could involve intercepting data from sensors or control systems to gather information about the environment or the state of the system. This information could then be used for further attacks or to gain unauthorised access to the network [43].

Moreover, the use of machine learning (ML) and artificial intelligence (AI) in IoT se-

curity presents both opportunities and challenges. While these technologies can enhance threat detection and response by analysing patterns and anomalies in IoT traffic, they also introduce new attack vectors. For example, adversaries could manipulate training data or exploit vulnerabilities in ML algorithms to bypass security mechanisms [44]. Adversaries can also employ techniques such as data poisoning, where they introduce malicious data into the training set, leading to compromised models that fail to detect certain types of attacks or even falsely classify malicious activities as benign. Another sophisticated attack method is adversarial ML, where attackers can force the network to lead to the misclassification of network traffic, allowing malicious activities to go undetected. This is particularly concerning for networks relying on AI for real-time threat detection and response, as it undermines the reliability and effectiveness of these networks.

1.4 Related Works

In the following subsections, we discuss the prior works related to the identified IoT security challenges in the context of the comprehensive security framework of IoT device detection, identification, and authentication in terrestrial and non-terrestrial networks.

1.4.1 Detection and Identification

In wireless communications, orthogonal multiple access (OMA) and NOMA are key resource allocation strategies. OMA assigns distinct resources to devices, whereas NOMA enables multiple users to share resources by leveraging power or code-domain multiplexing. In traditional grant-based OMA schemes, the maximum number of IoT devices being serviced is limited by the number of available orthogonal resources. Therefore, scheduling is required to allow the IoT devices to share the orthogonal resources. In contrast, grant-free NOMA allows IoT devices to transmit their data in an arrive-and-go manner by randomly choosing a resource block without going through the grant-access process [45, 46]. When multiple IoT devices choose the same resource block, a collision occurs, which requires retransmission. These collisions are significantly reduced due to the different multiple access signatures in NOMA [47]. Therefore, from a practical perspective, grant-free NOMA is considered an attractive solution for sporadic IoT traffic use cases.

The basic principle of grant-free NOMA is to allow the IoT devices to randomly access the resource blocks through multiple access signatures, such as power levels, spreading sequences, scrambling, and interleaving [45, Table. III]. Among these signatures, spreading

sequences are considered superior because they can efficiently mitigate multi-user interference [48]. The spreading sequences allow device-specific, low cross-correlation codes to enable grant-free communication. However, in spreading-based signatures, longer-length sequences are needed as the number of IoT devices increases. In this regard, complex spreading sequences, as proposed in multi-user shared access (MUSA) [49], enable support for a significantly larger number of IoT devices than pseudo-random sequences, i.e., a higher overloading factor (OF) without increasing the sequence length.

In spreading-based grant-free NOMA, each active device randomly and independently selects a spreading sequence from a predefined set [50]. Therefore, the key research challenge is to correctly detect the spreading sequences of the active IoT devices, also known as a multi-user detection (MUD) problem [51]. In this regard, identifying the total number of active IoT devices, also known as the active user detection (AUD) sub-problem, and the accuracy of correctly identified active IoT devices, which is the active user support set sub-problem, play a key role.

Identifying an IoT device depends on the quality of the active user support set. This can be derived from Active User Detection (AUD), which directly impacts the performance of IoT device detection and can be classified as a MUD problem. In many practical IoT use cases, while the total number of IoT devices is large, only a small percentage of the total IoT devices may be active in a given time frame [52–54]. Using this inherent sparsity of IoT devices, the AUD problem can be readily formulated as a sparse recovery problem, which can be solved using compressed sensing (CS) [55, 56] or ML [45]. Considering the inherent sparsity and the sporadic device activity, it is crucial to correctly model the activity pattern of IoT devices over a time frame. The activity pattern of IoT devices over a given time frame, whether independent or temporally correlated, greatly impacts the performance of MUD.

1.4.1.1 CS-based Solutions

CS-based solutions with frame-wise sparsity model: Many works have considered CS-based solutions for the MUD problem in spreading based grant-free NOMA with frame-wise sparsity [57–61]. In [57], the frame-wise joint sparsity model is exploited to achieve better performance of device detection using an iterative order recursive least square (IORLS) algorithm based on the orthogonal matching pursuit (OMP) algorithm. However, the authors considered prior knowledge of device sparsity level at the AP, which is typically unknown in practical scenarios. In [58], the authors proposed the alternative-direction-method-of-multipliers-(ADMM)-based CS to show improvement in the device

detection performance using a partial active device set as prior knowledge. However, obtaining the prior information on either the sparsity level, equivalent channel matrices, or both in practical systems is difficult. In [59], the device detection problem was modelled as a multiple measurement vector (MMV) problem, and a block sparsity adaptive subspace pursuit (BSASP) algorithm was used to solve it. However, pilot symbols are transmitted before every data packet, which leads to a significant system overhead. Similarly, the authors in [60, 61] developed greedy algorithms for joint device activity and data detection. However, these algorithms assume complete channel gain knowledge at the AP or pilot symbols for channel estimation.

CS-based solutions with burst sparsity model: Some recent works have considered CS-based solutions for the MUD problem in spreading-based grant-free NOMA with burst sparsity [62–64]. In [62], a dynamic CS-based multi-device detection was proposed, which utilised the temporal correlation between device transmissions in the previous frame to achieve the performance gain. This algorithm was developed based on the assumption that the device sparsity level is known, which requires a training stage to learn such information accurately. Alternatively, the prior-information aided adaptive subspace pursuit (PIAASP) algorithm was proposed in [63], which utilised the prior support according to the additional quality information (the number of common support sets shared in time slots). However, the preceding work is heavily dependent on the inertia of device support; thus, it is unsuitable when the active device support varies rapidly in adjacent time slots, as is often the case in practice. Similarly, the authors in [64] proposed an algorithm to take advantage of the temporal correlation, where the frame is divided into subframes. Each subframe contains adjacent time slots and considers the active and inactive devices sharing common support in all the time slots. Also exploiting the temporal correlation, the authors in [65] used $\ell_{2,1}$ minimisation to jointly detect the user activity and data.

1.4.1.2 ML-based Solutions

Recent works have adopted ML and demonstrated higher detection accuracy than conventional iterative algorithms [66–71]. The authors in [66] and [67] considered pseudo-random noise-based and complex spreading sequences, respectively, and proposed deep neural networks (DNN) for active user detection (D-AUD) in a grant-free NOMA system by using the received signal as the input to the DNN. However, since the preceding works utilised a vanilla DNN for this purpose, the temporal activity of the devices cannot be taken advantage of, leaving room for improvement. To tackle this, the authors in [68]

utilised a long short-term memory (LSTM) network to predict the activity of the devices based on their activation history. However, the dependence of LSTM on the previous activation history of devices makes the overall system prone to misclassification since the activation history is vaguely modelled. Adopting a different approach, the authors in [69] considered the use-case of generative networks to tackle the issue of detecting devices in different OFs with a single trained model. However, this work did not take the temporal correlation of device activity patterns into account. The authors in [71] provided a somewhat different approach by utilising power-domain NOMA instead of code-domain NOMA as the multiple access signature. However, the system faces extreme degradation due to this choice as the number of active devices in the cell increases. Furthermore, pilot symbols are included after every data symbol, drastically increasing the system's overhead. In a similar fashion, the authors in [70] utilised a bi-directional deep neural network for detection in a two-user power domain NOMA scenario. However, this differs from the grant-free NOMA scenario considered in this work since it does not use spreading-based signatures and assumes the connection of devices using prior access procedures. Similarly, the authors in [72] assigned nonorthogonal pilots to devices for transmission, leading to a larger system overhead as the number of devices increases. Thus, a more resilient approach is required in the context of deep learning, which can exploit the temporal correlation of active devices in the adjacent time slots whilst providing accurate detection of devices.

1.4.2 Authentication

IoT devices usually connect to a network through an AP. The conventional approach to establish secure communications between IoT devices and AP is to generate a shared secret key by exploiting the reciprocity of the random fading channel [73, 74]. Herein, the IoT devices measure highly correlated wireless channel characteristics (*e.g.*, channel impulse responses, or received signal strengths) and use them as shared random sources to generate a shared key. However, the low-cost and often resource-constrained IoT devices cannot facilitate physical-channel probing for a shared key generation due to the limited resources. Instead, these IoT devices rely on intermittent transmissions, which makes them highly susceptible to adversarial attacks [75].

Existing methods, such as upper-layer security protocols, suffer from high computational overhead [76]. Conversely, lightweight options are available, but these often rely on physical channel attributes [77, 78] and are unreliable in the presence of variations and noise. Furthermore, channel probing is challenging given the resource limitation of

the IoT devices. This underscores the need for fast (enabled by continuous authentication mechanism), reliable (no reliance on physical channel attributes), and lightweight authentication mechanisms for IoT devices.

Considering the adversaries, upper-layer security protocols have been increasingly studied in the literature [76,79]. However, they are not well suited for resource-constrained IoT devices due to their massive computational overhead and excessive latency. In this regard, low-complexity authentication schemes are desirable for resource-constrained IoT devices, complementing the overall network entropy by introducing additional measures for IoT device authentication in the lower layers [80, 81].

Physical layer security schemes based on keyless authentication [82–84] can provide lightweight security to the resource-constrained IoT devices by exploiting the inherent physical-channel attributes and/or device-specific features of IoT devices. By doing so, the overall network entropy can be improved while reducing IoT devices' computational cost and energy consumption. The authors in [82] introduced scheduling policies to utilise the physical channel characteristics for device authentication. The authors in [83] utilised the channel and phase noise of the physical channel between a transceiver pair utilising multiple antennas for hypothesis testing and device authentication. Similarly, the authors in [84] utilised the correlation of multiple channel impulse responses (CIR) from the physical channel for authentication. Recently, ML has also been applied to combine with physical layer authentication schemes to improve the robustness under channel fluctuations [77,78,85]. However, the reliance of these techniques on the physical channel for feature extraction results in unreliable authentication performance due to variations and noise present in complex dynamic environments.

In a different approach, to achieve continuous authentication, the authors in [86] used an authentication mechanism to create a learning-based kernel model that utilises multi-attributes from the physical channel for device authentication. Then, the authors in [87] utilised the multi-attribute design of the physical channel and support vector machine (SVM) to utilise pseudo-random binary access time slots for device authentication. However, due to the time-varying nature of the physical channel, especially in complex dynamic environments, and the low-cost components utilised by the IoT devices, the variations and noise cause unreliable seed acquisition. Additionally, since these works are based on the assumption of physical channel reciprocity, they will incur a high seed mismatch rate due to the half-duplex nature of the resource-constrained IoT devices; this results in multi-staged parity bits for seed reconciliation, which is against the deployment spirit of resource-constrained IoT devices. Moreover, since the IoT devices are resource-constrained, the physical channel probing process cannot be carried out due to

the inherent sporadic communication nature of the IoT devices.

1.4.3 Non-Terrestrial Networks

The assurance of security through robust authentication in LEO satellite-based IoT networks is indispensable for maintaining the integrity of these networks [88]. The growing dependence on LEO satellites for various applications, ranging from global internet coverage to critical communication infrastructure, underscores the importance of securing these networks against cyber threats. Currently, operational LEO satellite systems such as Starlink and OneWeb function within the 400-2000 km altitude range, with rapid orbital velocities, which necessitates frequent handoffs. This unique operational environment, characterised by high speeds and low earth orbit, introduces significant challenges in ensuring seamless and secure communication. The dynamic nature of these networks renders them vulnerable to adversarial attacks, such as man-in-the-middle and spoofing, potentially instigated by rogue Unmanned Aerial Vehicles (UAVs) [89]. These vulnerabilities highlight the need for innovative security solutions tailored to the unique characteristics of LEO satellite-based IoT networks.

In light of the 3rd Generation Partnership Project (3GPP) Release 17 guidelines, device authentication in 5G networks is secured via the Authentication and Key Management for Applications (AKMA) framework [90]. While this centralised approach is proficient within terrestrial 5G infrastructures, its direct application to LEO satellite-based IoT networks presents significant challenges, emphasising the necessity for a decentralised authentication strategy. Firstly, the inherent latency and connectivity variability characteristic of LEO satellite environments undermines AKMA's efficacy, introducing substantial delays in the authentication process that are unsuitable for latency-sensitive applications. Additionally, the scalability of AKMA is questionable in the face of exponentially growing IoT devices, as its centralised nature might not efficiently support the burgeoning network demand [91]. Further, the continuous authentication method employed by the AKMA framework conflicts with the energy-saving strategies of IoT devices, which are built to transmit data sporadically to save power. This mismatch unintentionally leads to higher energy usage and shorter device lifespans. Furthermore, AKMA's centralised scheme heightens security and privacy vulnerabilities due to the increased risk of man-in-the-middle attacks during prolonged key exchanges [92]. The infrastructure required to implement AKMA across LEO satellite networks significantly elevates operational costs and complexity, a limitation that decentralised methods could mitigate by leveraging existing network metrics for authentication, thus reducing infras-

tructure demands. Lastly, AKMA's rigidity complicates its adaptability to the dynamic nature of LEO satellite network topology, potentially causing authentication delays and service disruptions. Hence, pursuing innovative, decentralised authentication methods becomes imperative, designed to address LEO satellite-based IoT networks' unique operational demands and environmental challenges, ensuring scalable, energy-efficient, and secure communication.

Centralised authentication schemes typically rely on a single entity to manage authentication credentials and processes. A prominent example is 3GPP's Release 17, which highlights AKMA's framework for providing centralised security within the terrestrial 5G infrastructure [90, 92]. However, the AKMA authentication framework in LEO satellite-based IoT networks faces significant challenges. First, AKMA's centralised structure struggles with scalability amidst rapidly growing IoT device numbers, leading to inefficient network support. Its continuous authentication approach contradicts IoT devices' energy-saving strategies by causing increased energy consumption and reducing device lifespan due to the continuous shared key exchange requirement for every session [93]. Moreover, the centralised nature heightens security risks, notably from man-in-the-middle attacks during key exchanges, and escalates operational costs and complexity due to the extensive infrastructure required [94]. Thus, exploring decentralised authentication methods that address LEO satellite-based IoT networks' unique requirements and challenges is critical for ensuring scalable, energy-efficient, and secure communications.

In contrast, decentralised authentication schemes distribute the authentication process across multiple entities, reducing the dependency on a single point of failure and enhancing scalability. These schemes are particularly beneficial in environments with high mobility and dynamic topologies, such as LEO satellite networks. Blockchain technology offers a decentralised and immutable ledger for authentication, ensuring high security and trust [95]. The authors in [96] integrate blockchain with certificateless encryption to provide a secure and scalable authentication framework for LEO satellite networks. This approach mitigates the risks of man-in-the-middle attacks and ensures data integrity by leveraging the distributed nature of blockchain. Moreover, AI-oriented multifactor authentication schemes enhance security by using multiple authentication factors and AI for anomaly detection and response. For instance, the AI-oriented two-phase multifactor authentication in space-air-ground integrated networks employs AI to analyse communication patterns and authenticate devices based on dynamic and static factors [97].

Lightweight authentication is an important feature for centralised and decentralised schemes, mainly when dealing with IoT devices with limited computational and power ca-

pabilities. These schemes often combine cryptographic techniques with efficient key management to ensure security without imposing significant overhead. The authors in [87] utilised a multi-attribute design to utilise pseudo-random binary access time slots for device authentication. Then, the authors in [2, 7] extended this to a two-stage authentication strategy in a terrestrial environment. However, these works are prone to delays or failures in synchronisation, which can lead to discrepancies in user data, potentially allowing unauthorised access or denying access to legitimate devices [98]. Further, the scalability of these works to non-terrestrial networks is unclear due to their strict reliance on physical channels in the terrestrial networks.

1.5 Thesis Overview and Contributions

This thesis presents a comprehensive examination of security measures for IoT devices, highlighting novel strategies to enhance their detection, identification, and authentication across terrestrial and non-terrestrial networks. Figure 1.2 shows an overview of the thesis. The research encompasses cutting-edge machine-learning techniques and advanced cryptographic methods to address the unique challenges posed by IoT environments. The following sections detail the individual contributions of this work, starting with the sophisticated detection and identification processes and culminating in a future-oriented approach to IoT device security. The chapter-wise summary of the contributions of this thesis is given as follows:

1.5.1 Chapter 2: IoT Device Detection and Identification

The initial step in securing IoT devices involves their accurate detection and identification within a network. The first part of my research focuses on enhancing the capability of IoT systems to detect and identify connected IoT devices accurately using grant-free non-orthogonal multiple access (NOMA). This technology is particularly suited for environments with a high density of IoT devices, enabling efficient and scalable communications without the need for scheduled access. Employing advanced ML techniques, such as attention-based bidirectional long short-term memory (BiLSTM) networks, this work aims to improve the detection accuracy and reduce the collision rate among IoT devices, thus facilitating reliable device management and security enforcement. In this context, our main contributions are as follows.

- We design a BiLSTM network with an attention mechanism to carry out AUD. The BiLSTM network utilises two LSTM networks conjunctionally in opposite tempo-

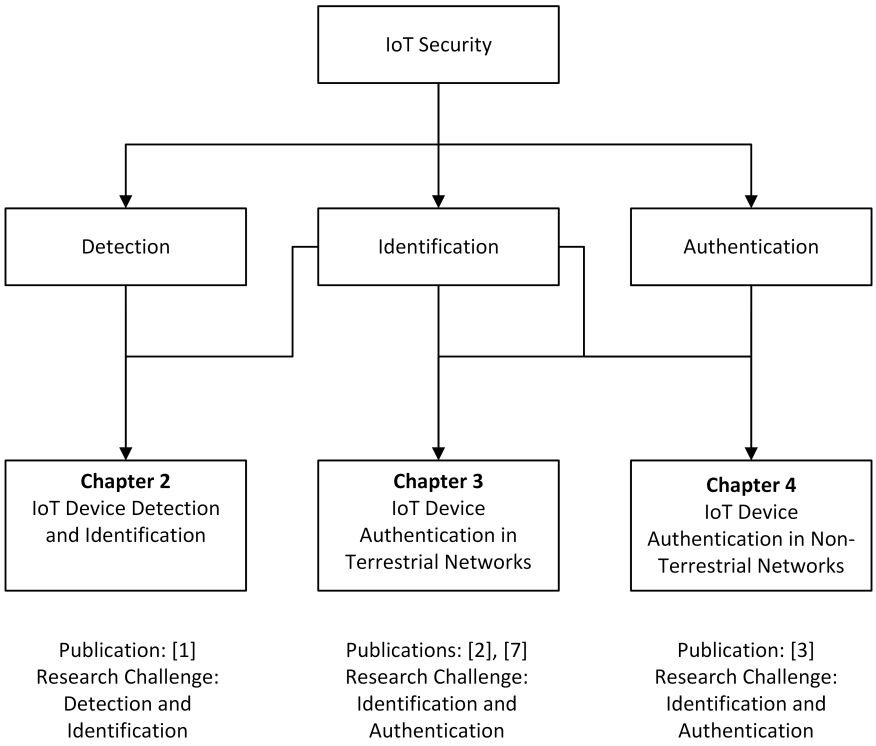


Figure 1.1: The overview of the three technical chapters of the thesis.

ral directions. The attention mechanism exploits the temporal correlation in the active user set and facilitates the BiLSTM network by providing context to the important activation history of the active IoT devices. By training the network in the offline stage, the proposed network maps the superimposed received signal and the active user support set, detecting a larger number of active IoT devices with higher accuracy.

- By detecting the active user support set using our proposed BiLSTM network, we then provide a framework to carry out blind data detection at the AP [99] without the need for explicit channel training. Using the estimated active user support set and complex spreading sequences, a blind minimum mean square error (MMSE) weight is obtained, from which the received signal is reconstructed without the explicit need for statistical channel information.

- Compared to the benchmark OMP scheme, our results show an improvement of around 30% when detecting the number of active IoT devices and an improvement of around 29% when identifying the active device support set. Additionally,

the proposed network achieves a gain of around 2.3 dB in bit-error-rate (BER) compared to the OMP scheme.

- Compared to the ML-aided LSTM-based CS scheme, our results show an improvement of around 10% when detecting the number of active IoT devices and an improvement of around 6% when identifying the active device support set. Additionally, the proposed network achieves a gain of around 0.9 dB in BER compared to the LSTM-based CS scheme. The computational complexity of the proposed network increases only marginally as the number of active IoT devices increases.

The results of this work have been presented in the following publication [1]:

J1. **S. Khan**, S. Durrani, M. B. Shahab, S. J. Johnson, S. Camtepe, "Joint User and Data Detection in Grant-Free NOMA With Attention-Based BiLSTM Network," *IEEE Open J. Commun.*, vol. 4, pp. 1499 – 1515, Jul. 2023.

1.5.2 Chapter 3: IoT Device Authentication in Terrestrial Networks

Following detection and identification, the second chapter addresses the authentication of these IoT devices. Given the resource constraints typical of IoT devices, traditional cryptographic methods can be impractical. Therefore, this research explores physical layer security techniques, which utilise the inherent properties of the communication medium as authentication metrics. This method not only reduces the computational load on IoT devices but also enhances the security of device communications against common threats such as spoofing and replay attacks. The main contributions of this work are summarised as follows.

- **Authentication scheme**: We propose a lightweight authentication scheme comprised of four processes: access time slots generation, spreading pool construction, seed generation, and authentication decision. The scheme provides continuous authentication by checking the access time slots and spreading pools of the IoT devices instead of generating and verifying shared keys.

- **Reduced overhead and latency**: The spreading sequences, utilised by the IoT devices as part of the grant-free NOMA transmission protocol, are used as the seed source for access time slot generation and IoT device authentication. Thus, our proposed scheme does not need seed verification and reconciliation processes, which incur massive overhead and latency.

- **Improved authentication performance**: Our results in the misdetection rate of illegitimate devices indicate a nearly threefold improvement, false alarm rate indicates state-of-the-art, and spreading sequence collision rate indicates superior performance in different settings while boasting a lower complexity compared to the benchmark schemes. Furthermore, our proposed scheme does not rely on the physical channel reciprocity assumption, which makes it a suitable authentication scheme for resource-constraint IoT devices.

The results of this chapter have been presented in the following publications [2, 7]:

J2. **S. Khan**, C. Thapa, S. Durrani, S. Camtepe, "Access-Based Lightweight Physical-Layer Authentication for the Internet of Things Devices," *IEEE Internet Things J.*, vol. 11, no. 7, pp.11312 – 11326, Nov. 2023.

C1. **S. Khan**, C. Thapa, S. Durrani, S. Camtepe, "Beyond Key-Based Authentication: A Novel Continuous Authentication Paradigm for IoTs," *in Proc. IEEE GLOBECOM Wkshps*, Kuala Lumpur, Malaysia, Dec. 2023.

1.5.3 Chapter 4: IoT Device Authentication in Non-Terrestrial Networks

The third part of the thesis extends the detection and authentication frameworks to non-terrestrial environments, specifically LEO satellite networks. This extension is crucial for IoT devices deployed in remote or inaccessible areas, where terrestrial network coverage is inadequate or absent. The research proposes a decentralised authentication framework suitable for the dynamic and challenging environment of satellite communications, addressing the unique demands such as frequent handoffs and variable latency, thereby ensuring continuous and secure device operation across global scales. The main contributions of this work are summarised as follows.

- We propose a decentralised and continuous authentication framework that adapts and modifies the AKMA framework to suit the operational requirements of LEO satellite-based IoT networks. We propose (i) novel seed generation and seed update and (ii) seed refreshing mechanisms for authentication in LEO satellite-based IoT networks. Notably, our framework accounts for the constraints of satellite communications by localising key refreshment mechanisms and employing distinct transmission patterns for IoT devices. By decentralising AKMA's key management process, we alleviate the synchronisation burden from IoT devices amidst the frequent handoffs due to orbiting satellites.

- Our proposed authentication framework enables the satellite to differentiate between legitimate and illegitimate devices through independent generation of transmission patterns. This approach minimises the reliance on physical channel properties and mitigates distance-related and correlated physical channel disparities.
- Our results demonstrate significant improvement in the authentication rate of legitimate IoT devices and a reduction in the misdetection rate of illegitimate devices. Our authentication framework achieves these superior results while ensuring lower bandwidth and computational cost than benchmark schemes.

The results of this chapter have been presented in the following publication [3]:

J3. **S. Khan**, S. Durrani, C. Thapa, S. Camtepe, "Modified AKMA for Decentralised Authentication in LEO Satellite-Based IoT Networks," *accepted in IEEE Internet Things J.*, Jan. 2025.

1.5.4 Chapter 5: Summary and Future Work

Finally, Chapter 5 provides a summary of the thesis results and makes suggestions for future research work.

Chapter 2

IoT Device Detection and Identification

2.1 Introduction

In the first technical chapter, we consider a complex spreading sequences-based grant-free NOMA scenario, where multiple devices communicate with the AP simultaneously in the uplink following a burst-sparsity model. To address the AUD problem, we design an attention-based BiLSTM network, which aims to create a mapping function between the superimposed received signal at the AP and the indices of active devices in the transmit signal. The proposed framework does not require active user sparsity or channel state knowledge to carry out AUD. Using the estimated active user support set, we then design a MUD framework to find the user sparsity and carry out blind data detection at the AP.

The choice of BiLSTM with attention over other architectures stems from a careful consideration of the problem requirements and the limitations of alternative methods. Traditional feedforward neural networks lack the capability to capture sequential dependencies inherent in the burst-sparsity model, making them less effective for AUD. Convolutional neural networks, while effective for spatial feature extraction, are not inherently designed to handle sequential data efficiently. Vanilla LSTMs, although capable of modeling temporal dependencies, fail to fully leverage bidirectional information, which is particularly valuable in detecting the indices of active devices in complex grant-free NOMA settings.

The rest of this chapter is organised as follows. In Section 2.2, we present the system model and MUD problem. Section 2.3 describes the proposed attention-based BiLSTM

Table 2.1: Important symbols used in this work.

Variable	Description	Dimension
K	Total number of IoT devices	1×1
N	Total subcarriers	1×1
S	Active number of IoT devices	1×1
J	Number of time slots	1×1
$\mathbf{C}$	Codebook matrix	$N \times NK$
$\mathbf{c}$	Spreading sequence	$N \times 1$
$\mathbf{g}$	Channel	$N \times 1$
$\mathbf{v}$	Synthesis of transmit symbol and channel	$NK \times 1$
$\widetilde{\mathbf{v}}$	Stacked synthesis of transmit symbol and channel	$NJK \times 1$
$\mathbf{x}$	Sparse transmit signal	$NJK \times 1$
$\mathbf{y}$	Received signal	$NJ \times 1$
$\widetilde{\mathbf{y}}$	Stacked received signal	$NJ \times 1$
$\acute{\mathbf{y}}$	Transformed sparse signal	$K \times 1$
$\widehat{\boldsymbol{\chi}}$	Bits of reconstructed sparse signal	$K \times 1$
$\boldsymbol{\Gamma}$	Active device support set	$K \times 1$
$\widehat{\boldsymbol{\Upsilon}}$	Estimated active device support set	$K \times 1$
ξ	Rearranged codebook matrix	$NJ \times NJK$

scheme and describes the neural network's architecture. Section 2.4 discusses the network's training details and complexity analysis. Finally, in Section 2.5, we present the simulation results to verify the performance gain of the proposed technique. Table 2.1 summarizes the important symbols used in this work, including the dimensions of vectors and matrices.

2.2 System Model

We consider a spreading-based uplink grant-free NOMA system comprising of an AP and K IoT devices, as shown in Fig. 2.1. Without loss of generality, all devices and the AP are assumed to be equipped with a single antenna. We consider an overloaded system where the number of resource blocks N is less than the number of IoT devices, i.e., $N < K$. During transmission, a subset of the K devices sporadically and randomly become active when they have data to transmit. We adopt the burst-sparsity model in this work, i.e., some transmissions continue for several consecutive time slots while others last for one-time slot only [62–64].

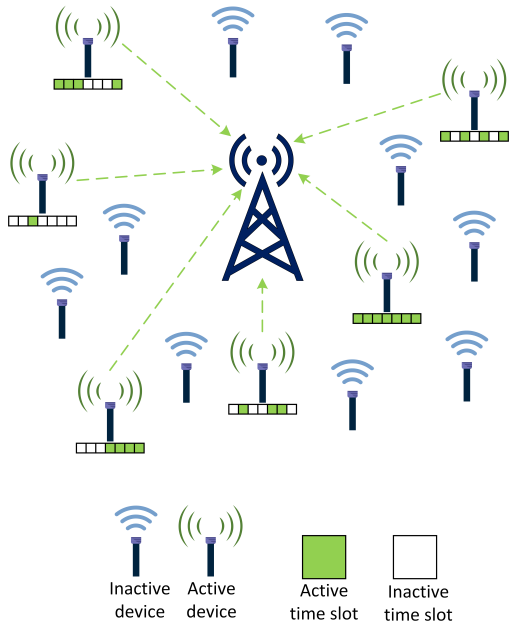


Figure 2.1: Illustration of a typical uplink grant-free NOMA system.

2.2.1 Signal Model

Considering an arbitrary symbol interval, an active device transmits its complex modulated signal towards the AP, which are independent random variables drawn from standard symmetric discrete constellation set M. For inactive devices, their transmit symbol is equal to zero. In this work, we consider that the device symbols are spread with a family of short complex-valued spreading sequences with low cross-correlation values [49]. These short complex-valued spreading sequences can be generated naturally based on the binary sequence elements. For instance, for $M = 3$, each element of the complex spreading sequence is taken from the set $\{-1, 0, 1, -1+i, i, 1+i, -1-i, -i, 1-i\}$ [49,67].

After modulation, the symbol s_k from the k^{th} device is spread onto a spreading sequence $\mathbf{c}_k = [c_{1k}, c_{2k}, \ldots, c_{Nk}]^T \in \mathbb{C}^{N\times 1}$ which is randomly and independently selected from a pre-defined set. The received signal $\mathbf{y}$ at the AP is the superposition of all signals, given as

$$\mathbf{y} = \sum_{k=1}^{K} \text{diag}(\mathbf{c}_k)\mathbf{g}_k s_k + \mathbf{w} \ = \mathbf{C}\mathbf{v} + \mathbf{w}, \tag{2.1}$$

where $\mathbf{g}_k = [g_{1k}, g_{2k}, \ldots, g_{Nk}]^T \in \mathbb{C}^{N\times 1}$ denotes the channel vector between the AP and the k^{th} device over N sub-carriers, and $\mathbf{w} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ represents the complex Gaussian noise vector. Moreover, $\mathbf{C} = [\text{diag}(\mathbf{c}_1), \text{diag}(\mathbf{c}_2), \ldots, \text{diag}(\mathbf{c}_K)] \in \mathbb{C}^{N\times NK}$ refers to the

codebook matrix of all devices, and $\mathbf{v} = [\mathbf{v}_1^T, \ldots, \mathbf{v}_K^T]^T = [(s_1\mathbf{g}_1)^T, \ldots, (s_K\mathbf{g}_K)^T]^T \in \mathbb{C}^{NK\times 1}$ is the synthesis of the transmit symbols and channel vectors.

2.2.2 Consecutive-Time Slot Dynamic Model

Exploiting the sparsity in the data transmission (i.e., only a subset of devices wake up to transmit) and the temporal correlation of the device activity pattern (i.e., data transmission is bursty in general), we can formulate the vector $\mathbf{v}$ as a sparse vector and extend our system model in (2.1) to a continuous-time slot model.

The idea is to utilise the bursty nature of $\widetilde{\mathbf{v}} = \left[\mathbf{v}^{[1]}, \mathbf{v}^{[2]}, \ldots, \mathbf{v}^{[J]}\right]^T \in \mathbb{C}^{NJK\times 1}$ where $\mathbf{v}^{[j]}$ is the signal at the j-th time slot, to retrieve it from the received signals $\widetilde{\mathbf{y}} = \left[\mathbf{y}^{[1]}, \mathbf{y}^{[2]}, \ldots, \mathbf{y}^{[J]}\right]^T \in \mathbb{C}^{NJ\times 1}$, in the J successive time slots. This formulation helps in capturing the temporal correlation of the active devices by detecting the transmit signals $\mathbf{v}$ in the continuous-time slots. The stacked received signal vector $\widetilde{\mathbf{y}}$ can be represented as

$$\widetilde{\mathbf{y}} = \begin{bmatrix} \mathbf{C}^{[1]} & 0 & \cdots & 0 \\ 0 & \mathbf{C}^{[2]} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{C}^{[J]} \end{bmatrix} \begin{bmatrix} \mathbf{v}^{[1]} \\ \mathbf{v}^{[2]} \\ \vdots \\ \mathbf{v}^{[J]} \end{bmatrix} + \begin{bmatrix} \mathbf{w}^{[1]} \\ \mathbf{w}^{[2]} \\ \vdots \\ \mathbf{w}^{[J]} \end{bmatrix}, \tag{2.2}$$

where $\mathbf{C}$ is the equivalent code-book matrix of all devices, which contains the complex spreading sequences of all K devices, $\mathbf{v}$ is the composite of the transmitted symbol and channel vector, and $\mathbf{w}$ is the Gaussian noise vector.

The AP receives a multi-device vector $\widetilde{\mathbf{y}}$ with no knowledge of the active transmitting devices or locations of the non-zero symbols. The active device support set $\mathbf{\Gamma}^{[j]}$ varies over different time slots considering the device's random transmission in a grant-free fashion. With this in mind, let $\mathbf{u}^{[j]} = [s_1, s_2, \ldots, s_K]^T$ correspond to the total devices in the j-th time slot. Then, the active device support set[1] $\mathbf{\Gamma}^{[j]}$ of the signal $\mathbf{x}^{[j]}$ in the j-th time slot is defined as [63]

$$\mathbf{\Gamma}^{[j]} = k \mid \mathbf{u}_k^{[j]} \neq 0, 1 \leq k \leq K. \tag{2.3}$$

From this, the number of transmitting active devices is defined through the cardinality

[1]The actual device support set $\mathbf{\Gamma}^{[j]}$ is utilised as ground truth to reduce the misclassification rate during network training only. In the online deployment, the proposed BiLSTM network is used to estimate the active device support set. This will be explained in the following sections.

of the active device support set $\mathbf{\Gamma}^{[j]}$, given as [63]

$$S^{[j]} = \left|\left|\mathbf{\Gamma}^{[j]}\right|\right|_0. \tag{2.4}$$

Since IoT traffic is not entirely random and often consists of data bursts and traffic patterns, in this work, we consider the burst-sparsity model where only a subset of active devices in the previous time slot also transmit in the next time slot. That is, only a subset of indices in $\mathbf{\Gamma}^{[j-1]}$ are present in $\mathbf{\Gamma}^{[j]}$. Therefore, to quantify the commonality of active devices transmitting in consecutive time slots, we define η as the level of temporal correlation between the previous time slot $\mathbf{\Gamma}^{[j-1]}$ and the current time slot $\mathbf{\Gamma}^{[j]}$. It is given as

$$\eta = \frac{\left|\left|\mathbf{\Gamma}^{[j-1]} \bigcap \mathbf{\Gamma}^{[j]}\right|\right|_0}{\left|\left|\mathbf{\Gamma}^{[j]}\right|\right|_0}. \tag{2.5}$$

Note that in (2.5), η characterises the overlapping level of the active devices transmitting in consecutive time slots. For instance, when $\eta = 0.5$, half of the devices transmit in consecutive time slots ≥ 2, whereas the remaining transmit only once during the whole process. In Section V, we will show how the variation of temporal correlation η affects the overall system performance.

2.2.3 Multi-User Detection Problem

When multiple active devices communicate with the AP simultaneously in a grant-free manner, the first task for the AP is to detect the active devices that contributed to the received signal. Therefore, the identification of active devices leads to the problem of finding the support of the transmitted signal.

In this regard, the rows in (2.2) can be rearranged. We also introduce an active device criterion $\delta \in 0, 1$, where $\delta = 1$ and $\delta = 0$ correspond to active and inactive devices, respectively [66]. Using this, the stacked received signal vector $\widetilde{\mathbf{y}}$ can be written as

$$\widetilde{\mathbf{y}} = \begin{bmatrix} \xi_1 & \cdots & \xi_K \end{bmatrix} \begin{bmatrix} \delta_1 \mathbf{x}_1 \\ \vdots \\ \delta_K \mathbf{x}_K \end{bmatrix} + \begin{bmatrix} \mathbf{w}^{[1]} \\ \vdots \\ \mathbf{w}^{[J]} \end{bmatrix} = \xi \mathbf{x} + \mathbf{w}, \tag{2.6}$$

where $\xi = [\xi_1, \xi_2, \ldots, \xi_K] \in \mathbb{C}^{NJ \times NJK}$, and $\mathbf{x} = [\delta_1 \mathbf{x}_1^T, \delta_2 \mathbf{x}_2^T, \ldots, \delta_K \mathbf{x}_K^T]^T \in \mathbb{C}^{NJK \times 1}$, such that for any k^{th} device, $\mathbf{x}_k = [(s_k^{[1]} \mathbf{g}_k^{[1]})^T, (s_k^{[2]} \mathbf{g}_k^{[2]})^T, \ldots, (s_k^{[J]} \mathbf{g}_k^{[J]})^T]^T$ and $\xi_k = [\text{diag}(\mathbf{c}_k^{[1]}), \text{diag}(\mathbf{c}_k^{[2]}), \ldots, \text{diag}(\mathbf{c}_k^{[J]})]$, respectively. From (2.6), it is inferred that out of K, only a subset of devices, say S, are active. This means that the sparse vector $\mathbf{x}$ has S

nonzero blocks corresponding to the S active devices. Therefore, $\widetilde{\mathbf{y}}$ in (2.6) can be represented as a linear combination of S submatrices of $\xi_1, \ldots, \xi_K$ perturbed by the noise [66]. We assume the codebook entries of ξ are available at the AP [45]. However, the AP does not know which spreading sequence is chosen by the different active devices. Thus, the AP needs to identify the sub-matrices ξ_S, which are analogous to ξ, by processing $\widetilde{\mathbf{y}}$.

From this, the MUD problem becomes a 2-dimensional CS problem, which is common in the CS paradigm [100]. With this in mind, the following MUD problem is readily articulated as an active device support estimation problem, given as

$$\boldsymbol{\Upsilon} = \underset{|\boldsymbol{\Upsilon}|=S}{\arg\min} \ \frac{1}{2} ||\widetilde{\mathbf{y}} - \xi_{\boldsymbol{\Upsilon}} \mathbf{x}_{\boldsymbol{\Upsilon}}||_2^2 . \tag{2.7}$$

This detection problem in (2.7) can be solved using classical CS approaches. The approaches based on exhaustive searches, such as ℓ_1-minimisation [101], provide theoretical performance gains but suffer from heavy computational complexity. The approaches based on greedy algorithms [102] have comparably lower complexity but result in a sub-optimal solution and require a larger number of measurements for signal recovery. The biggest drawback of conventional CS-based schemes is that they assume perfect knowledge of the channel and active device sparsity levels. Furthermore, the enormous computational complexity and the latency of iterative algorithms make them a practical solution only for a small number of active devices. When there is a larger number of active devices, the performance of conventional CS-based schemes degrades due to their sole dependence on the residual vector in each iteration[2]. Due to this, as the number of active devices increases, conventional CS-based schemes are not suitable solutions to facilitate grant-free communication. This motivates us to pursue a ML-aided solution presented in the next section.

2.3 Deep Learning aided MUD

To tackle the MUD problem in Section II-C, we propose a solution using deep learning. In essence, we aim to delineate a nonlinear mapping using deep learning to create a pattern between the stacked received signal $\widetilde{\mathbf{y}}$ and the support of $\mathbf{x}$ and perform MUD

[2]A nonzero submatrix of ξ with an index chosen at the i-th iteration is given as $\epsilon = \underset{k=1,\cdots,K}{\arg\max} \ \frac{1}{2} \left|\left|\xi_k^H \mathbf{r}^{i-1}\right|\right|_2^2$, where $\mathbf{r}^{i-1} = \mathbf{y} - \xi_{\Upsilon}^{i-1} \widehat{\mathbf{x}}^{i-1}$ is the i-th residual vector and $\widehat{\mathbf{x}}^{i-1} = \xi_{\Upsilon^{i-1}}^{\dagger} \mathbf{y}$ is an approximate of the transmitted signal $\mathbf{x}$ in the $(i-1)$-th iteration. It is of understanding that the performance of active user support identification is influenced primarily by ξ, which is generated through the codebook $\mathbf{C}$, and residual vector $\mathbf{r}^{(\cdot)}$.

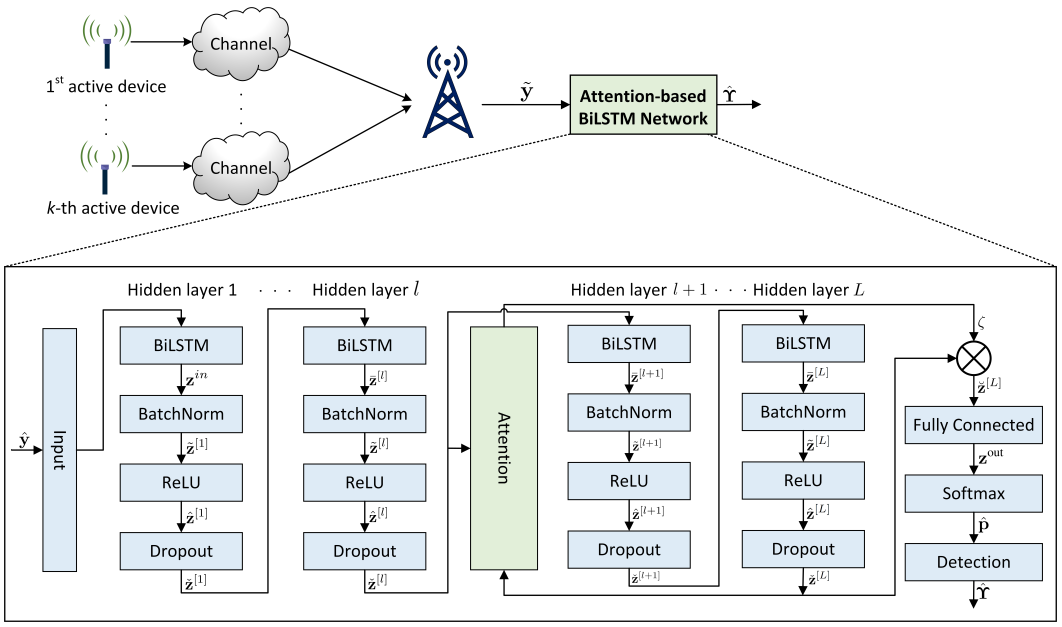


Figure 2.2: Detailed architecture and working of the proposed attention-based BiLSTM network.

at the AP. The resulting active device support estimation problem $\widehat{\boldsymbol{\Upsilon}}$ is then defined as

$$\widehat{\boldsymbol{\Upsilon}} = g(\widetilde{\mathbf{y}}, \Theta), \tag{2.8}$$

where Θ represents the weights and corresponding biases of the learning architecture.

2.3.1 Learning Architecture

In this work, we adopt an attention-based BiLSTM network to solve the MUD problem, as illustrated in Fig. 2.2. The attention mechanism is discussed in Section III-B, while in this section, we discuss the BiLSTM network. The motivation for adopting the BiLSTM network is as follows.

Standard unidirectional LSTM networks undertake sequences in forwarding temporal order, ignoring future context. This is because unidirectional LSTM only preserves the information of the previous time steps since it has exclusively seen inputs from the past. On the other hand, BiLSTM networks take unidirectional LSTM networks one step further by setting up a second LSTM layer, where the gradients in the hidden connections flow in the opposite temporal direction. That is, BiLSTM runs the inputs in two ways, one from past to future (left to right, i.e., forward) and another one from future to past

(right to left, i.e., backward). This gives BiLSTMs the ability to exploit more information, thereby simultaneously obtaining contextual features from forward and reverse temporal directions. In essence, more features from both directions are captured for mapping active devices transmitting in consecutive time slots. The LSTM in the reverse direction is calculated in the same fashion as the forward direction. Noticeably, since the direction is reversed, the time information is passed from future to past.

For input $\widetilde{\mathbf{y}}_t$ at the current time step t, the BiLSTM network calculation is given by

$$\overrightarrow{\mathbf{h}_f} = \sigma(\mathbf{W}_f \widetilde{\mathbf{y}}_t + \mathbf{W}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \tag{2.9}$$

$$\overleftarrow{\mathbf{h}_r} = \sigma(\mathbf{W}_r \widetilde{\mathbf{y}}_t + \mathbf{W}_r \mathbf{h}_{t+1} + \mathbf{b}_r), \tag{2.10}$$

where σ represents the activation function, $t-1$ and $t+1$ represent the forward and reverse direction time steps respectively, $\mathbf{h}_{t-1}$ and $\mathbf{h}_{t+1}$ represent the previous and next hidden states respectively, $\mathbf{W}_f$ and $\mathbf{W}_r$ represent the forward and reverse direction input weights respectively, and $\mathbf{b}_f$ and $\mathbf{b}_r$ represent the forward and reverse direction learnable bias parameter respectively. $\overrightarrow{\mathbf{h}_f}$ and $\overleftarrow{\mathbf{h}_r}$ represent the forward and reverse direction LSTM network outputs respectively. Finally, the output of the BiLSTM $\mathbf{z}_t$ is

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \overrightarrow{\mathbf{h}_f} \oplus \mathbf{W}_z \overleftarrow{\mathbf{h}_r} + \mathbf{b}_z) \ = \sigma(\mathbf{W}_z \widetilde{\mathbf{h}}_t + \mathbf{b}_z), \tag{2.11}$$

where $\mathbf{W}_z$ represents the BiLSTM output weights, $\mathbf{b}_z$ represents the BiLSTM output learnable bias parameter, and $\widetilde{\mathbf{h}}_t$ is the concatenated hidden state of the forward and reverse direction LSTMs.

Fig. 2.2 shows the proposed attention-based BiLSTM network applied to our MUD problem. For each training iteration, we use U training data copies $\widetilde{\mathbf{y}}^{(1)}, \cdots, \widetilde{\mathbf{y}}^{(U)}$. Next, since the stacked received signal $\widetilde{\mathbf{y}}^{(u)}$ is a complex-valued modulated vector, we split the real and imaginary parts and use $\widehat{\mathbf{y}}^{(u)} = [\Re(\widetilde{y}_1^{(u)}) \cdots \Re(\widetilde{y}_N^{(u)}), \Im(\widetilde{y}_1^{(u)}) \cdots \Im(\widetilde{y}_N^{(u)})]$ as an input vector to the network. With this in mind, the unit output in (2.11) is substituted as

$$\mathbf{z}_t^{\text{in},(u)} = \sigma(\mathbf{W}_z^{\text{in}} \widehat{\mathbf{y}}_t^{(u)} + \mathbf{b}_z^{\text{in}}), \quad \text{for } u = 1, \cdots, U, \tag{2.12}$$

where $\mathbf{W}_z^{\text{in}} \in \mathbb{R}^{\alpha \times 2NJ}$ is the initial weight, $\widehat{\mathbf{y}}_t^{(u)} \in \mathbb{R}^{2NJ \times 1}$ is the input vector, and $\mathbf{b}_z^{\text{in}} \in \mathbb{R}^{\alpha \times 1}$ is the input learnable bias term.

In this work, we employ batch normalisation to help coordinate the update of multiple layers by standardising the inputs of each layer to have fixed means and variances. This is important because when active devices experience different wireless channels and

transmit their data in a grant-free manner, the resulting stacked received signal $\widetilde{\mathbf{y}}$ has substantial variations. These significant variations make it difficult for the network to learn the device activation pattern. By standardising the inputs of each layer, batch normalisation reduces the variations and helps to overcome this difficulty. Thus, the output vectors U from (2.12) are put together in the mini-batch $\mathbf{B} = [\mathbf{z}_t^{(1)} \cdots \mathbf{z}_t^{(U)}]^T$. Once arranged in a mini-batch, these vectors are scaled and shifted using their respective hidden weights and batch normalised. The output for each element $z_{t,i}^{\text{in},(u)}$ of the batch normalisation (BatchNorm) is given as

$$\widetilde{z}_{t,i}^{(u)} = \beta \left(\frac{z_{t,i}^{\text{in},(u)} - \mu_{\mathbf{B},t,i}}{\sqrt{\sigma_{\mathbf{B},t,i}^2}} \right) + \gamma, \quad \text{for } i = 1, \cdots, \alpha, \tag{2.13}$$

where $\mu_{\mathbf{B},t,i} = \frac{1}{U} \sum_{u=1}^{U} z_{t,i}^{\text{in},(u)}$ calculates the batch-wise mean, $\sigma_{\mathbf{B},t,i}^2 = \frac{1}{U} \sum_{u=1}^{U} (z_{t,i}^{\text{in},(u)} - \mu_{\mathbf{B},t,i})^2$ calculates the batch-wise variance, β is used as a scaling parameter, γ is used as the shifting parameter, and α represents the width of the hidden layers.

The proposed scheme learns to create a mapping function between the stacked received signal $\widetilde{\mathbf{y}}$ and the current active device support set $\mathbf{\Gamma}$. However, the estimate of the current active device support set $\widehat{\mathbf{\Upsilon}}$ is vastly agitated by the activation patterns of the neurons, which in turn are dependent on perturbations and precision errors. This issue is further compounded as the spreading sequences in the sensing codebook matrix ξ are correlated. Accordingly, the estimate of the current active device support set $\widehat{\mathbf{\Upsilon}}$ might not be accurate and will misclassify in the presence of random perturbations. In addition, when the device activity pattern is similar in consecutive time slots, the network is more prone to overfitting due to the unchanging device activation pattern. We use the ReLU activation function and dropout layer to address these issues. By using the ReLU activation function, the computed weights at every iteration are ranged, i.e., $f(x) = \max(x, 0)$, which is then used as

$$\widehat{\mathbf{z}}_t^{(u)} = f(\widetilde{\mathbf{z}}_t^{(u)}). \tag{2.14}$$

In the dropout layer, the activated neurons in a hidden layer are randomly halted with a probability ρ_{drop}, given as

$$\check{\mathbf{z}}_t^{(u)} = \widehat{\mathbf{z}}_t^{(u)} \odot \mathbf{d}^{(u)}, \quad d_i^{(u)} \sim Bern(\rho_{\text{drop}}), \tag{2.15}$$

where $d_i^{(u)}$ is the i-th element of the dropout vector $\mathbf{d}^{(u)}$, and $\odot$ is the Hadamard prod-

uct. $Bern(\rho_{\text{drop}})$ is the Bernoulli random variable which takes the value 0 with the dropout probability ρ_{drop} and 1 with the probability $1 - \rho_{\text{drop}}$. The dropout mechanism deliberately makes the training process noisy by deactivating neurons randomly, forcing the remaining neurons to take more responsibility in creating a different path for the gradient flow. This random dilution of neurons provides rigorous circumstances where network layers co-adapt to rectify mistakes from prior layers, which helps create a more generalised network capable of estimating the current active device support set with more accuracy. Therefore, removing incoming and outgoing connections of the dropped neurons with a random probability systematically resolves the active device activation patterns' similarity among correlated support sets.

After the dropout layer, the output vector $\check{\mathbf{z}}$ makes its way through multiple hidden layers[3]. In subsequent, every hidden layer comprises the BiLSTM layer, a BatchNorm layer to reduce the variation of $\overline{\mathbf{z}}^{[l]}$, a ReLU activation function applied to $\widetilde{\mathbf{z}}^{(u)}$ to determine whether the information $(\widehat{z}_1^{[l]}, \ldots, \widehat{z}_\alpha^{[l]})$ generated by the hidden unit is activated or not, and finally a dropout layer to overcome overfitting of the network is applied (see Fig. 2.2). The output of the l^{th} hidden layer's BiLSTM $\overline{\mathbf{z}}_t^{[l]}$ is

$$\overline{\mathbf{z}}_t^{[l]} = \mathbf{W}^{[l]} \left(\sum_{i=1}^{l-1} \check{\mathbf{z}}_t^{[i]} \right) + \mathbf{b}^{[l]}, \tag{2.16}$$

where $\mathbf{W}^{[l]} \in \mathbb{R}^{\alpha \times \alpha}$ and $\mathbf{b}^{[l]} \in \mathbb{R}^{\alpha \times 1}$ are the weight and bias in the l^{th} hidden layer, respectively.

2.3.2 Attention Mechanism

Fig. 2.3 shows the working of the attention mechanism and its integration with the BiLSTM network architecture. The motivation for adopting the attention mechanism is twofold: (i) a neural network that creates a mapping function for the active device detection problem in (2.7) by analyzing the whole input at every step ignores the temporal correlation of the device activity pattern and (ii) with the increasing number of active devices, it becomes difficult for a neural network to learn its activation pattern due to its inherent sequential path architecture, causing problems such as vanishing and exploding gradients [103].

The attention mechanism allows the neural network to apply context to specific parts of the data at every time step. That is, instead of finding the active devices in all of the input vectors altogether, a neural network with an attention mechanism breaks down the

[3]For notational simplicity, in the proceeding sections, the training data index u has been omitted.

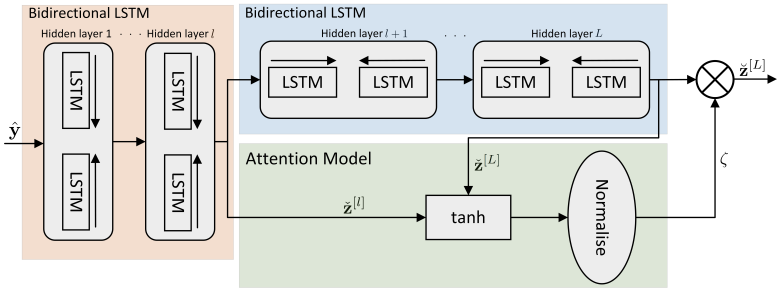


Figure 2.3: The proposed BiLSTM module with an attention mechanism.

data, applies a contextual vector to it, and then gives a score to the parts where active devices are present and transmitting consecutively. This mechanism brings additional temporal-based reasoning into the overall architecture for active device detection, helping the neural network load more active devices for detection.

The output of the BiLSTM network is computed as a weighted summation of the output of the BiLSTM network $\check{\mathbf{z}}_t^{(L)}$ at time step t as

$$\breve{\mathbf{z}}_t^{[L]} = \sum_{k=1}^{K+1} \zeta_k \check{\mathbf{z}}_{t-(k-1)}^{[L]}, \tag{2.17}$$

where ζ_k is the temporal attention value at time step $t-(k-1)$, computed as

$$\zeta_k = \frac{e^{s_k}}{\sum_{k=1}^{K} e^{s_k}}, \tag{2.18}$$

where the scores $\mathbf{s} = [s_1 \cdots s_K]^T$ indicate the repeated activation pattern of active devices in the time slots, which is obtained as

$$s_k = W_{\text{rel}} \tanh(W_a \check{\mathbf{z}}_t^{[l]} + Z_a \check{\mathbf{z}}_t^{[L]} + b_a). \tag{2.19}$$

where $\check{\mathbf{z}}^{[l]}$ is the output of the previous hidden layers, and W_a and Z_a are the attention learnable parameters that learn to project each context element and hidden state into a latent space and W_{rel} denotes the relevance parameter [104].

Evidently from (2.18) and (2.19), at time step t, ζ depends on the input $\check{\mathbf{z}}_t^{(l)}$. Furthermore, ζ is also dependent on the hidden variables $\check{\mathbf{z}}_t^{[L]}$ in the previous and current time step t. The attention value ζ can also be regarded as activating the active device detection gate. That is, the amount of information flow into the BiLSTM network is controlled by setting the gates. With this in mind, the final prediction result is influenced

by a larger activation value, which results in a larger flow of information. It should also be noted that the standard LSTM network cannot detect many active devices concerning the previous activation pattern due to the large memory overhead occurring. The BiLSTM network with an attention-based mechanism can capture device activation patterns in consecutive time slots with long-range dependencies. The information not required can be suppressed to improve the accuracy and efficiency of active device detection.

After passing through the L BiLSTM layers and the attention-based mechanism, the FC layer at the output produces K values corresponding to the total number of devices. Thereby, the output vector $\mathbf{z}^{\text{out}}$ is produced as

$$\mathbf{z}^{\text{out}} = \mathbf{W}^{\text{out}} \sum_{l=1}^{L} \breve{\mathbf{z}}^{[l]} + \mathbf{b}^{\text{out}}, \tag{2.20}$$

where $\mathbf{W}^{\text{out}}$ is the corresponding weight and $\mathbf{b}^{\text{out}}$ the bias, respectively. The softmax layer then maps K output values into K probabilities $(\widehat{p}_1, \cdots, \widehat{p}_K)$ representing the likelihood of being the true support element in the estimated active device support set $\widehat{\boldsymbol{\Upsilon}}$. The k^{th} probability $\widehat{p}_i$ calculated through softmax is given as

$$\widehat{p}_k = \frac{e^{z_k^{\text{out}}}}{\sum_{k=1}^{K} e^{z_k^{\text{out}}}}. \tag{2.21}$$

Finally, an estimate of the active device support set $\widehat{\boldsymbol{\Upsilon}}$ is obtained by picking from the K elements those having a probability greater than the threshold τ, given as

$$\widehat{\boldsymbol{\Upsilon}} = \begin{cases} 1 & \widehat{p}_k \geq \tau \\ 0 & \text{otherwise} \end{cases}. \tag{2.22}$$

Once $\widehat{\boldsymbol{\Upsilon}}$ in (2.22) is obtained, the estimated support $\widehat{S}$ is then extracted through the cardinality of the estimated active device support set, i.e., $\widehat{S} = ||\widehat{\boldsymbol{\Upsilon}}||_0$ for the j-th time slot. We later show how the estimated active device support set and estimated support are used to evaluate the MUD performance and device identification accuracy.

2.4 Model Training, User Detection, and Complexity Analysis

In this section, we discuss the model training, which in turn is used for signal reconstruction, and find the computational complexity of the proposed attention-based BiLSTM

network.

2.4.1 Model Training

During the offline training phase, the network's parameters set Θ^* are computed by minimising the loss function $\mathcal{J}(\Theta)$ (i.e., $\Theta^* = \arg\min_{\Theta} \mathcal{J}(\Theta)$). During every training iteration, the network parameters are updated using the gradient descent method when the loss function $\mathcal{J}(\Theta)$ is differentiable. Specifically, using the Adam optimiser, the network parameters Θ_i are updated in the direction of the steepest descent in the i-th training iteration, given as

$$\Theta_i = \Theta_{i+1} - \frac{\psi m_i}{\sqrt{v_i + \epsilon}}, \tag{2.23}$$

where ψ is the learning rate determining the step size, and ϵ is a smoothing term that prevents division by zero. Furthermore, m_i and v_i are estimates of the mean and uncentered variance of the gradients, respectively, defined as [9]

$$\begin{aligned} m_i &= \delta_1 m_{i-1} + (1-\delta_1)\nabla\mathcal{J}(\Theta_i) \\ v_i &= \delta_2 v_{i-1} + (1-\delta_2)\nabla[\mathcal{J}(\Theta_i)]^2, \end{aligned} \tag{2.24}$$

where δ_1 and δ_2 are the decay rates of the moving average. The moving average parameters help in controlling the step size of the optimiser in order to identify the global optimum solution of the training set correctly and prevents the network from looping in a local solution when the training data is not sparse [9].

Recalling that the final output of the attention-based BiLSTM network is the K-dimensional vector $\widehat{\mathbf{p}}$ whose element represents the probability of being the estimated support element from the estimated active device support set $\widehat{\boldsymbol{\Upsilon}}$. In this regard, $\widehat{\mathbf{p}} = [\widehat{p}_1 \cdots \widehat{p}_K]$ needs to be compared against the true probability $\mathbf{p}$ in the loss function calculation. We employ the cross entropy loss $\mathcal{J}(\mathbf{p}, \widehat{\mathbf{p}}, \Theta)$ for network training, defined as [103]

$$\mathcal{J}(\mathbf{p}, \widehat{\mathbf{p}}, \Theta) = -\frac{1}{K}\left(\sum_{k=1}^{K} p_k \log \widehat{p}_k\right) + \lambda \sum_{k=1}^{K} \Theta_k^2, \tag{2.25}$$

where p_k is the ground truth (actual active device), $\widehat{p}_k$ is the estimate (estimated active device) of the attention-based BiLSTM network, and λ is the L_2 regularisation term which is used for weight decaying and in turn, improves the generalisation performance of the network.

2.4.2 Blind Data Detection of Active Devices

In grant-free transmission, the codebook or the spreading sequence is unknown before the signal is detected. This can significantly increase the detection complexity. However, finding the active devices from the received signal can also recover their adopted spreading sequences since a local copy of the spreading sequences is available at the AP. Thus the detection and decoding computational complexity can be reduced significantly while keeping the practical constraints of grant-free NOMA systems intact.

First, AUD is carried out as in (2.22) where the estimated active device support set $\widehat{\boldsymbol{\Upsilon}}$, and the estimated sparsity level $\widehat{S}$ is obtained from the attention-based BiLSTM network using the stacked received signal $\widetilde{\mathbf{y}}$ at the AP. Next, using this estimated active device support set $\widehat{\boldsymbol{\Upsilon}}$, the stacked received signal $\widetilde{\mathbf{y}}$ is transformed into a sparse signal $\acute{\mathbf{y}}$, which contains received data for the estimated active devices. Having knowledge of the estimated active devices and their received bits, the spreading sequences employed by these active devices are obtained by selecting the estimated $\widehat{S}$ spreading sequences having the highest correlation probability with the spreading sequences at the AP[4] [107].

Once the spreading sequences employed by the active devices are calculated, blind detection can be carried out. In blind detection, the active device channels are unknown, while the spreading sequences are known. Therefore, based on the sparse signal $\acute{\mathbf{y}}$, which includes the statistical information of channels of all active devices, the blind MMSE weight $\mathbf{w}$ can be obtained without the knowledge of device channels. Thereby, the MMSE weight can be calculated as [99]

$$\mathbf{w}^T = (\widehat{\mathbf{g}}^T\widehat{\mathbf{g}} + \sigma^2 I)^{-1}\widehat{\mathbf{g}}^T, \tag{2.26}$$

where $\widehat{\mathbf{g}}$ is the estimated channel between the AP and the devices. After rearranging (2.6), the transmitted bits of the reconstructed sparse signal $\widehat{\boldsymbol{\chi}}$ for the active devices can then be estimated as

$$\widehat{\boldsymbol{\chi}} = \mathbf{w}^T\frac{\acute{\mathbf{y}}}{\xi}. \tag{2.27}$$

By doing so, the active devices' bits are estimated, and the sparse signal $\acute{\mathbf{y}}$ is reconstructed without the need for explicit channel estimation. The entire process is summarised in Algorithm 1.

[4]Due to the spreading sequences being randomly selected from a pool, there is a possibility of two or more active devices selecting the same spreading sequence, causing spreading sequence collision [105]. However, since we have employed complex spreading sequences, and the use of channels of different devices are different, such spreading sequence collision does not have a large impact on the performance of blind detection [106].

Algorithm 1 The Proposed Attention-based BiLSTM Network.

Input Received signal $\widehat{\mathbf{y}}$

Output Estimated active user support set $\widehat{\boldsymbol{\Upsilon}}$, estimated sparsity level $\widehat{S}$, bits of reconstructed sparse signal $\widehat{\boldsymbol{\chi}}$

Initialisation $\acute{\mathbf{y}} = 1, \cdots, K$

Active device support and sparsity estimation

for $j = 1$ to J **do**
  Obtain $\widehat{\mathbf{p}}$ by passing $\widehat{\mathbf{y}}^{[j]}$ into the attention-based BiLSTM network
  $\widehat{\boldsymbol{\Upsilon}}^{[j]} = k \mid \widehat{p}_k \geq \tau, 1 \leq k \leq K$
  $\widehat{S}^{[j]} = \left|\left|\widehat{\boldsymbol{\Upsilon}}^{[j]}\right|\right|_0$

Sparse signal reconstruction

  for $k = 1$ to K **do**
    if $\widehat{\mathbf{p}}_k \geq \tau$ **then**
      $\acute{\mathbf{y}}_k^{[j]} = \widehat{\mathbf{y}}_k^{[j]}$
    else
      $\acute{\mathbf{y}}_k^{[j]} = 0$

Blind data detection of active devices

  $\mathbf{w}^T = (\widehat{\mathbf{g}}^T\widehat{\mathbf{g}} + \sigma^2 I)^{-1}\widehat{\mathbf{g}}^T$
  $\widehat{\boldsymbol{\chi}}^{[j]} = \mathbf{w}^T \frac{\acute{\mathbf{y}}^{[j]}}{\xi}$ mapped to the nearest symbol

Return $\widehat{\boldsymbol{\Upsilon}}$, $\widehat{S}$, $\widehat{\boldsymbol{\chi}}$

2.4.3 Computational Complexity

In this subsection, we evaluate the computational complexity of the proposed attention-based BiLSTM network. We evaluate the complexity using the floating-point operations per second (flops) [66], taking into account the complexity of the hidden and deep learning layers of the proposed BiLSTM network at the j-th time slot.

In the first layer of the attention-based BiLSTM network, the input vector has a dimension of $\widehat{\mathbf{y}} \in \mathbb{R}^{2N\times 1}$, whereas the weight and bias have the dimensions $\mathbf{W}^{\text{in}} \in \mathbb{R}^{\alpha\times 2N}$ and $\mathbf{b}^{\text{in}} \in \mathbb{R}^{\alpha\times 1}$ respectively. Furthermore, we know that BiLSTM has four gates, which do a forward pass and a backward pass, thereby bringing the generic flop computation per BiLSTM block to

$$\mathcal{C}_{\text{in}} = 8 \times (4N-1)\alpha + \alpha = 32N\alpha - 7\alpha. \tag{2.28}$$

Next, in the BatchNorm, the element-wise scalar multiplication and addition are carried out twice for normalisation. Thereby, the complexity $\mathcal{C}_{\text{bn}}$ of BatchNorm is given

Table 2.2: Computational complexity comparison for different sparsity levels (the total number of potential devices $K = 200$, the number of subcarriers $N = 100$, the number of hidden layers $L = 3$, width of hidden layer $\alpha = 1000$).

Technique	Floating point operations (flops)	Complexity for different sparsity levels			
		$S = 10$	$S = 20$	$S = 30$	$S = 40$
LS-OMP	$2SN^2K + \frac{S^4+6S^3+7S^2+2S}{12}N^3 + S(S+1)N^2 - S$	1.41×10^9	1.76×10^{10}	8.15×10^{10}	2.46×10^{11}
D-AUD	$2L\alpha^2 + (4N + 7L + 2N + 4)\alpha + (S+3)K - \frac{S(S+1)}{2} - 1$ $+2N + S(\frac{14}{3}N^3 + N^2 - N)$	5.33×10^7	1.00×10^8	1.46×10^8	1.93×10^8
LSTM-CS	$2\alpha^2(4L+3) + 2\alpha(8N + K) + \alpha(3L - 1) + 3K - 1$ $+2N + S(\frac{14}{3}N^3 + N^2 - N)$	7.87×10^7	1.25×10^8	1.72×10^8	2.19×10^8
Proposed	$2\alpha^2(8L+3) + 2\alpha(16N + K) - \alpha(L+3) + 3K - 1$ $+2N + S(\frac{14}{3}N^3 + N^2 - N)$	1.04×10^8	1.51×10^8	1.97×10^8	2.46×10^8

as

$$\mathcal{C}_{\text{bn}} = 4\alpha. \tag{2.29}$$

Subsequently, in the proceeding hidden layers' BiLSTMs, the hidden weight $\mathbf{W} \in \mathbb{R}^{\alpha \times \alpha}$ is multiplied with the input vector and then the bias term $\mathbf{b} \in \mathbb{R}^{\alpha \times 1}$ is added to it. Next, after passing through the subsequent BatchNorm for each element, the weights are passed through the ReLU activation function. Next, for generalisation, the dropout vector $\mathbf{d}$ is multiplied by the ReLU output $\widehat{\mathbf{z}}$. Consequently, the complexity of the L hidden layers $\mathcal{C}_{\text{hide}}$ is given as

$$\mathcal{C}_{\text{hide}} = L(8 \times (2\alpha - 1)\alpha + \alpha + 4\alpha + \alpha + \alpha) \ = 16L\alpha^2 - L\alpha. \tag{2.30}$$

Following this, the Attention layer performs weighted matrix multiplications with the input and previously sampled data, adds a bias term to the latent data, and multiplies the learnable parameters matrix to compute the scores. Thus, the complexity $\mathcal{C}_{\text{atten}}$ of the Attention layer is

$$\mathcal{C}_{\text{atten}} = \alpha(2\alpha - 1) + 4\alpha^2 + \alpha = 6\alpha^2 \tag{2.31}$$

Next, the FC layer at its output has its weights $\mathbf{W}^{\text{out}} \in \mathbb{R}^{K \times \alpha}$ and bias term $\mathbf{b} \in \mathbb{R}^{K \times 1}$ multiplied with the weights from the L hidden layers and the Attention mechanism. Thereby, the FC layer at the output has a complexity $\mathcal{C}_{\text{out}}$ given as

$$\mathcal{C}_{\text{out}} = (2\alpha - 1)K + K = 2K\alpha. \tag{2.32}$$

The softmax layer computes the K probabilities of potential devices, as in (2.21). By doing so, the softmax complexity $\mathcal{C}_{\text{sm}}$ is given as

$$\mathcal{C}_{\text{sm}} = 3K - 1. \tag{2.33}$$

From (2.28) to (2.33), the final complexity of the proposed attention-based BiLSTM network is

$$\begin{aligned}\mathcal{C}_{\text{ABA}} =& \mathcal{C}_{\text{in}} + \mathcal{C}_{\text{bn}} + \mathcal{C}_{\text{hide}} + \mathcal{C}_{\text{atten}} + \mathcal{C}_{\text{out}} + \mathcal{C}_{\text{sm}} \\ =& 2\alpha^2(8L+3) + 2\alpha(16N+K) - \alpha(L+3) + 3K - 1.\end{aligned} \tag{2.34}$$

For an unbiased analysis, we compare the proposed attention-based BiLSTM network with two deep learning-based techniques, D-AUD [66] and LSTM-CS [108], and a conventional technique, least squares orthogonal matching pursuit (LS-OMP) [109] for complexity comparison in Table 2.2. In addition, for a fair comparison, the MMSE estimation term has been added to D-AUD and LSTM-CS techniques for signal detection purposes, such that $\mathcal{C}_{MMSE} = 2N + S(\frac{14}{3}N^3 + N^2 - N)$. We examine the computational complexity in flops for different sparsity levels. We observe that the complexity of the proposed attention-based BiLSTM network is slighter higher than D-AUD and LSTM-CS but much lower than that of conventional approaches. This is because the D-AUD technique utilises vanilla FC layers for its network, which do not exploit the temporal correlation of data. Due to this reason, the performance of such networks might degrade with a higher number of active devices. The LSTM-CS uses unidirectional LSTM and therefore has lower computational complexity than the proposed attention-based BiLSTM network. However, as shown in Section V, this results in performance degradation. It is important to note that the complexity of ML-based techniques depends heavily on the network parameters (L and α), but not the system parameters (the number of active devices S, and the total number of devices K). Thus, when S increases from 10 to 20, the computational complexity of ML networks increases marginally, but that of LS-OMP increases sharply. Therefore, in a practical grant-free NOMA setting with a higher number of active devices, the ML schemes are more competitive in computational complexity than conventional schemes.

2.4.4 Convergence

We examine the validation loss $\mathcal{J}_v(\Theta)$ for a different number of hidden layers L for the proposed network, as shown in Fig. 2.4. We can see that a lower L results in a network being unstable during training, whereas a higher L results in a more stable network but with a slower convergence rate. Thus, we adopt $L = 3$ for training dataset generation and also the simulations in this work. Note that the sudden increase in validation loss for the $L = 1$ curve is due to the model overfitting to the training data, causing it to fit noise and outliers and perform poorly on the validation set.

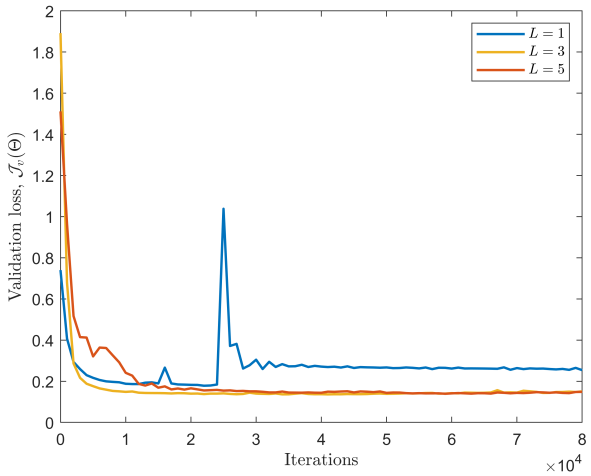


Figure 2.4: Validation loss $\mathcal{J}_v(\Theta)$ for different number of hidden layers L, with total number of devices $K = 200$, number of subcarriers $N = 100$, and number of active devices $S = 20$.

2.4.5 Training Dataset Generation

In order to determine the optimal network mapping function g^* for the stacked received signal $\widetilde{\mathbf{y}}$ and support of the $\mathbf{x}$, a comprehensive training dataset is required. A good option in this regard is acquiring a dataset produced using real received signals; however, there is no open-source dataset for the grant-free NOMA scenario at this stage.

In this work, the training set is generated artificially by sampling values from (2.6) while keeping the realistic system constraints intact. In essence, during the offline training stage, U training data copies of the stacked received signal vector $\widetilde{\mathbf{y}}$ and the support of $\mathbf{x}$ are used as the dataset for network training. With sufficient training data copies, a balanced dataset is obtained, which captures the practical transmission nature of IoT devices. The balance is achieved by ensuring equal representation of active and inactive device scenarios, varying SNRs, and diverse transmission patterns. These factors are carefully considered to mirror the real-world operational characteristics of IoT devices. In this work, the dataset consists of 132,000 artificially generated training samples and 33,000 validation samples.

The parameter values used for training dataset generation are summarised in Table 2.3. In particular, in the training dataset, the number of active devices is $S = 20$ among the $K = 200$ potential devices, and the number of subcarriers is $N = 100$. The temporal transmission nature of devices is captured based on (2.5), and η is set to 0.5.

Table 2.3: Parameter values used in generating the training dataset.

Parameter	Value
Modulation	QPSK
Total devices, K	200
Total subcarriers, N	100
Active devices, S	20
Channel, g_k	Rayleigh fading, $\mathcal{CN}(0,1)$
Multiple access signature	Random selection
Time slots, J	7
Temporal correlation, η	0.5
Signal-to-Noise Ratio (SNR) distribution	0 dB to 20 dB
Detection threshold, τ	0.5
Hidden layers, L	3
Hidden layer width, α	1000
Activation layer, σ	ReLU
Learning rate, ψ	0.001
Batch size, B	20
Dropout probability, ρ_{drop}	0.3
Validation split	20%
Moving average decay rate, δ_1, δ_2	$\delta_1 = 0.9$, $\delta_2 = 0.99$

The length of the time frame is $J = 7$. The number of hidden BiLSTM layers is set as $L = 3$, each with a width of $\alpha = 1000$, each followed by a ReLU activation function. The attention mechanism is placed before the final hidden layer. The output layer is preceded by an FC layer whose width corresponds to the number of classes. The dropout probability for the dropout layer is set to $\rho_{\text{drop}} = 0.3$. The batch size is set as 20, while Adam is the optimiser. The value for the latent learning rate ψ is set to 0.001.

2.5 Results and Discussion

In this section, we evaluate the performance of the proposed attention-based BiLSTM network in solving the MUD problem. We also plot the performance of four benchmark solutions: two traditional CS solutions, LS-OMP [109] and dynamic CS-based MUD method [62], one ML-based LSTM-CS MUD method [108] and the Oracle least squares (LS) algorithm.

The motivation for considering these four benchmarks is as follows. We consider the LS-OMP as it is the standard CS technique that is always considered as one of the benchmarks in this research field. The dynamic CS-based and ML-based LSTM-CS methods

are considered because they take temporal correlation into account during MUD. Additionally, the ML-based LSTM-CS method demonstrates the advantage gained from considering the proposed BiLSTM over vanilla LSTM. The Oracle LS algorithm is considered as it provides the theoretical performance lower bound, although it is impractical in real-world situations where perfect knowledge is unavailable.

For a fair comparison, we make the following assumptions in the implementations of the four benchmark schemes:

- For the two traditional benchmark solutions, the sparsity level is assumed to be known at the AP due to the assumption of the channels being perfectly known; only the sparse support location is unknown at the AP.
- For the ML-based LSTM-CS MUD method, the core working of the method is adopted from [108], but the LSTM layer is adapted to our architecture (as in Fig. (2)) for a fair comparison.
- For the ML-based LSTM-CS MUD method, we assume that it does not need any channel state information (CSI), i.e., it is unaware of the sparsity level, sparse support location and the channels.
- For the Oracle LS algorithm, we assume perfect knowledge of the CSI, user sparsity level and sparse support location.

In the simulations, unless otherwise stated, $K = 200$ potential devices simultaneously share $N = 100$ orthogonal resource blocks. Thus, the OF[5] is 200%. The number of active devices is in the range $S = 10 - 40$. We employ $M = 4$-ary complex spreading sequences, where both the real and imaginary parts take values from the set $\{-2, -1, 0, 1\}$. For every time slot, there are S number of active devices, where the active device support set $\mathbf{\Gamma}^{[j]}$ in each time slot has $S/2$ devices transmitting in the next time slot, i.e., $\eta = 0.5$, while the remaining are randomly selected from $\{1, 2, \cdots, K\}$. The number of time slots is fixed at $J = 7$ to conform to the LTE-Advanced protocol [110]. The signals being transmitted are modulated by Quadrature Phase Shift Keying (QPSK). Furthermore, all channels are assumed to follow an independent Rayleigh fading, and the channel fading coefficient is generated following $g_{n,k} \sim \mathcal{CN}(0, 1)$ [111–113]. The path loss between the AP and the k-th device is modeled as $128.1 + 37.6 \log_{10}(d_i)$, where d_i is the distance (in km) [110]. The results are averaged over 1000 Monte Carlo trials.

[5]The OF is defined as the ratio of the number of potential devices to the number of available resource blocks in the system, i.e., OF (%) $= \frac{K}{N} \times 100$.

The simulations are carried out on the Gadi supercomputer of the National Computational Infrastructure (NCI), Australia, utilising 48 cores of Intel Xeon Platinum 8274 (Cascade Lake) processors, 192GB of random access memory and NVIDIA V100 GPU. The simulations are carried out on MATLAB 2021b.

2.5.1 Performance Metrics

In order to appropriately evaluate the performance, including the quality of support estimation, device identification, and multi-device data detection, we use the following metrics: the detection probability (ρ_d), the accuracy, and the average BER as performance metrics. Given the bits of the reconstructed sparse signal $\widehat{\boldsymbol{\chi}}_k^{[j]}$ for device k at the j-th time slot, the performance metrics are defined as follows.

- Detection probability: This metric evaluates the performance of support estimation. It is defined as the ratio of the number of detected active devices to the number of *all* active devices, given as

$$\rho_d = \frac{1}{S} \sum_{k \in \Gamma^{[j]}} \mathbb{1}_{\widehat{\boldsymbol{\chi}}_k^{[j]} \neq 0}. \tag{2.35}$$

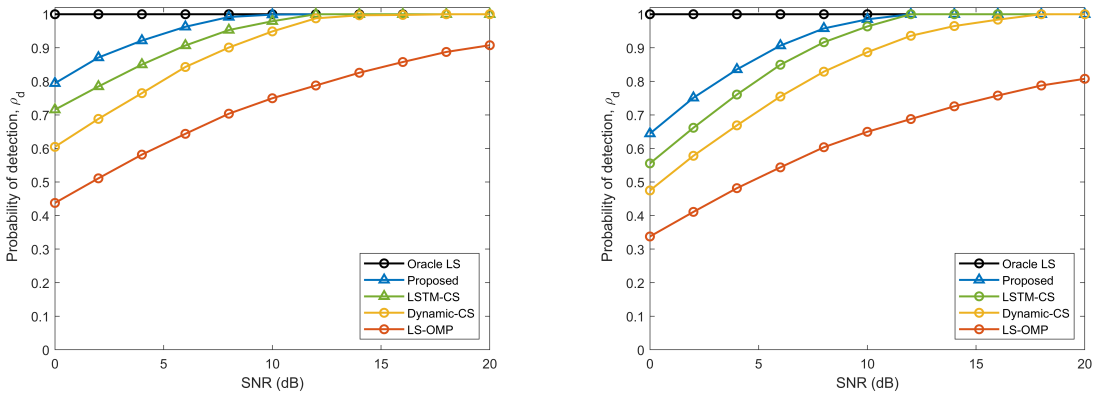
- Accuracy: This metric evaluates the performance of the quality of support estimation for device identification. It is defined as the ratio of the number of correctly identified active devices to the number of *all* active devices, expressed as a % and given as

$$\text{Accuracy } (\%) = \frac{1}{S} \sum_{k \in \Gamma^{[j]}} \mathbb{1}_{\boldsymbol{\Gamma}^{[j]} == \widehat{\boldsymbol{\Upsilon}}^{[j]}} \times 100. \tag{2.36}$$

- Average BER: This metric evaluates the performance of multi-device data detection. It is defined as the ratio of incorrectly recovered bits transmitted by the active devices to all bits transmitted by the active devices. It should be noted that the average BER includes a penalty for decoding the wrongly detected active devices.

2.5.2 Support Estimation

Fig. 2.5 plots the detection performance, ρ_d, versus the SNR (dB) for $S = 10$ and $S = 20$, with $K = 200$, and $N = 100$. The following trends can be observed from the figure. The Oracle LS gives the theoretical best performance (100% detection probability for the considered scenario), which is the same for all SNR values. As the SNR



(a) $S = 10$ (b) $S = 20$

Figure 2.5: Probability of detection, ρ_d, versus SNR (dB) for the number of active devices S, with the total number of potential devices $K = 200$, and the number of subcarriers $N = 100$.

increases, the performance of all the schemes slowly approaches to that of the Oracle LS. The LS-OMP performs the worst since it ignores the temporal correlation in the device activation history. The dynamic CS-based MUD method performs better than LS-OMP since it considers the temporal correlation in the device activation history. The ML-based LSTM-CS method performs better than the two traditional algorithms but cannot perform similarly to the proposed BiLSTM network due to its unidirectional architecture. The proposed attention-based BiLSTM network outperforms all these benchmark algorithms, i.e., it exhibits a higher detection probability of successfully identifying the correct number of active devices against all other schemes. For instance, the proposed attention-based BiLSTM network achieves the Oracle LS detection performance at SNR = 8 dB and SNR = 12 dB, respectively, for $S = 10$ and $S = 20$ active devices. It should be noted that the proposed attention-based BiLSTM network is unaware of the device sparsity level and detects the active devices based on the received signal only, compared to other traditional algorithms, which are based on the assumption of the known channels and device sparsity level. As the number of active devices S increases from 10 to 20, the detection performance of the proposed attention-based BiLSTM network decreases gradually. The decrease in performance is attributed to the introduction of additional interference, variability, and overlapping patterns. These complexities pose challenges for the model to effectively capture and learn the underlying patterns and relationships

Table 2.4: Device identification accuracy versus the number of active devices S, with the total number of potential devices $K = 200$, the number of subcarriers $N = 100$, and SNR = 6 dB.

# of Active Devices	Accuracy (%)			
	LS-OMP	Dynamic-CS	LSTM-CS	Proposed
10	71.92	84.57	90.25	94.85
15	64.10	79.29	86.51	92.54
20	59.75	77.47	80.09	84.84
25	44.40	72.74	73.84	74.99
30	36.52	47.94	54.16	62.38
35	4.29	32.84	40.21	46.97
40	0	15.14	22.52	29.31

within the data[6].

2.5.3 Device Identification

Device identification can help the AP prioritise service provision considering the available resources and provide access to devices based on their priority. Table 2.4 shows the accuracy of correctly identified active devices at $K = 200$, $N = 100$, and SNR = 6 dB. It should be noted again that the traditional schemes in this regard assume complete knowledge of the device sparsity level and that their accuracy is based on identifying the actual active device support set only. On the contrary, the proposed attention-based BiLSTM network follows a practical approach where the active device sparsity level is first estimated. Then, the actual active device support set is identified based on the estimated sparsity level.

We can see from the figure that the trends between the various benchmark schemes are the same as in Fig. 5. The proposed attention-based BiLSTM network outperforms the benchmark schemes by correctly identifying the actual active device support set with higher accuracy. The ML-based LSTM-CS method cannot correctly identify all the active devices because it relies on forward direction architecture only. On the contrary, due to its forward and reverse direction architecture, the proposed BiLSTM network can identify more active devices correctly. It can be seen that with the increasing number of active devices, the accuracy of correctly identifying the actual active device support set decreases, which is to be expected when grant-free NOMA systems operate in overloaded

[6]Note that in order to further enhance the network detection performance, data augmentation techniques can be introduced to control the variations and to improve the model robustness, enabling it to capture complex relationships better. This is outside the scope of this work.

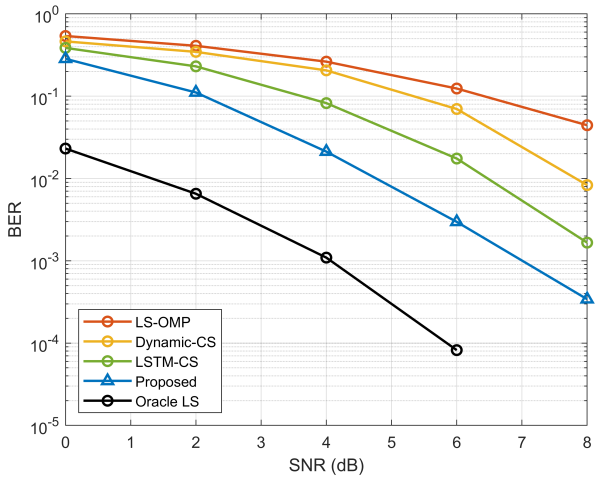


Figure 2.6: Average BER versus the SNR (dB), with the total number of potential devices $K = 200$, the number of subcarriers $N = 100$, and the number of active devices $S = 20$.

conditions.

2.5.4 Multi-User Data Detection

Fig. 2.6 plots the average BER of the considered algorithms against the SNR (dB), with $K = 200$, $N = 100$, and $S = 20$. In all scenarios, our proposed attention-based BiLSTM network outperforms the benchmark schemes over the whole considered range of SNR, including the ML-based LTSM-CS method. For SNR $>$ 4 dB, the gap between the proposed attention-based BiLSTM network and the Oracle LS algorithm is about 3 dB only. This performance gap with the Oracle LS algorithm is because it fully assumes the active device sparsity level and active device support set. The inaccurate active device estimation causes the performance gap as a side effect of the grant-free NOMA system.

Fig. 2.7 plots the average BER against the active device sparsity S, with $K = 200$, $N = 100$, and SNR = 6 dB. Unlike the computational complexity of the proposed network in Section IV-C, the BER performance is impacted by the number of active users. For all methods, the BER decays as the active devices increase. Even so, the proposed attention-based BiLSTM network exhibits consistently lower BER than the benchmark schemes throughout the whole considered range of SNR. The ML-based LSTM-CS method performs better than traditional methods initially but saturates with a high number of active devices since it cannot capture their temporal activation pattern due to its unidirectional architecture. The consistent performance gains of the proposed attention-based BiLSTM

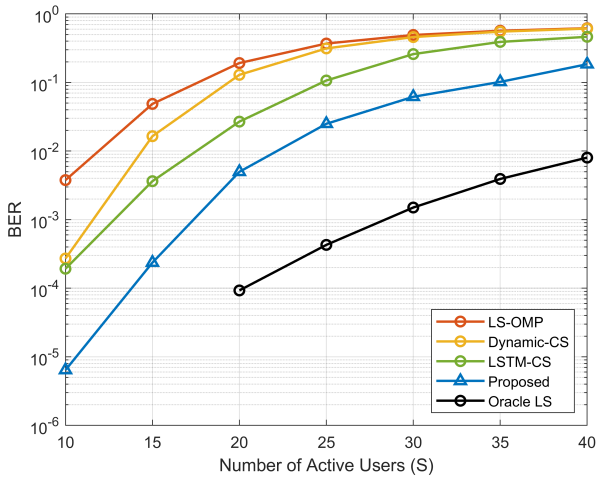


Figure 2.7: Average BER versus the number of active devices S, with total number of potential devices $K = 200$, the number of subcarriers $N = 100$, and SNR = 6 dB.

network show that the network has precisely mapped the underlying relationship between device activity and received signals, given that the network is trained for $S = 20$ active devices.

Fig. 2.8 plots the average BER against the SNR (dB) for varying OFs, with $N = 100$ and $S = 20$. It is evident that the average BER for all benchmark techniques increases with a higher OF as the potential devices K are increased, making the system prone to correlation errors. Even so, the average BER of the proposed attention-based BiLSTM network compared to conventional techniques is lower, manifesting that the proposed attention-based BiLSTM network can load more devices with the same training configuration. This is because the proposed attention-based BiLSTM network has higher tolerance and robustness against increased OFs due to decoupled correlated activation patterns.

Fig. 2.9 plots the average BER against the temporal correlation parameter, η, with $K = 200$, $N = 100$, $S = 20$ and SNR = 6 dB. Note that the result for $\eta = 1$ corresponds to the special case of frame-wise joint sparsity, i.e., devices' activity remains constant over an entire data frame. We can see that the proposed network performs well for all values of η. Herein, the LS-OMP algorithm performs poorly because it does not utilise the extra information present in the previous time slots for temporal activity. Conversely, the BER of the dynamic CS-based method is also relatively higher due to its dependence on devices' activity in the $(j-1)$ time slot only. The ML-based LSTM-CS method performs better than the dynamic CS-based method because it takes the

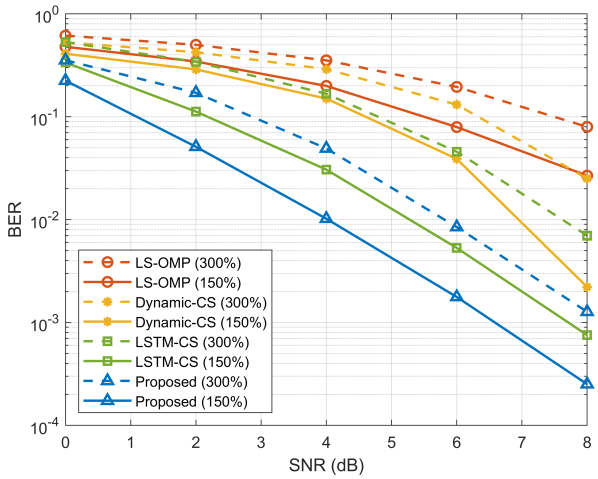


Figure 2.8: Average BER versus SNR (dB) for varying OF, with number of subcarriers $N = 100$, and number of active devices $S = 20$.

temporal activity of devices in all time slots. However, because the ML-based LSTM-CS utilises a forward direction LSTM only, it does not completely capture the activation pattern of active devices. On the contrary, it can be seen that the increasing temporal correlation parameter η enhances the BER performance of the proposed attention-based BiLSTM network. This is because the proposed attention-based BiLSTM network has bidirectional LSTM units, which successfully capture the underlying mapping of the stacked received signal $\widetilde{\mathbf{y}}$ with the temporal correlation of device activity between different time-slots using the estimated support of $\mathbf{x}$. This further testifies to the generability of the proposed attention-based BiLSTM network in different transmission patterns, showing that the proposed attention-based BiLSTM network is not limited to the burst sparsity model and is designed to perform robustly across a range of transmission scenarios. The Oracle LS algorithm outperforms the proposed algorithm and remains consistent since it assumes a complete active device support set.

2.5.5 Discussion on Robustness, Scalability and Generalisation

The results in Figs. 2.5-2.9 show that the proposed attention-based BiLSTM network, which is trained on $S = 20$, $N = 100$, $K = 200$ and $\eta = 0.5$, is robust to changes in the key system parameters. We can see that the trained BiLSTM network still performs well when there is a change in the number of active devices (Figs. 2.5 and 2.7), the number of potential devices or, equivalently, the overloading factor (Fig. 2.8) or temporal correlation

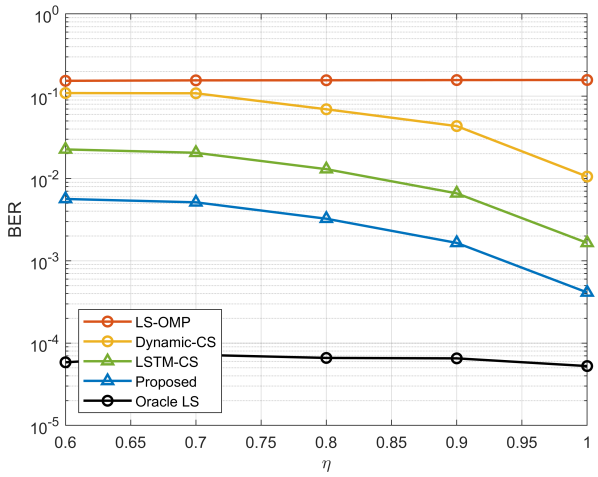


Figure 2.9: Average BER versus the temporal correlation parameter η, with total number of devices $K = 200$, number of subcarriers $N = 100$, number of active devices $S = 20$, and SNR = 6 dB.

model (Fig. 2.9), and does not need to be retrained for the considered practical range of considered values ($10 \leq S \leq 40$, $0.5 \leq \eta \leq 1$ and $150 \leq K \leq 300$). This is because training the network at $\eta = 0.5$ allows it to learn the important features of the device activation patterns, and it still performs well when the parameters change. This shows that the proposed network is generalisable to different system parameters. However, in the event of significant environmental changes (complete channel mismatch), the model would need to be retrained to accommodate entirely new channel characteristics.

In addition, the proposed network is a good solution for grant-free NOMA systems to provide faster access for massive IoT devices. As demonstrated in Table 2.2, the proposed network's computational complexity is comparable to the state-of-the-art ML-based solution and does not heavily depend on the system parameters. Thus, when the number of active devices increases or the number of potential devices in the system becomes large, the computational complexity increases only marginally. Thus, the proposed scheme is scalable and is suitable for faster access in massive IoT device scenarios.

2.6 Summary

In this work, we proposed an attention-based BiLSTM network for AUD in an uplink grant-free NOMA system by exploiting the temporal correlation of active user support sets. First, a BiLSTM network is used to create a pattern of the device activation history

in its hidden layers, whereas the attention mechanism provides essential context to the device activation history pattern. Then, the complex spreading sequences are utilised for blind data detection without explicit channel estimation from the estimated active user support set. Thus, the proposed mechanism is efficient and does not depend on impractical assumptions, such as prior knowledge of active user sparsity or channel conditions. Through simulations, we demonstrated that the proposed mechanism outperforms several existing benchmark MUD algorithms and maintains lower computational complexity. In this work, we have applied the proposed framework to spreading based grant-free NOMA scheme.

Chapter 3

IoT Device Authentication in Terrestrial Networks

3.1 Introduction

The transition from detection and identification (Chapter 2) to authentication (Chapter 3) reflects a natural progression in the IoT security framework. While detection and authentication differ in purpose, there are shared foundations. Both rely on leveraging inherent properties of IoT device transmissions—such as sparsity and access patterns—to achieve lightweight and efficient solutions tailored for resource-constrained IoT environments. By building on these principles, this chapter introduces a novel authentication scheme optimized for terrestrial networks, addressing challenges like computational overhead, misdetection rates, and robustness. In this second technical chapter, we propose a novel lightweight and continuous authentication scheme for resource-constrained IoT devices by identifying the pre-arranged access time slots and spreading pools of each IoT device, which provides high uncertainties for the spoofers and supplies seamless protections for legitimate communications. In our proposed scheme, the access time slots are pre-agreed between a pair of IoT devices and the AP, which are difficult for the adversaries to predict and do not require additional hardware for implementation [87]. The access time slots are generated using the spreading pools available at the AP and IoT devices. The access time slots for every IoT device are generated independently at the AP and the IoT devices, thereby obeying the grant-free NOMA protocol for a practical massive IoT deployment. If the access time slot and spreading pool of an IoT device are different from the access time slot and spreading pool at the AP, it will be identified as an illegitimate device by the AP. *To our best knowledge, this is the first work to authenticate*

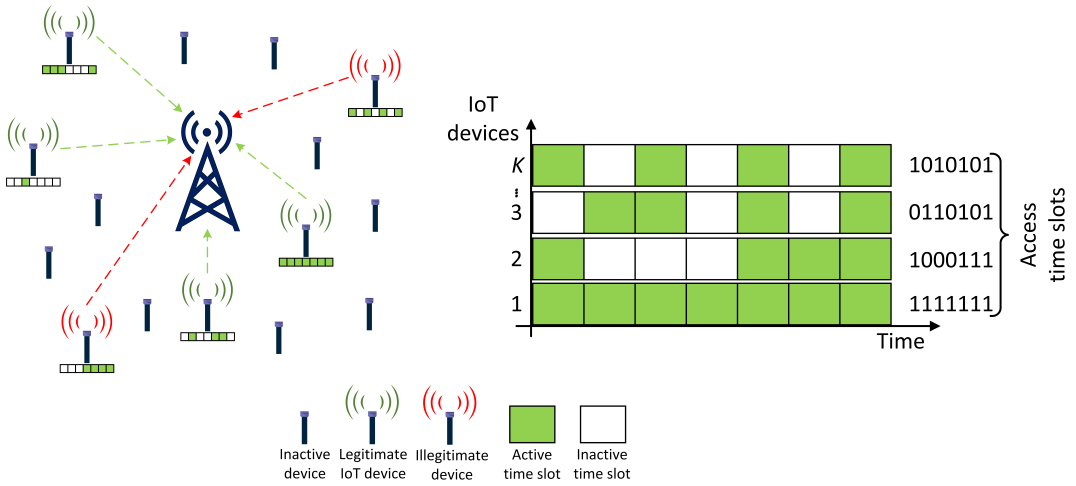


Figure 3.1: Illustration of our system model. The transmission between the IoT devices and the AP is carried out by following the pre-agreed access time slots.

multiple resource-constrained IoT devices utilising grant-free NOMA protocol by utilising their spreading pools and pre-arranged access time slots as the source for authentication.

The rest of this chapter is organised as follows. In Section 3.2, we review the related studies of authentication schemes for IoTs. In Section 3.3, we present the system model and the authentication problem. In Section 3.4, we describe the proposed authentication scheme and provide a detailed description of the different phases of device authentication. Finally, in Section 3.5, we derive the performance analysis of the proposed authentication scheme. In Section 3.6, we present the simulation results to verify the performance gain of the proposed technique.

3.2 System Model

In this work, we consider the scenario where IoT devices wake up sporadically and transmit their data to the AP in a grant-free manner, as depicted in Fig. 3.1. Thus, we consider a spreading-based uplink grant-free NOMA system comprising of an AP and K IoT devices with limited computing capabilities. The AP has relatively powerful computing capabilities and is at a fixed location. The AP and IoT devices are assumed to be equipped with a single antenna, and their clocks are synchronised[1]. We assume that upper-layer security mechanisms are utilised initially to establish system parameters be-

[1]Practically, clock synchronisation can be achieved via methods described in [114–116] to achieve energy-efficient communications for IoT devices. However, this is outside the scope of this work.

Table 3.1: Important symbols used in this work.

Variable	Description	Dimension
K	Total number of IoT devices	1×1
N	Total subcarriers	1×1
S	Active number of IoT devices	1×1
J	Number of time slots	1×1
$\mathbf{c}$	Spreading sequence	$N \times 1$
$\mathbf{h}$	Channel	$N \times 1$
$\mathbf{x}$	Transmit signal	$K \times 1$
$\mathbf{w}$	Gaussian noise	$N \times 1$
$\mathbf{y}$	Received signal	$N \times 1$
$\mathbf{G}$	Synthesis of channel vector and spreading sequences	$N \times K$
$\mathbf{H}$	Channel matrix	$N \times K$
$\mathbf{C}$	Codebook matrix	$N \times K$
$\mathbf{X}$	Transmit signal (continuous time slots)	$K \times J$
$\overline{\mathbf{G}}$	Synthesis of channel vector and spreading sequences (continuous time slots)	$N \times K$
$\mathbf{W}$	Gaussian noise (continuous time slots)	$N \times J$
$\mathbf{Y}$	Received signal (continuous time slots)	$N \times J$
$\overline{\mathbf{\Gamma}}$	Authenticated devices' indicator	$K \times J$
$\widetilde{\mathbf{X}}$	Authenticated devices' data	$K \times J$

tween the AP and IoT devices [85]. During transmission, a subset of the K IoT devices sporadically and randomly become active when they have data to transmit. We consider an overloaded system where the number of resource blocks N is less than the number of IoT devices in a cell, *i.e.*, $N < K$.

3.2.1 Threat Model

In the system model, as depicted in Fig. 3.1, we assume that illegitimate devices can be present anywhere in a cell, including in close proximity to legitimate IoT devices, and therefore, their physical channels can be correlated. As a result, the AP can receive transmissions from both legitimate IoT and illegitimate devices, where the illegitimate devices attempt to access the network by conducting spoofing attacks, such as man-in-the-middle attacks and replay attacks. With this in mind, apart from the codebook matrix[2], we assume that the illegitimate devices utilise the same system parameters and

[2]Generally, the AP can refresh the codebook matrix in a cell to enhance communication using different methods [117, 118]. However, this is a separate research topic and is, therefore, outside the scope of this

upper-layer signalling as the legitimate IoT devices, as detailed in Table 3.1. We further assume that the illegitimate devices can remain active at all times and can scan the network to learn the transmission pattern of legitimate IoT devices. Thus, illegitimate devices can be resourceful and more computationally capable than legitimate IoT devices.

3.2.2 Signal Model

Considering an arbitrary symbol interval, an IoT device randomly wakes up and transmits its complex modulated signal towards the AP, which are independent random variables drawn from a standard symmetric discrete constellation set. After modulation, the transmitted symbol x_k from the k-th IoT device is spread onto a spreading sequence $\mathbf{c}_k$ of length N. The received signal y on the n-th subcarrier at the AP is given as

$$y_n = \sum_{k=1}^{K} h_{nk} c_{nk} x_k + w_n, \tag{3.1}$$

where h_{nk} refer to the n-th subcarrier of the k-th IoT device's channel vector $\mathbf{h}_k = [h_{1k}, h_{2k}, \ldots, h_{Nk}]^T \in \mathbb{C}^{N\times 1}$, c_{nk} refer to the n-th component of the spreading sequence $\mathbf{c}_k = [c_{1k}, c_{2k}, \ldots, c_{Nk}]^T \in \mathbb{C}^{N\times 1}$, and w_n is the Gaussian noise on the n-th subcarrier with zero mean and variance σ^2. By combining the received signals overall N subcarriers, the received signal vector $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T \in \mathbb{C}^{N\times 1}$ is given as

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{w}, \tag{3.2}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_K]^T \in \mathbb{C}^{K\times 1}$ is the transmitted signal vector for all K devices and $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T \in \mathbb{C}^{N\times 1}$ is the noise vector. $\mathbf{G} \in \mathbb{C}^{N\times K}$ is the synthesis of the channel vectors and spreading sequences, given as

$$\mathbf{G} = \mathbf{H} \odot \mathbf{C}, \tag{3.3}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N\times K}$ is the channel matrix, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K] \in \mathbb{C}^{N\times K}$ is the codebook matrix, and $\odot$ is the Hadamard product, *i.e.*, $g_{nk} = h_{nk} c_{nk}$.

3.2.3 Transmission Model

Different works [66, 69, 119] have assumed that the active IoT devices remain unchanged in an entire frame. However, in practical grant-free systems, the IoT devices access or

work.

leave the system randomly [45]. Moreover, once active, due to the size of their data payload, some IoT devices transmit their data in consecutive time slots. From this, it concurs that the nature of data transmission by IoT devices is generically random and not deterministic. Therefore, we consider a scenario where the IoT devices become active or inactive in different time slots, which is a more practical scenario in IoT applications with sporadic communications. Motivated by this, we can extend the signal model in (3.2) from a single time slot transmission model to a continuous time-slots transmission model.

The transmitted signals $\mathbf{X} = \left[\mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \ldots, \mathbf{x}^{[J]}\right] \in \mathbb{C}^{K \times J}$ are recovered from the received signals $\mathbf{Y} = \left[\mathbf{y}^{[1]}, \mathbf{y}^{[2]}, \ldots, \mathbf{y}^{[J]}\right] \in \mathbb{C}^{N \times J}$ in J continuous time slots, based on the LTE-Advanced standard protocol [110]. Thus, the continuous time-slots transmission model for the j-th time slot is given as

$$\mathbf{y}^{[j]} = \mathbf{G}^{[j]}\mathbf{x}^{[j]} + \mathbf{w}^{[j]}, \quad j = 1, 2, \ldots, J, \tag{3.4}$$

where $\mathbf{G}^{[j]} \in \mathbb{C}^{N \times K}$ is the synthesis of the channel vectors and spreading sequences in the j-th time slot and $\mathbf{w}^{[j]}$ is the equivalent Gaussian noise vector in the j-th time slot.

3.2.4 Problem Statement

The sporadic nature of the IoT devices allows the illegitimate devices to impersonate the legitimate IoT devices to spoof the AP and gain access to the core network. Assuming that an IoT device transmits to the AP in the j-th time slot, the objective at the AP is to authenticate the device if the message originated from a legitimate IoT device. In order to achieve this, the AP and the legitimate IoT devices can agree on specific transceiver features or characteristics, which can be used to distinguish legitimate IoT devices from illegitimate devices. Let $\mathbf{\Gamma}^{[j]}$ represent the authenticated devices indicator in the j-th time slot; then, the authentication problem is given as

$$\mathbf{\Gamma}^{[j]} = \begin{cases} 1 & \text{if } \mathcal{H}_0 \\ 0 & \text{if } \mathcal{H}_1 \end{cases}, \tag{3.5}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ represents the received signal $\mathbf{y}^{[j]}$ in the j-th time slot, originated from a legitimate IoT device and an illegitimate device, respectively, and act as the hypothesis for IoT device authentication. The conventional schemes [83, 120, 121] rely on quantisation-based thresholds in (3.5) for decision making. However, the authentication performance significantly declines due to the quantisation errors introduced by

the algorithms. Additionally, it is challenging to obtain optimal values for the detection thresholds to maintain continuous authentication when a large number of IoT devices are involved since exhaustive search methods are utilised to obtain these values.

Another downside to these conventional schemes is that they rely on the physical channel for seed acquisition, verification, reconciliation, and IoT device authentication [80,81,122]. However, reliance on the physical channel for device authentication does not explicitly apply to resource-constrained IoT devices. The reasons for this are as follows.

- A transceiver pair cannot probe the physical channel simultaneously for seed acquisition due to the half-duplex nature of the radio. The resource-constrained IoT devices are assumed to probe the physical channel for seed acquisition and authentication in the conventional physical-channel-based schemes. This is impractical since the resource-constrained IoT devices cannot probe the physical channel due to their limited battery; therefore, the conventional schemes result in excessive battery loss and time lag due to the radio distance turnaround time.

- The reconciliation overhead due to imperfect physical-channel reciprocity increases with the increased key length for seed generation. This means that to achieve a higher authentication rate (by increasing key length), the parity bit information to correct errors is also increased. This is against the spirit of authentication mechanisms for resource-constrained IoT devices, where channel training/probing of IoT devices should be minimised due to their limited resources.

- A transceiver pair separated by a greater than half wavelength distance does not guarantee independent physical channels for seed acquisition [123]. This means that there is no clear safeguard distance to ensure the secrecy of the generated key, as typically assumed in the physical-channel-based seed acquisition techniques [124].

From this discussion, we can conclude that (i) conventional physical-channel-based authentication techniques exhibit these intrinsic limitations, which limits their effectiveness in situations where a transceiver pair experiences spoofing attacks, and (ii) the conventional physical-channel-based seed acquisition techniques are not practical for resource-constrained IoT devices. Therefore, access to a coherent source for identical and lightweight seed generation is crucial for continuous authentication between the AP and resource-constrained IoT devices.

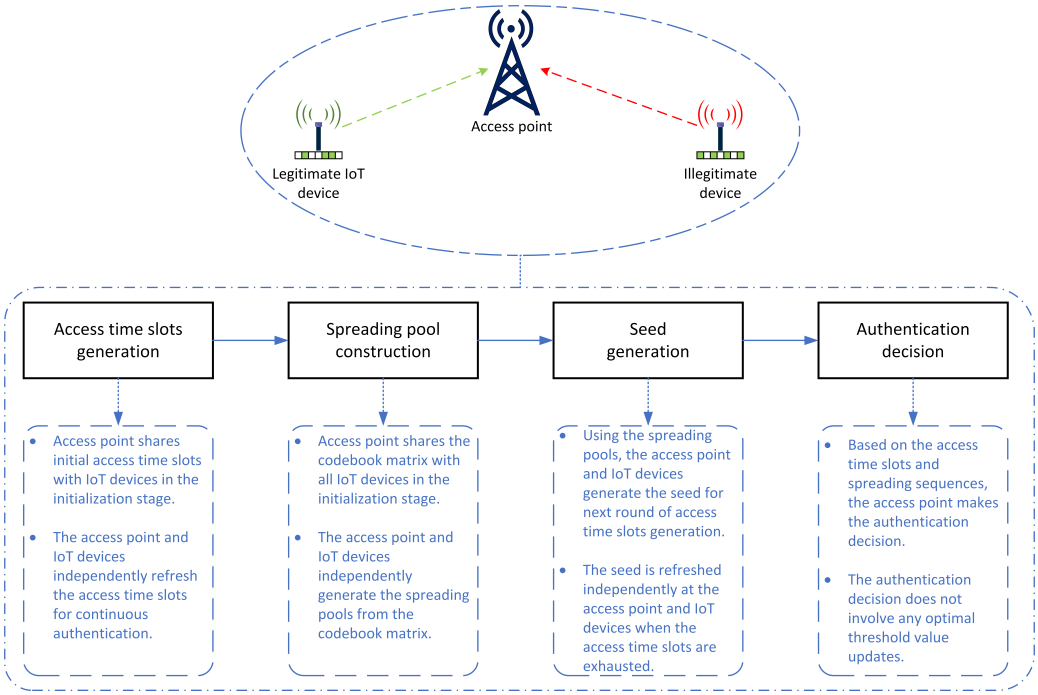


Figure 3.2: Proposed authentication scheme comprises four processes: access time slots generation, spreading pool construction, seed generation, and authentication decision.

3.3 Proposed Authentication Scheme

With the sporadic nature of transmission of IoT devices in mind, the objective at the AP is to authenticate the legitimate IoT devices from the received signal $\mathbf{y}^{[j]}$ in the j-th time slot. Therefore, to achieve authentication, the generated seeds must adhere to the policies as follows [87]: 1) a transceiver pair must generate an identical seed stemming from an identical feature for authentication at the AP; 2) seeds should be undisclosed to any other devices, making the generated feature unpredictable by illegitimate devices; and 3) seeds should be proactively refreshed to maintain continuous authentication while preserving uncertainty for illegitimate devices. To meet these requirements, we use the transmission nature of the grant-free NOMA in (3.4) as the seed source instead of relying on physical-channel attributes for seed acquisition. Then, we use the seed to generate the access time slots for IoT device authentication.

The proposed authentication scheme consists of four processes: (A) access time slots generation, (B) spreading pool construction, (C) seed generation, and (D) authentication decision, and is summarised in Fig. 3.2. The four processes form a cohesive, secure, continuous authentication system between an AP and IoT devices. The AP shares initial

access time slots and a codebook matrix with the IoT devices in the initialisation stage using symmetric encryption, which is per standard by 3GPP and is not unique to the proposed authentication framework. The AP and IoT devices then independently refresh the access time slots and generate spreading pools from the codebook matrix. These spreading pools generate the seed for the next round of access time slots. When the current access time slots are exhausted, the seed is refreshed independently at the AP and IoT devices. Finally, the AP uses the access time slots and spreading sequences to make authentication decisions without needing optimal threshold value updates. By combining these processes, the system ensures that the AP and IoT devices can securely communicate, authenticate each other, and maintain continuous authentication over time. This interaction of the proposed authentication scheme with the grant-free NOMA system is illustrated in Fig. 3.3. The four processes are further explained in detail below.

3.3.1 Access Time Slots Generation

The access time slots for IoT device transmission are divided into recurring time slots of fixed length [87], as depicted in Fig. 3.1. The IoT devices transmit their signals to the AP in time slots pre-agreed upon between the IoT devices and the AP. Therefore, the AP can quickly identify an illegitimate device based on its time slot access. If the seeds are hidden from illegitimate devices, the access time slots are highly unpredictable. More importantly, a seed can generate several access time slots, allowing each IoT device at the AP to be identified continuously for an extended period. Unlike conventional key-based physical-channel schemes, authentication via access time slots does not entail complex computation or high latencies because, in key-based schemes, access to a coherent key is required for every message transmission. In contrast, the access time slots do not require a shared key for every transmission since the transmission schedules are followed by the IoT device and verified by the AP. Thereby, continuous and lightweight authentication between a transceiver is achieved.

The access time slots are generated using linear feedback shift registers, which entails a statistical behaviour close to truly random sequences and does not entail expensive exponential or modulo operations [125]. Therefore, the access time slots are lightweight and challenging to predict by adversaries who do not know the pseudo-random transmission schedules. The access time slots are generated using a monic polynomial of degree μ, which is a prime number with $2^{\mu} - 1$ the maximum length of the generated access

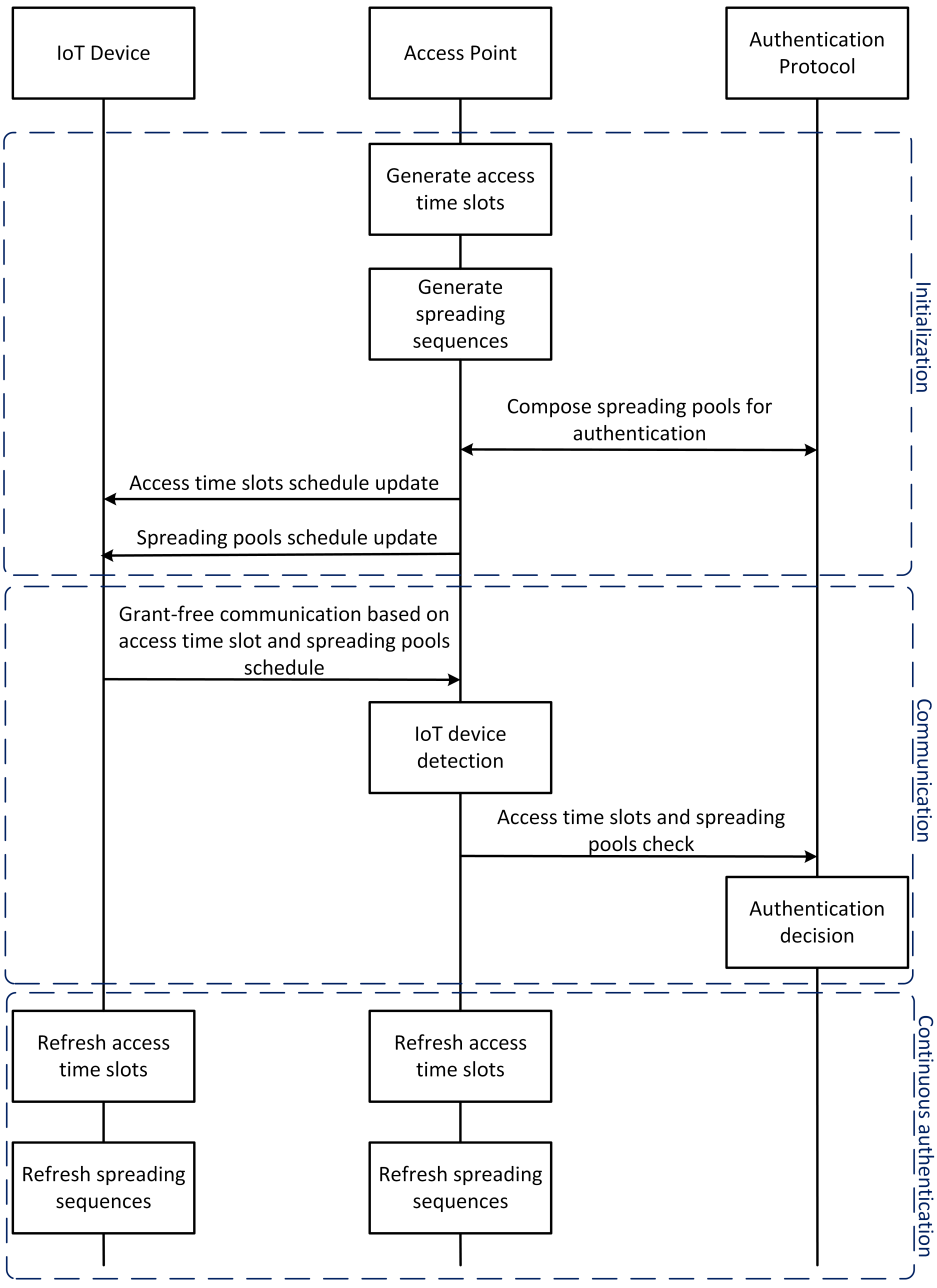


Figure 3.3: Flowchart of proposed authentication scheme and its interaction with grant-free NOMA system model considered in this work.

time slots [126]. The generating monic polynomial for a generic variable $\varkappa$ is given as

$$f(\varkappa) = C_0 + C_1\varkappa + C_2\varkappa^2 + \ldots + C_\mu\varkappa^\mu = \sum_{i=1}^{i=\mu} C_i\varkappa^i, \quad (C_0 = C_\mu = 1), \tag{3.6}$$

where $C_0, C_1, \ldots, C_\mu$ are the constant coefficients of the polynomial. It should be noted that it is difficult for illegitimate devices to predict the monic polynomial function used for the access time slot generation since the AP and IoT devices can refresh the monic polynomial, further enhancing the authentication performance. Furthermore, the process of access time slots generation using the monic polynomial in (3.6) is repeated independently at the AP and k-th IoT device to renew the transmission schedule for continuous authentication, provided they have access to an identical seed. Therefore, a transceiver pair does not have to carry out complex hash function operations for seed concealment and sharing, testifying to the low complexity and lightweight nature of the proposed authentication scheme.

3.3.2 Spreading Pool Construction

In this work, we consider that the transmission symbols of the IoT devices are spread with a family of short complex-valued spreading sequences with low cross-correlation values [1, 49], as shown in (3.2). This allows for loading more IoT devices in a resource block and reducing implementation complexity. Let $\mathbf{C}_{(4,6)}$ represent a complex-valued codebook matrix to support $K = 6$ devices using $N = 4$ resource blocks in an overloaded[3] scenario, given as

$$\mathbf{C}_{(4,6)} = \begin{bmatrix} w_0 & w_4 & w_3 & w_1 & w_6 & w_5 \\ 0 & w_2 & w_6 & w_4 & w_5 & w_0 \\ w_4 & w_7 & w_0 & w_3 & w_0 & 0 \\ w_3 & w_0 & w_2 & w_4 & w_3 & w_6 \end{bmatrix}, \tag{3.7}$$

where w_n is the non-zero elements of the codeword. The non-binary and complex-valued spreading sequences in (3.7) allow for a higher degree of freedom for loading a larger number of IoT devices, thus providing much more flexibility in spreading sequences design,

[3]The OF is defined as the ratio of the number of potential IoT devices to the number of available resource blocks in the system, *i.e.*, OF (%) $= \frac{K}{N} \times 100$.

which is reflected by a high OF and demonstrates a true sense of grant-free transmission[4].

Conventionally in grant-free systems, the codebook matrix in (3.3) is stored locally with the AP and shared with all IoT devices independently in the initialisation stage, which is later utilised by the IoT devices for data transmission [45, 130]. With this sense of practicality in mind, we propose constructing a lightweight mechanism to utilise the codebook matrix in (3.3) for enhanced authentication. This involves constructing spreading pools from the codebook matrix in (3.3) for every IoT device in a cell. Let γ_k denote the spreading pool constructed using the codebook matrix $\mathbf{C}_{(4,6)}$ in (3.7) for the k-th IoT device. As such, for the overloaded scenario in (3.7), the respective spreading pools for $K = 6$ IoT devices can be constructed as

$$\begin{aligned}
\gamma_1 &= \{w_0, 0, w_4, w_3\},\\
\gamma_2 &= \{w_4, w_2, w_7, w_0\},\\
\gamma_3 &= \{w_3, w_6, w_0, w_2\},\\
\gamma_4 &= \{w_1, w_4, w_3, w_4\},\\
\gamma_5 &= \{w_6, w_5, w_0, w_3\},\\
\gamma_6 &= \{w_5, w_0, 0, w_6\}.
\end{aligned} \tag{3.8}$$

Once the spreading pools are constructed, the access time slots are superimposed over the spreading pools for intelligent transmission and enhanced authentication. Thus, the spreading pools in (3.8) can therefore be rewritten as

$$\begin{aligned}
\gamma_1 &= \{\overbrace{w_0}^{1}, \overbrace{0}^{1}, \overbrace{w_4}^{0}, \overbrace{w_3}^{0}\},\\
\gamma_2 &= \{\overbrace{w_4}^{1}, \overbrace{w_2}^{1}, \overbrace{w_7}^{0}, \overbrace{w_0}^{1}\},\\
\gamma_3 &= \{\overbrace{w_3}^{1}, \overbrace{w_6}^{0}, \overbrace{w_0}^{0}, \overbrace{w_2}^{0}\},\\
\gamma_4 &= \{\overbrace{w_1}^{1}, \overbrace{w_4}^{1}, \overbrace{w_3}^{1}, \overbrace{w_4}^{0}\},\\
\gamma_5 &= \{\overbrace{w_6}^{1}, \overbrace{w_5}^{0}, \overbrace{w_0}^{0}, \overbrace{w_3}^{0}\},\\
\gamma_6 &= \{\overbrace{w_5}^{1}, \overbrace{w_0}^{1}, \overbrace{0}^{1}, \overbrace{w_6}^{0}\}.
\end{aligned} \tag{3.9}$$

From (3.9), it can be seen that by jointly utilising the spreading pools and access

[4]The design of the codebook matrix can be carried out in different ways [45, 127–129] to enhance the OF of the system further. However, this is outside the scope of this work.

time slots, an enhanced security mechanism can be developed, which provides a higher degree of system efficiency (reduction in spreading sequence collision due to intelligent transmission) and security entropy (a two-step mechanism for device authentication). The utilised spreading pools by the respective IoT devices are then used for seed and refreshed access time slot generation. Herein, it should be noted that a longer length of spreading pool and access time slots results in a higher authentication entropy. However, a shorter length results in lower bit-error-rate (BER) performance. This demonstrates a trade-off between authentication and BER performance which can be controlled based on the network requirements.

3.3.3 Seed Generation

Once the spreading pools and their tagged access time slots are exhausted, the AP and IoT devices need to recreate newer spreading pools and access time slots for continuous authentication. In this regard, the k-th IoT device can use its current spreading pool to generate a seed value for the newer pools. Let $(c_{1k}, c_{2k}, \ldots, c_{Nk})$ represent the length of the spreading sequences inside a spreading pool γ_k, and $(l_{1k}, l_{2k}, \ldots, l_{Lk})$ represent the access time slots of the k-th IoT device. Then, we generate the seed by taking the XOR of the access time slots and calculating the ℓ_2 norm of the tagged spreading sequences. This process for an arbitrary spreading pool γ_k of the k-th IoT device is as follows:
Step 1: Take the original spreading pool and its superimposed access time slots

$$\gamma_k = \{\overbrace{w_0}^{1}, \overbrace{0}^{1}, \overbrace{w_4}^{0}, \overbrace{w_3}^{0}\} \tag{3.10}$$

Step 2: Take XOR of the access time slots

$$\gamma_k = \{\overbrace{w_0}^{0}, \overbrace{0}^{0}, \overbrace{w_4}^{1}, \overbrace{w_3}^{1}\} \tag{3.11}$$

Step 3: Nullify the spreading sequences under 0's

$$\gamma_k = \{\overbrace{0}^{0}, \overbrace{0}^{0}, \overbrace{w_4}^{1}, \overbrace{w_3}^{1}\} \tag{3.12}$$

Step 4: Take the sum and ℓ_2 norm of the spreading sequences under 1's to obtain preliminary seed

$$\Theta = ||w_4 + w_3||_2 \tag{3.13}$$

Step 5: Take the square of the preliminary seed to obtain the final seed

$$seed = \Theta^2. \tag{3.14}$$

This process is performed independently at the AP and the IoT devices. It should be noted that steps 4 and 5 depend on the resource availability of the IoT devices. That is to say; if the IoT devices are extremely resource-constrained, the preliminary Θ can be used for access time slots generation since it averts computationally expensive $\mathcal{O}(L^2)$ operation in step 5, as well as results in a shorter key length. However, step 5 provides a longer key length for increased authentication, thereby providing prolonged authentication. The choice of seed in steps 4 and 5 demonstrates a trade-off between the computational performance and security performance of a transceiver pair. Hence, this process should be well-designed to achieve a better trade-off. Furthermore, it should be noted that, unlike the conventional physical-channel-based schemes, the proposed authentication scheme does not rely on channel probing for seed acquisition, seed reconciliation, or authentication. This means that the seed verification phase, which is required in the conventional physical-channel-based authentication schemes due to either imperfect channel probing or quantisation errors, is not needed in the proposed authentication scheme, thus paving the way for a practical, lightweight, and independent authentication mechanism in a grant-free NOMA system.

3.3.4 Authentication Decision

The conventional physical-channel-based authentication schemes rely on quantisation-aided hypothesis testing as a decision criterion in (3.5). However, such benchmarks rely on static statistical properties of the physical channel and cannot account for varying attributes of fast-fading physical-channel characteristics, resulting in misdetection. As opposed to this, the proposed authentication scheme does not rely on a quantisation-based threshold as an authentication criterion. Instead, the proposed scheme utilises a two-step authentication decision process, where the AP first matches the access time slots of the transceiver pair and then compares the spreading sequences of the following transmitting schedule. The two-step authentication process enables mitigating misdetection at the AP and averts false alarms. This authentication process is summarised in Algorithm 1, and the main procedure is presented as follows.

1. *Line 2:* The sparse transmitted signal vector in the j-th time slot is estimated and

detected at the AP by the least squares algorithm as [62]:

$$\widehat{\mathbf{x}}^{[j]} = \left(\mathbf{G}^{[j]}\right)^{\dagger} \mathbf{y}^{[j]}. \tag{3.15}$$

2. *Line 3:* The codebook matrix $\mathbf{C}^{[j]}$ utilised by the IoT devices in the j-th time slot is extracted by applying Hadamard division on the channel matrix as:

$$\mathbf{C}^{[j]} = \mathbf{G}^{[j]} \oslash \mathbf{H}^{[j]}. \tag{3.16}$$

3. *Line 6:* The spreading pools and the transmission schedule of the K IoT devices is extracted from the codebook matrix in the j-th time slot as:

$$\gamma_k^{[j](l)}[\text{device}] = \mathbf{C}^{[j]}(:, k). \tag{3.17}$$

4. *Line 7-12:* The l-th access time slot of the k-th IoT device $\gamma_k^{[j](l)}[\text{device}]$ in the j-th time slot is compared with the l-th access time slot of the AP $\gamma_k^{[j](l)}[\text{AP}]$ in the j-th time slot. If the access time slot matches, the authenticated devices indicator function $\mathbf{\Gamma}_k^{[j]}$ for the k-th device in the j-th time slot is set to 1. Otherwise, the indicator function records a 0, deeming the k-th device as illegitimate.

5. *Line 13-17:* The l-th spreading sequence of the k-th IoT device $\gamma^{[j]}(k, l)[\text{device}]$ from the extracted spreading pool in the j-th time slot is compared with the l-th spreading sequence of the AP $\gamma^{[j]}(k, l)[\text{AP}]$ in the j-th time slot. If the spreading sequence matches, the authenticated devices indicator function $\mathbf{\Gamma}_k^{[j]}$ for the k-th device in the j-th time slot is set to 1. Otherwise, the indicator function records a 0, deeming the k-th device as illegitimate.

6. *Line 20:* The authenticated devices data $\widetilde{\mathbf{x}}^{[j]}$ in the j-th time slot is determined by calculating the Hadamard product between the estimated sparse transmitted signal vector $\widehat{\mathbf{x}}^{[j]}$ and the authenticated devices indicator function $\mathbf{\Gamma}^{[j]}$ in the j-th time slot, given as:

$$\widetilde{\mathbf{x}}^{[j]} = \widehat{\mathbf{x}}^{[j]} \odot \mathbf{\Gamma}^{[j]}. \tag{3.18}$$

At the end of the iteration, the authenticated devices data $\widetilde{\mathbf{x}}^{[j]}$ in the j-th time slot is transformed into a sparse vector, where the data of the illegitimate devices is replaced with 0's, whereas the authenticated devices data is recovered.

Algorithm 2 The Proposed Authentication Scheme.

Input:

Received signals: $\mathbf{Y} = \left[\mathbf{y}^{[1]}, \mathbf{y}^{[2]}, \ldots, \mathbf{y}^{[J]}\right]$;

Equivalent channel matrices: $\overline{\mathbf{G}} = \left[\mathbf{G}^{[1]}, \mathbf{G}^{[2]}, \ldots, \mathbf{G}^{[J]}\right]$.

Output:

Authenticated devices indicator: $\overline{\mathbf{\Gamma}} = \left[\mathbf{\Gamma}^{[1]}, \mathbf{\Gamma}^{[2]}, \ldots, \mathbf{\Gamma}^{[J]}\right]$;

Authenticated devices symbols: $\widetilde{\mathbf{X}} = \left[\widetilde{\mathbf{x}}^{[1]}, \widetilde{\mathbf{x}}^{[2]}, \ldots, \widetilde{\mathbf{x}}^{[J]}\right]$.

Device detection

1: **for** $j = 1$ to J **do**

2: $\quad \widehat{\mathbf{x}}^{[j]} = \left(\mathbf{G}^{[j]}\right)^{\dagger} \mathbf{y}^{[j]}$

3: $\quad \mathbf{C}^{[j]} = \mathbf{G}^{[j]} \oslash \mathbf{H}^{[j]}$

Device authentication

4: $\quad$ **for** $l = 1$ to L **do**

5: $\quad\quad$ **for** $k = 1$ to K **do**

6: $\quad\quad\quad \gamma_k^{[j](l)}[\text{device}] = \mathbf{C}^{[j]}(:, k)$.

Step 1: (Access time slot check)

7: $\quad\quad\quad$ **if** $\gamma_k^{[j](l)}[\text{AP}] == \gamma_k^{[j](l)}[\text{device}]$ **then**

8: $\quad\quad\quad\quad \mathbf{\Gamma}_k^{[j](l)} = 1$.

9: $\quad\quad\quad$ **else**

10: $\quad\quad\quad\quad \mathbf{\Gamma}_k^{[j](l)} = 0$.

11: $\quad\quad\quad\quad$ Skip to line 17.

Step 2: (Spreading sequence check)

12: $\quad\quad\quad$ **if** $\gamma^{[j]}(k, l)[\text{AP}] == \gamma^{[j]}(k, l)[\text{device}]$ **then**

13: $\quad\quad\quad\quad \mathbf{\Gamma}_k^{[j](l)} = 1$.

14: $\quad\quad\quad$ **else**

15: $\quad\quad\quad\quad \mathbf{\Gamma}_k^{[j](l)} = 0$.

16: $\quad \widetilde{\mathbf{x}}^{[j]} = \widehat{\mathbf{x}}^{[j]} \odot \mathbf{\Gamma}^{[j]}$.

Return:

$\overline{\mathbf{\Gamma}} = \left[\mathbf{\Gamma}^{[1]}, \mathbf{\Gamma}^{[2]}, \ldots, \mathbf{\Gamma}^{[J]}\right]$;

$\widetilde{\mathbf{X}} = \left[\widetilde{\mathbf{x}}^{[1]}, \widetilde{\mathbf{x}}^{[2]}, \ldots, \widetilde{\mathbf{x}}^{[J]}\right]$.

3.4 Security Performance Analysis

The performance of any new authentication scheme can be assessed using security analysis. A comprehensive formal security analysis often necessitates sophisticated modelling, which entails using advanced mathematical frameworks and cryptographic primitives to replicate potential threat scenarios and evaluate system vulnerabilities. In such modelling, formal methods and symbolic representations are employed to capture and analyse the intricate dynamics of potential attacks and the protective countermeasures of the system. This intricate modelling process aims to uncover hidden vulnerabilities, test the system's resilience against various threats, and derive insights for strengthening the system's defence mechanisms [131]. However, a formal security analysis is outside the scope of this work. Instead, similar to [87], the effectiveness of our proposed authentication scheme can be assessed rigorously using performance metrics such as entropy, key space, and computational efficiency. Here is why these metrics are employed:

- Entropy: This metric indicates a system's resilience against unauthorised access. Specifically, greater entropy suggests that an illegitimate device would be computationally arduous to predict or deduce the system's state.

- Key Space: This metric represents the total set of potential keys that could be employed within the system, offering a quantifiable measure of its complexity against brute-force attacks.

- Lightweight: This metric aims to minimise computational demands and resource consumption while maintaining stringent security standards.

By focusing on these metrics, we can demonstrate the robustness and security performance of the proposed authentication scheme.

3.4.1 Entropy

Legitimate IoT devices go through periodic updates of the access time slots and spreading pools; therefore, it is challenging for illegitimate devices to spoof the AP. Furthermore, since a transceiver pair independently but identically utilises multiple spreading sequences from the spreading pool for seed generation, they are difficult for illegitimate devices to predict. Following this, it is clear that the seed is concealed from an adversary if it does not know the access time slots and the corresponding spreading pools. Furthermore, updating the access time slots and spreading pools will provide further protection for

legitimate IoT devices by renewing their access sequences over time. Hence, the proposed authentication scheme provides enhanced protection against spoofing attacks and pertains to legitimate communications between IoT devices and the AP.

With this understanding, entropy is defined as a metric that measures the uncertainty associated with the randomness of a system [132] and is used to evaluate the security strength of the authentication scheme. Thus, entropy is defined as

$$E_{total} = \sum_{r=1}^{R} E_r, \tag{3.19}$$

where

$$E_r = -p_{r0} \log p_{r0} - (1 - p_{r0}) \log(1 - p_{r0}). \tag{3.20}$$

R represents the total length of the shared key, and p_{r0} denotes the posterior probability of the r-th bit when it is 0 from the illegitimate devices' knowledge.

Claim 3.1 *The entropy of the proposed authentication scheme is higher than that of the physical-channel key generation schemes of [132–135].*

Proof: We provide proof using heuristic arguments as follows. Assuming R denotes the length of the access time slots and the key in the physical-channel key generation schemes, we denote p_{r0}^{I} and p_{r0}^{II} as their posterior probabilities of the r-th bit when it is 0 from the illegitimate devices' knowledge, respectively. It should be noted that the proposed authentication scheme relies on multiple attributes, *i.e.*, it utilises N spreading sequences for seed generation. On the contrary, the physical-channel key generation schemes rely on a single attribute for shared key generation. Since multiple attributes are being utilised in the proposed authentication scheme and legitimate IoT devices follow the pre-agreed access time slots for transmission, it is difficult for illegitimate devices to spoof the AP. Then,

$$\left| p_{r0}^{\mathrm{I}} - \frac{1}{2} \right| < \left| p_{r0}^{\mathrm{II}} - \frac{1}{2} \right| \tag{3.21}$$

holds [87], which means the illegitimate devices have less knowledge that the r-th bit is 0 in the proposed authentication scheme. Let E_r^{I} denote the entropy of the proposed authentication scheme, and E_r^{II} denote the entropy of the physical-channel key generation schemes. Then, from (3.21), we can concur that

$$E_r^{\mathrm{I}} > E_r^{\mathrm{II}} \tag{3.22}$$

holds. This completes the proof.

3.4.2 Key Space

Due to their limited computational resources, the resource-constrained IoT devices cannot compute shared keys for every data transmission, required by conventional encryption methods. To overcome this inherent issue, resource-constrained IoT devices rely on shortened keys to reduce the computational overhead. However, shortened keys can be more vulnerable to malicious attacks as they can be easily cracked by attackers using brute force. This is because sophisticated attackers with rapidly growing processing power can compromise the short-length keys within a much shorter time than before, for example, by using exhaustive search approaches [124]. Therefore, an additional layer of security based on low computational cost is required. Based on multi-factor attributes, the proposed authentication method complements the overall security paradigm by acting as another source of randomness to provide additional entropy to the system. This authentication at the lower layer compensates for entropy loss due to the use of shortened keys in the higher layers in resource-constrained IoT devices.

Claim 3.2 *The key space of the proposed authentication scheme is higher than that of the physical-channel key generation schemes of [132–135].*

Proof: We provide proof using heuristic arguments as follows. Assuming that R represents the length of the key in the proposed authentication scheme and physical-channel key generation schemes, we denote κ_R^{I} and κ_R^{II} as the upper bound of the key search space, respectively. We know that the proposed authentication scheme utilises the access time slots and the complex spreading sequences for IoT device authentication. On the other hand, the physical-channel key generation schemes rely on the attribute of the physical channel for key generation. Thus, in Table 3.2, we demonstrate the key search space versus the key length of the proposed authentication scheme and physical-channel key generation schemes. It is evident that the proposed authentication scheme achieves a higher search space than the physical-channel key generation schemes for the same key length. This is because the proposed technique utilises complex spreading sequences and access time slots, which adds another source of randomness to the system for key generation. Therefore, the proposed authentication scheme is less susceptible to brute force attacks than the physical-channel key generation schemes for the same key length. Thus, from Table 3.2, it is concurred that

$$\kappa_R^{\mathrm{I}} > \kappa_R^{\mathrm{II}} \tag{3.23}$$

holds. This completes the proof.

Table 3.2: Key length versus search space complexity of physical-channel-based and proposed techniques.

Key length	Physical-channel key generation schemes		Proposed authentication scheme	
	Search space	Authentication complexity	Search space	Authentication complexity
9	512	$\mathcal{O}(\mathcal{N})$	8192	$\mathcal{O}(1)$
11	2048		32768	
13	8192		131072	
15	32768		524288	
17	131072		2097152	

Since the proposed authentication scheme introduces more randomness into the network, the total system entropy E_{total} is higher than physical-channel key generation schemes. Hence, the proposed authentication scheme can be integrated into the network to provide additional entropy for improving the system's resistance to attacks.

3.4.3 Lightweight

The proposed authentication scheme utilises the transmission parameters and access time slots for IoT device authentication. Conversely, the proposed authentication scheme does not rely on physical-channel probing for IoT device authentication. As a result, the seed verification phase is not required in our proposed authentication scheme. More importantly, the proposed schemes provide continuous authentication by checking the spreading sequences and access time slots of the IoT devices instead of generating and verifying shared keys repeatedly. As a result, as shown in Table 3.2, compared to the physical-channel-based key generation schemes, the proposed authentication schemes achieve a lower authentication complexity for $\mathcal{N}$ times of authentication, which validates the lightweight nature of the proposed authentication scheme.

3.5 Results and Discussion

In this section, we evaluate the performance of the proposed authentication scheme in solving the device authentication problem. We plot the performance of three physical-channel-based authentication benchmark solutions: using binary hypothesis testing (BHT) [84], using ML-based SVM [136], and using deep neural network-based (NN) detection [137]. For these three benchmark solutions, the core architectures are borrowed from the respective works but their input configurations have been adjusted to our sys-

Table 3.3: Access time slots generation using seed.

State	Observations
Spreading pool utilised between a transceiver pair	$\gamma = \{-4-4i, -0+8i, 1-1i, \ldots, -2+2i, 4\}$
Seed extracted by the AP	1111010000000
Seed extracted by the IoT	1111010000000
Access time slots at the AP and IoT	10000100000011101101001010110011 11000111000010010011111011101001110 11111000111001011110011111110010011 11011001111001000001

tem model for a fair comparison. For these benchmark solutions, the estimates of the received signal strength indicator (RSSI), the channel impulse response (CIR), and the channel frequency response (CFR) are used as attributes from the physical channel for authentication [122]. Specifically, due to the correlation of adjacent CIRs and CFRs on the same path, the temporal process of the i-th subpath at the j-th time slot is given as [122]

$$h_i(j) = \zeta h_i(j-1) + \sqrt{(1-\zeta^2)\sigma_i^2} u_i(j-1), \tag{3.24}$$

where $\zeta \in [0, 1]$ represents the physical-channel correlation of two successive subpaths and u_i is a driving noise which is modelled as a zero-mean complex Gaussian random variable with unit variance [84]. The path loss between the AP and the k-th IoT device is modelled as $128.1 + 37.6 \log_{10}(d_i)$, where d_i is the distance (in km) [110]. Additionally, for the benchmark schemes, the physical channels of the illegitimate devices are assumed to be independent of the legitimate IoT devices, meaning the illegitimate devices are assumed to be at a distance greater than half wavelength from the legitimate IoT devices.

Assuming initial authentication between a transceiver pair in the j-th time slot, their observation characteristics are shown in Table 3.3. As detailed in section III-C, the AP and IoT device independently extract the seed by utilising the spreading pool used for data transmission. Since the seed source is the spreading pool, extracted from the codebook matrix and available with the transceiver pair locally, there is no requirement for seed verification. Therefore, once the seed is acquired, the AP and IoT independently generate the access time slots required for transmission. In this work, we utilise the following monic polynomial for the access time slots generation

$$f(\varkappa) = 1 + \varkappa^1 + \varkappa^3. \tag{3.25}$$

3.5.1 Experimental Setup

In the simulations, unless otherwise stated, $K = 200$ potential devices simultaneously share $N = 100$ resources. Thus, the OF is 200%. For every time slot, there is $S = 20$ number of active devices randomly selected from the set $\{1, 2, \ldots, K\}$. $S = 20$ was chosen as it represents a moderate and practical number of active devices commonly encountered in resource-constrained IoT networks, balancing network sparsity and congestion. This selection aligns with configurations in related literature, ensuring both relevance and a realistic baseline for performance evaluation. The number of time slots is fixed at $J = 7$. The transmitted signals are modulated by QPSK. The SNR range is set between 0 to 25 dB. The oracle least squares algorithm is utilised for device detection.

The simulations are carried out on the Gadi supercomputer of the National Computational Infrastructure (NCI), Australia, utilising 48 cores of Intel Xeon Platinum 8274 (Cascade Lake) processors and 192GB of random access memory. The simulations are carried out on MATLAB 2021b. The results are averaged over 1000 Monte Carlo trials.

3.5.2 Performance Metrics

In order to appropriately evaluate the authentication performance, we use the following metrics: the false alarm rate (ρ_{fa}), the misdetection rate (ρ_{md}), and the spreading sequence collision rate (ρ_{sc}) as performance metrics. Given the transmit signal $\mathbf{x}$, authenticated devices data $\widetilde{\mathbf{x}}$, the authenticated devices indicator $\mathbf{\Gamma}$, and the spreading pool γ for the k-th IoT device in the j-th time slot, the performance metrics are defined as follows.

- False alarm rate: This metric evaluates the rate of legitimate IoT devices being falsely detected as illegitimate devices, given as

$$\rho_{\text{fa}} = \frac{1}{K} \sum_{k \in \mathbf{x}^{[j]}} P\left\{ \mathbf{\Gamma}_k^{[j]} = 0 \mid \mathbf{x}_k^{[j]} = 1 \right\}. \tag{3.26}$$

- Misdetection rate: This metric evaluates the rate of illegitimate IoT devices being misdetected, given as

$$\rho_{\text{md}} = \frac{1}{K} \sum_{k \in \mathbf{x}^{[j]}} P\left\{ \mathbf{\Gamma}_k^{[j]} = 1 \mid \mathbf{x}_k^{[j]} = 0 \right\}. \tag{3.27}$$

- Spreading sequence collision rate: This metric evaluates the rate of legitimate IoT devices utilising the same spreading sequence in the same access time slot, given

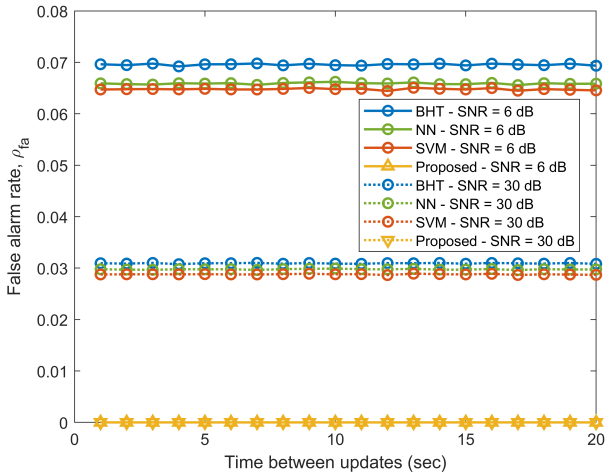


Figure 3.4: False alarm rate, ρ_{fa}, versus the time between updates (sec), with the total number of potential devices $K = 200$, the number of resources $N = 100$, and the number of active devices $S = 20$.

as:

$$\rho_{\text{sc}} = \frac{1}{K} \sum_{k \in \widetilde{\mathbf{x}}^{[j]}} P\left\{ \gamma_k^{[j](l)} == \gamma_{i \neq k}^{[j](l)} \right\}. \tag{3.28}$$

3.5.3 Authentication Performance

Fig. 3.4 plots the false alarm rate, ρ_{fa}, versus the time between updates (sec) for $K = 200$, $N = 100$, and $S = 20$. The false alarm events are avoided in the proposed authentication scheme due to the spreading sequences-based seed generation technique proposed in this work. The spreading sequences-based seed generation allows AP and IoT devices to independently acquire identical seeds for the access time slots generation. In essence, the access time slots generated in the proposed authentication scheme between the AP and an IoT device are identical and do not require parity bits for seed reconciliation. On the contrary, since the benchmark schemes rely on estimates of multiple attributes of the physical channel, false alarm events are inevitable due to the imperfect and time-varying nature of the physical channel encountered due to reliance on the randomness of the channel for seed acquisition. Moreover, lower SNR could lead to a higher false alarm rate in physical-channel-based schemes since its performance explicitly relies on observing physical-channel attributes.

Fig. 3.5 plots the misdetection rate, ρ_{md}, versus SNR (dB) for $K = 200$, $N = 100$,

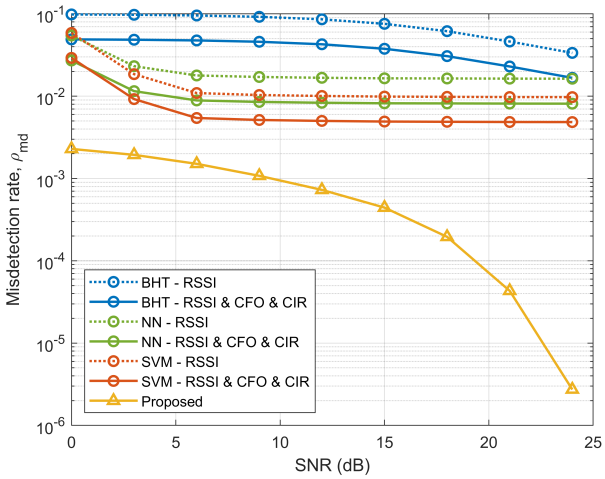


Figure 3.5: Misdetection rate, ρ_{md}, versus SNR (dB), with the total number of potential devices $K = 200$, the number of resources $N = 100$, and the number of active devices $S = 20$.

and $S = 20$[5]. We can observe that in the entire SNR range, the proposed authentication scheme's misdetection rate decreases and achieves a near-threefold performance gain against the benchmark schemes at the higher SNR range. For instance, the performance gain is around 10 dB compared to the traditional BHT-based authentication scheme at SNR = 6 dB. This trend is because the AP and IoT devices identically but independently generate the access time slots using the spreading sequences. These spreading sequences and the access time slots are then used for IoT device authentication. Hence, the proposed authentication scheme is robust in the noisy wireless communication environment. Fig. 3.5 also demonstrates the authentication performance of the benchmark schemes for single and multiple attributes, which rely on estimates of these attributes from the physical channel for device authentication. It can be seen that the benchmark schemes have a higher misdetection rate at lower SNR, which is due to the imperfect physical-channel mismatch between the AP and IoT devices, which requires the continuous updating of the decision boundary. More importantly, the reliance of the proposed authentication scheme on spreading sequences for continuous authentication adds an additional element to the authentication mechanism and generally makes it more difficult for an illegitimate device to spoof the AP under the proposed authentication protocol. By employing our proposed authentication scheme, the AP gains the ability to differentiate between legitimate and

[5]Fig. 3.5 is simulated with 100,000 Monte Carlo trials to evaluate its performance for the entire SNR range. This simulation took 19 hours to execute on the Gadi NCI supercomputer.

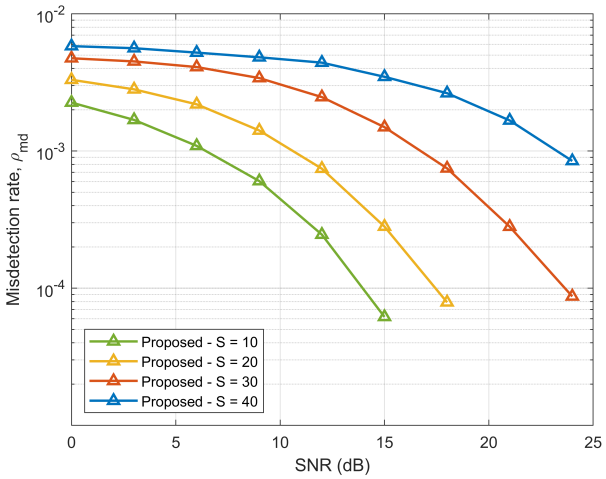


Figure 3.6: Misdetection rate, ρ_{md}, versus SNR (dB) for the varying number of active devices S, with the total number of potential devices $K = 200$, and the number of resources $N = 100$.

illegitimate devices based on their utilisation of spreading sequences and transmission characteristics. Consequently, our proposed authentication scheme eliminates the disparities introduced by distance-related factors when distinguishing between legitimate and illegitimate devices, and therefore, the correlated physical channel characteristics do not play a role in spoofing the AP.

3.5.4 Robustness in Different Configurations

Fig. 3.6 plots the misdetection rate, ρ_{md}, versus SNR (dB) for the varying number of active devices S, with $K = 200$, and $N = 100$. It can be seen that the proposed authentication scheme is capable of handling a variety of active transmitting devices S. This is because the proposed authentication scheme does not rely on physical channels for binary testing as a decision boundary, which requires an update to the decision boundary for every change in the number of active devices S. Since the proposed authentication scheme relies on the spreading sequences extracted from the codebook matrix, the proposed authentication scheme can adapt to any number of active transmitting devices S. It should be noted that the reduction in misdetection rate ρ_{md}, caused by the increase in the number of active transmitting devices S is due to the device estimation errors, which is a side effect of the grant-free NOMA system.

Fig. 3.7 plots the spreading sequence collision rate, ρ_{sc}, versus the number of active devices S for different OF settings, with $K = 200$. The spreading sequence collision rate

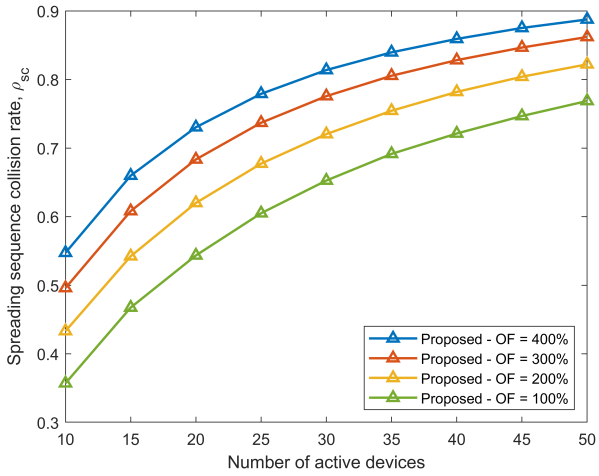


Figure 3.7: Spreading sequence collision rate, ρ_{sc}, versus the varying number of active devices S, with the total number of potential devices $K = 200$.

ρ_{sc} increases with the number of active devices S. It is also evident that a low number of resources N results in a higher OF, which also increases the spreading sequence collision rate ρ_{sc}. This is because when more active devices S transmit simultaneously with shared resources N, the probability of the two or more active devices using the same resource for transmission increases, which increases the spreading sequence collision rate ρ_{sc}. It should be noted that these collisions result from the system's bottleneck due to the inherent nature of the grant-free NOMA systems. Even so, the proposed authentication scheme can handle various active devices S and therefore is robust to different system settings.

Fig. 3.8 plots the misdetection rate, ρ_{md}, versus the time between updates (sec) for $K = 200$, $N = 100$, and $S = 20$. It can be seen that with the increase in the length L of the access time slots, the misdetection rate ρ_{md} of the proposed authentication scheme decreases. This is because the longer length of access time slots results in a more randomised transmission pattern for legitimate IoT devices, which is difficult for an illegitimate device to predict and spoof the AP. However, shorter lengths of access time slots, which result in a higher misdetection rate, are less computationally expensive to generate. Therefore, the choice between the length of the access time slots and the system's computational requirements is a trade-off that can be carefully chosen, depending on the requirement of the network.

Fig. 3.9 plots the computational cost versus the time between updates (sec) for $K = 200$, $N = 100$, and $S = 20$. It can be seen that the proposed authentication scheme

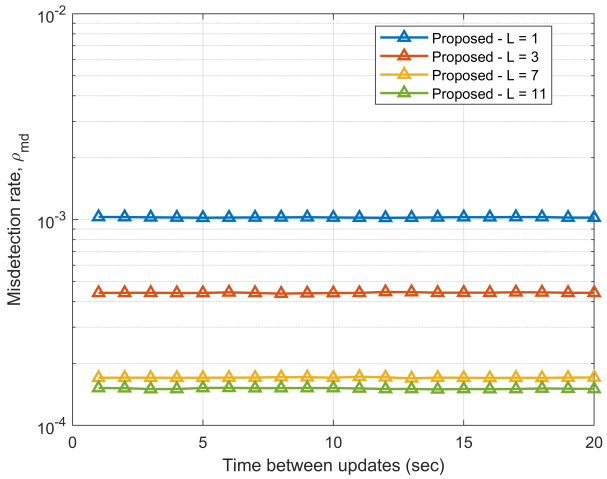


Figure 3.8: Misdetection rate, ρ_{md}, versus the time between updates (sec) for the varying length of authentication sequence L, with the total number of potential devices $K = 200$, the number of resources $N = 100$, and the number of active devices $S = 20$.

attains a lower computational cost than the benchmark schemes. This is because the proposed authentication scheme relies on the access time slots and the spreading pools as its source of IoT device authentication. Since the codebook matrix, which is utilised to derive the spreading pools, is managed by the AP and does not require creating any threshold boundaries, the proposed scheme has a lower computational cost. On the contrary, the physical-channel-based benchmark schemes rely on a computationally expensive exhaustive search to derive decision boundaries for IoT device authentication. Furthermore, methods such as SVM and hypothesis testing are required for continuous parameter updates due to the time-varying nature of the physical channel for device authentication.

3.6 Summary

In this work, we proposed a secure and efficient continuous authentication scheme for IoT devices. Our scheme utilised the grant-free NOMA protocol's transmission characteristics as a source for seed generation and device authentication. By utilising pre-arranged access time slots and spreading sequences of IoT devices at the AP, the proposed scheme eliminated the need for channel probing, seed reconciliation, and authentication. Simulation results demonstrated the effectiveness of the proposed scheme, with a near three-fold reduction in misdetection rate and close to zero false alarm rate in various system config-

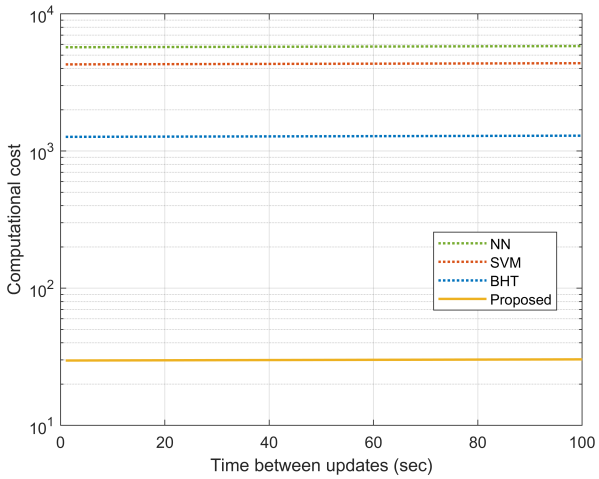


Figure 3.9: Computational cost versus the time between updates (sec), with the total number of potential devices $K = 200$, the number of resources $N = 100$, and the number of active devices $S = 20$.

urations. Additionally, our proposed scheme offered computational efficiency compared to benchmark schemes based on SVM and BHT utilising physical channel information, with at least half the computational cost.

Chapter 4

IoT Device Authentication in Non-Terrestrial Networks

4.1 Introduction

This chapter builds upon the traditional AKMA framework by tailoring it specifically for IoT networks operating over LEO satellites. To provide context, the AKMA framework is a protocol designed to ensure mutual authentication between devices and the network and provide secure key management. In its traditional implementation, AKMA operates by pre-registering devices with a central authentication server, where a shared secret key is established. This key serves as the basis for deriving session keys during authentication. When a device initiates a connection, it sends an authentication request containing cryptographic data derived from the shared key. The authentication server validates the request and responds with a session key that facilitates secure communication. This framework's design makes it well-suited for mobile devices but not for resource-constraint IoT devices since it assumes a centralized infrastructure and static terrestrial network, which limits its applicability in dynamic and decentralized environments like LEO satellite-based IoT networks.

Extending the authentication framework from terrestrial (Chapter 3) to non-terrestrial networks, this chapter tackles the unique challenges posed by LEO satellite-based IoT systems. While both frameworks share a reliance on physical layer security principles, the dynamic topology, high mobility, and latency-sensitive nature of LEO networks necessitate a fundamentally different approach. Building on the authentication scheme discussed in Chapter 3, this chapter proposes modifications to decentralize the process, adapt to frequent handovers, and mitigate vulnerabilities specific to satellite networks.

In this third technical chapter, we introduce an enhanced version of the AKMA authentication framework, tailored explicitly for use within LEO satellite-based IoT networks. Our proposed framework provides robust authentication for legitimate communications by employing pre-arranged access time slots for each IoT device, significantly increasing the complexity for potential spoofers. Unlike traditional approaches, our framework facilitates these access time slots based on mutual agreements between IoT devices and the corresponding serving satellite without necessitating additional hardware—leveraging the AKMA key and codebook matrix for slot generation. This setup ensures unpredictability for adversaries [2] and supports scalable deployments in massive IoT environments. The generation of access time slots occurs independently at both the serving satellite and IoT devices, allowing for seamless operation. Any discrepancy in the access time slots between an IoT device and the satellite is detected by the satellite, which then flags the device as illegitimate. *To our best knowledge, this is the first work to adapt the AKMA authentication framework for concurrent authentication across multiple IoT devices, utilising the AKMA key and codebook matrix as the foundational authentication mechanism.*

The rest of this chapter is organised as follows. In Section 4.2, we review the related studies of authentication schemes for LEO satellite-based IoT networks. In Section 4.3, we present the system model and the authentication problem. In Section 4.4, we describe the proposed authentication scheme and provide a detailed description of the different phases of device authentication. Finally, in Section 4.5, we present the simulation results to verify the performance gain of the proposed technique.

4.2 System Model

In this work, we consider a LEO satellite-based IoT network. We assume that K IoT devices are in a remote location and do not have access to a fixed AP in their vicinity. The LEO satellite network provides coverage to IoT devices. Each satellite has a footprint/coverage area on the ground. IoT devices talk to the satellite, which offers the best coverage. Here, without the loss of generality, we consider one such satellite providing coverage to K IoT devices, as illustrated in Fig 4.1[1]. We assume that IoT devices have limited computing and battery capabilities and rely on intermittent transmission to conserve resources. During communication sessions, a subset M of the K devices in-

[1]Although this work primarily focuses on a single satellite scenario, the proposed authentication scheme can extend to multiple satellites in a constellation by leveraging inter-satellite communication protocols to maintain continuous authentication state across satellite handovers. This aspect is outside the current scope of this work and will be the subject of future work.

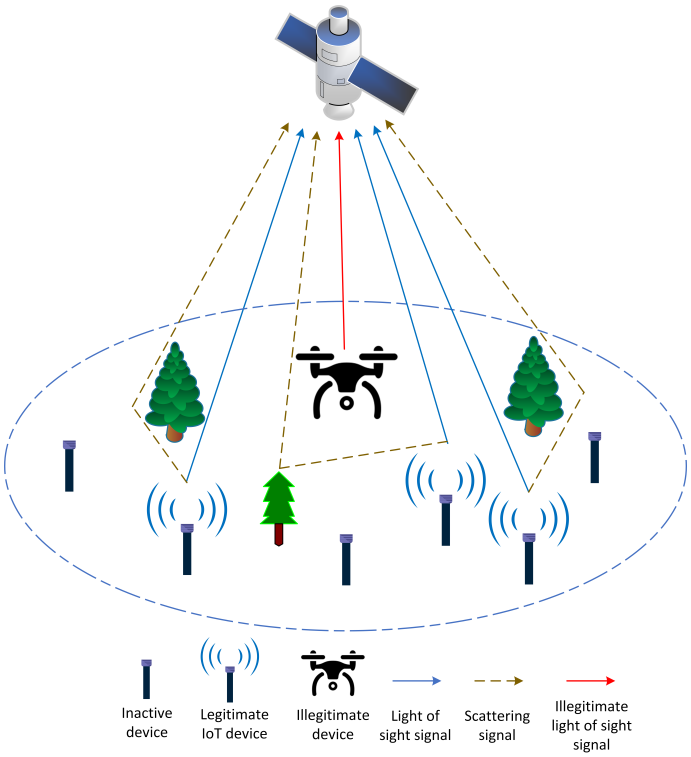


Figure 4.1: Illustration of the system model for LEO satellite-based IoT network. The transmission from the IoT devices to the serving satellite includes a line-of-sight and a scattering component. The transmission from the UAV to the serving satellite is line-of-sight.

termittently becomes active to transmit data, employing the irregular repetition slotted ALOHA (IRSA) protocol [138] for medium access, which accommodates the sporadic nature of IoT device activity. We consider an overloaded system where the number of resource blocks N is less than the number of IoT devices in a cell, i.e., $N < K$.

4.2.1 Threat Model

We assume the presence of an unauthorised device in the satellite's coverage area. Further, we assume that this unauthorised device is a UAV situated above the legitimate IoT devices, leading to a correlated physical channel environment. Consequently, the satellite might concurrently receive transmissions from legitimate IoT devices and illegitimate UAV devices. This illegitimate device may attempt unauthorised network access through sophisticated strategies like man-in-the-middle and spoofing attacks. We assume that the illegitimate device possesses superior computational resources compared to the

Table 4.1: Important symbols used in this paper.

Variable	Description
K	Total number of IoT devices
N	Total subcarriers
M	Active number of IoT devices
J	Number of time slots
$\mathbf{c}$	Spreading sequence
$\mathbf{h}$	Channel
$\mathbf{x}$	Transmit signal
$\mathbf{w}$	Gaussian noise
$\mathbf{y}$	Received signal
$\mathbf{G}$	Synthesis of channel vector and spreading sequences
$\mathbf{H}$	Channel matrix
$\mathbf{C}$	Codebook matrix
$\mathbf{X}$	Transmit signal (continuous time slots)
$\mathbf{Y}$	Received signal (continuous time slots)
$\overline{\mathbf{\Gamma}}$	Authenticated devices' indicator
l	Duplicate replicas of packet
K_d	AKMA key
K'_d	Modified AKMA key
ω	Length of slots within a frame
n	Size of the codebook matrix

legitimate IoT devices, i.e., it adopts identical system parameters and upper layer signalling as legitimate IoT devices, as detailed in Table 4.1, and also maintains persistent network surveillance to discern transmission patterns of legitimate IoT devices.

4.2.2 Signal Model

Considering an arbitrary symbol interval, an IoT device randomly wakes up and transmits its complex modulated signal toward the satellite, which is independent random variables drawn from a standard symmetric discrete constellation set. After modulation, the transmitted symbol x_k from the k-th IoT device is spread onto a spreading sequence $\mathbf{c}_k$ of length N. The received signal y on the n-th subcarrier at the satellite is given as

$$y_n = \sum_{k=1}^{K} h_{nk} c_{nk} x_k + w_n, \tag{4.1}$$

where h_{nk} refers to the channel gain of the k-th IoT device's n-th subcarrier, c_{nk} refers to the n-th component of the spreading sequence $\mathbf{c}_k$, and w_n is the Gaussian noise on the n-th subcarrier with zero mean and variance σ^2.

By combining the received signals over all N subcarriers, the received signal vector $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T \in \mathbb{C}^{N\times 1}$ at the LEO satellite is given as

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{w}, \tag{4.2}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_K]^T \in \mathbb{C}^{K\times 1}$ is the transmitted signal vector for all K devices and $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T \in \mathbb{C}^{N\times 1}$ is the noise vector. $\mathbf{G} \in \mathbb{C}^{N\times K}$ is the synthesis of the channel vectors and spreading sequences, given as

$$\mathbf{G} = \mathbf{H} \odot \mathbf{C}, \tag{4.3}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N\times K}$ is the channel matrix, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K] \in \mathbb{C}^{N\times K}$ is the codebook matrix, and $\odot$ is the Hadamard product, i.e., $g_{nk} = h_{nk}c_{nk}$ [1].

4.2.3 Channel Model

We adopt the state-of-the-art line-of-sight based satellite channel model in [139–141]. The channel h_{nk} of the k-th IoT device on the n-th subcarrier to the LEO satellite is modelled as

$$h_{nk} = h_{\text{ray}} e^{j\phi_{\text{ray}}} + h_{\text{los}} e^{j\phi_{\text{los}}}, \tag{4.4}$$

where the channel between the k-th IoT device and the associated satellite encompasses two primary components: a scattering component and an LOS component. The scattering component's amplitude, h_{ray}, adheres to a Rayleigh distribution while its phase, ϕ_{ray}, is uniformly distributed within the range $[-\pi, \pi]$. In contrast, the LOS component is characterised by amplitude, h_{los}, that follows a Nakagami-m distribution indicative of a quasi-static channel with a specified m-factor, and its phase ϕ_{los} remains constant. It is pertinent to mention that the variability in the amplitude h_{los} predominantly stems from shadowing effects, as elaborated in [141].

Further, the free-space path loss is used to estimate the loss of signal strength in the free-space environment since, in satellite communications, the majority of the signal path is unobstructed. Accordingly, the path loss between the k-th IoT device and the serving satellite is represented by $20\log_{10}(d_i) + 20\log_{10}(f) - 147.55$, where d_i denotes the distance in kilometres and f represents the centre frequency. Additionally, due to the higher orbital velocities of LEO satellites, we introduce a Doppler frequency shift (DFS)

in the form of phase frequency offset, represented by $f_d = v \times f/c$, where v is the velocity of the satellite, c is the speed of light.

4.2.4 Medium Access Model

Due to their inherent stochastic transmission patterns, IoT devices employ IRSA as their medium access protocol for satellite communications [138]. Within this framework, a k-th IoT device will dispatch l duplicate replicas of its packet in a single frame. To further elucidate, each replica's header embeds information regarding the positions of its counterparts. If a replica undergoes successful decoding, it becomes feasible to reconstruct and subsequently subtract its associated replicas from the frame.

With these considerations in mind, it becomes pertinent to augment the signal model, as portrayed in (4.2), transforming it from a single time slot transmission paradigm to one that encompasses continuous time slots. Accordingly, the transmitted signals $\mathbf{X} = \left[\mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \ldots, \mathbf{x}^{[J]}\right] \in \mathbb{C}^{K \times J}$ are recovered from the received signals $\mathbf{Y} = \left[\mathbf{y}^{[1]}, \mathbf{y}^{[2]}, \ldots, \mathbf{y}^{[J]}\right] \in \mathbb{C}^{N \times J}$ in J continuous time slots. Thus, the continuous time-slots transmission model for the j-th time slot is given as

$$\mathbf{y}^{[j]} = \mathbf{G}^{[j]}\mathbf{x}^{[j]} + \mathbf{w}^{[j]}, \quad j = 1, 2, \ldots, J, \tag{4.5}$$

where $\mathbf{G}^{[j]} \in \mathbb{C}^{N \times K}$ is the synthesis of the channel vectors and spreading sequences in the j-th time slot and $\mathbf{w}^{[j]}$ is the equivalent Gaussian noise vector in the j-th time slot.

4.2.5 Problem Statement

In the context of IoT devices engaging in communication with LEO satellite networks, the intermittent communication patterns of IoT devices provide a window for unauthorised entities to masquerade as legitimate devices, thereby compromising the integrity of the satellite link and infiltrating the core network. When an IoT device initiates a transmission to a satellite in the j-th time slot, the primary task of the satellite network is to verify the message's authenticity, ensuring it originates from a verified IoT device. This necessitates a predefined consensus between the satellite network and the authentic IoT devices on specific signal characteristics or transceiver attributes that are instrumental in differentiating legitimate devices from illegitimate devices. Let $\mathbf{\Gamma}^{[j]}$ symbolise the authentication status of devices in the j-th time slot, where the authentication challenge

is mathematically formulated as

$$\mathbf{\Gamma}^{[j]} = \begin{cases} 1 & \text{if } \mathcal{H}_0 \\ 0 & \text{if } \mathcal{H}_1 \end{cases}, \tag{4.6}$$

with $\mathcal{H}_0$ and $\mathcal{H}_1$ denoting the hypothesis that the received signal $\mathbf{y}^{[j]}$ in the j-th time slot originates from a legitimate or an illegitimate IoT device, respectively. Traditional methodologies [83, 120, 121] depend on quantisation thresholds for authentication decisions. Nevertheless, the accuracy of these methods substantially decreases due to quantisation inaccuracies, and pinpointing optimal detection thresholds for persistent authentication across an extensive array of IoT devices is impractical due to the exhaustive search required to ascertain these thresholds.

Moreover, the reliance of these traditional approaches on processes such as seed generation, verification, reconciliation, and authentication of IoT devices [80, 81, 122], while effective within terrestrial 5G infrastructures, poses significant challenges when applied directly to non-terrestrial environments, necessitating a reevaluation towards a more decentralised authentication strategy. The challenges include:

- The continuous authentication protocol employed by AKMA conflicts with the energy conservation strategies of IoT devices, which are designed to minimise transmission frequency to conserve power. This leads to increased energy consumption and reduced device longevity.

- The centralised nature of AKMA amplifies security and privacy risks, particularly the susceptibility to man-in-the-middle attacks during the key exchange process, thereby compromising the security integrity of the network.

- Lastly, the static framework of AKMA struggles to adapt to the dynamic topologies characteristic of LEO satellite networks, potentially leading to authentication bottlenecks and service interruptions.

- Implementing AKMA across LEO satellite networks entails significant complexity, driven by the infrastructure requirements of a centralised authentication system. Decentralised methods could alleviate these concerns by utilising network-based metrics for authentication, thereby simplifying infrastructure needs.

These considerations conclude that the AKMA framework, while robust within its intended terrestrial scope, is ill-suited for the distinctive operational paradigms of LEO

satellite-based IoT networks. Therefore, exploring and developing innovative, decentralised authentication methods becomes essential. Such methods should be specifically tailored to meet the unique requirements of LEO satellite environments, ensuring scalable and secure device communication.

4.3 Proposed Modified AKMA Framework

At the satellite, the primary goal is to authenticate IoT devices using the received signal $\mathbf{y}^{[j]}$. To achieve this robust mutual authentication, the transceiver pair must generate a unique seed, undisclosed to external devices. This seed must be periodically refreshed to ensure its continued validity. Finally, each transceiver pair must manage its authentication process in a decentralised manner. Such an approach guarantees that a legitimate IoT device can integrate into the network during any satellite handover without impacting the authentication status of other devices already in the network. Accordingly, our proposed authentication framework unfolds in two distinct steps.

1. **Seed Generation:** A concatenation of diversified functions is employed to compute a unique seed value for the initial and subsequent transmission of IoT devices.
2. **Transmission Pattern Generation:** Utilising the seed value derived in the previous step, a specific transmission pattern is generated. This pattern instructs IoT devices on their transmission when communicating with satellites.

It is pivotal to note that, unlike terrestrial AKMA, the seed calculation and transmission pattern generation processes proposed in our modified AKMA framework are independently executed at the satellite and the IoT devices. Such independence ensures system robustness, significantly mitigating potential false alarms. This interaction of the proposed authentication scheme is illustrated in Fig. 4.2. The steps are described in detail below.

4.3.1 Seed Generation

The proposed seed generation mechanism has three parts.

4.3.1.1 Initial Slot Selection

An IoT device randomly transmits l packets in a frame to minimise packet collision and maximise throughput. We use this randomness to our advantage by using it as a source

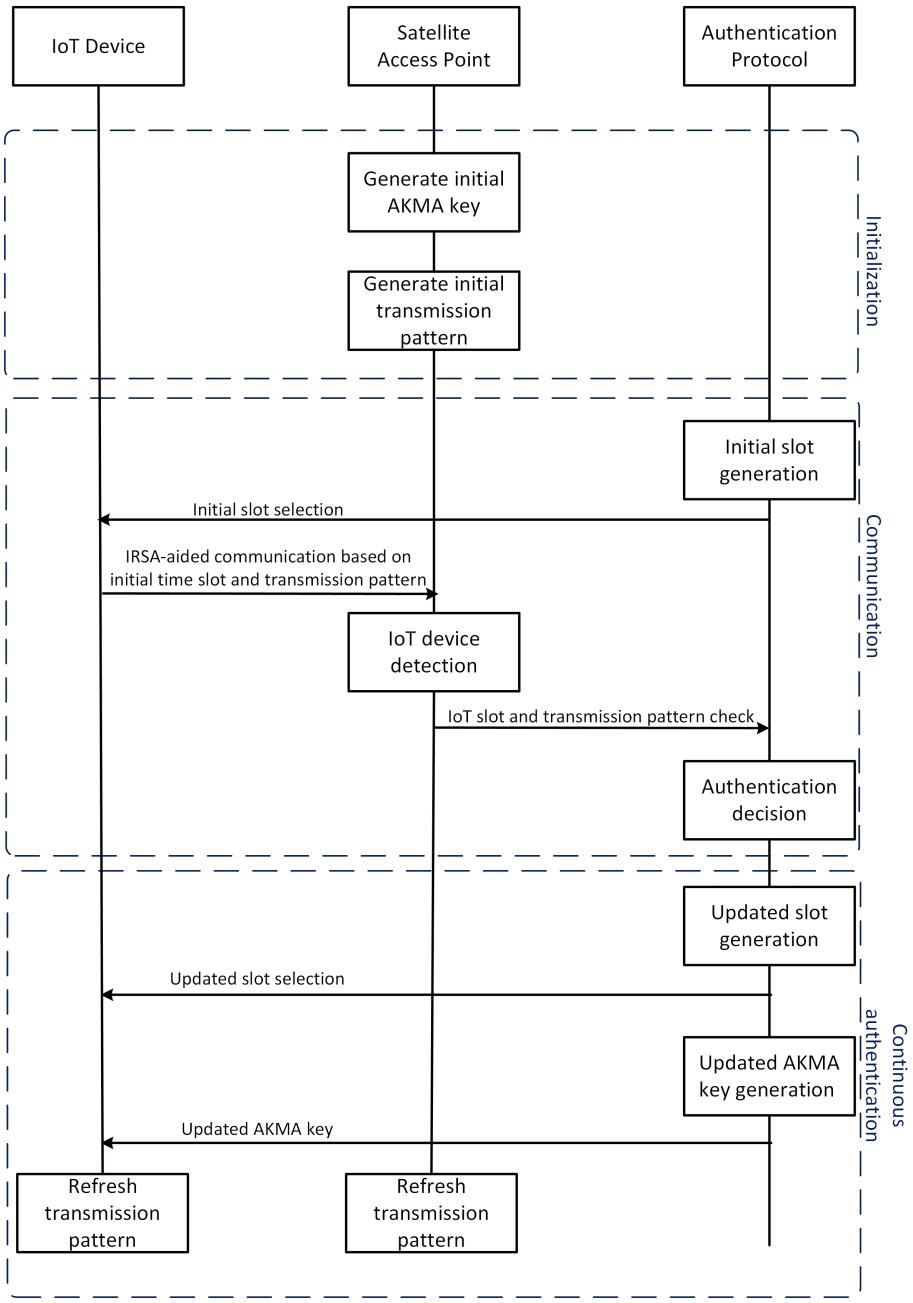


Figure 4.2: Flowchart of proposed authentication framework and its interaction with IRSA-based transmission model considered in this work.

for seed generation. Accordingly, the primary objective of this phase is to derive an initial seed that determines the starting slot in a communication frame for data packet transmission. The mathematical representation for this is given by:

$$S_{\text{initial}} = f_1(K_d, \omega, \mathbf{C}), \tag{4.7}$$

where $\mathbf{C}$ is the codebook matrix in (4.3) and the function f_1 is defined as:

$$f_1(K_d, \omega, n) = (K_d \oplus \mathfrak{H}(\omega)) \times n, \tag{4.8}$$

where K_d denotes the unique AKMA key, ω specifies the length of slots within a frame, $\mathfrak{H}(\cdot)$ represents a hashing function, and n indicates the size of the spreading sequences codebook matrix. The initial slot selection function, f_1, determines the seed by performing an XOR operation between K_d and the hash of the slot size. This result is then multiplied with n to ensure that the derived slot number, S_{initial}, remains within the valid slot range.

4.3.1.2 Slot Update Mechanism

To maintain continuous authentication, the IoTs and satellite must refresh the seed independently. Accordingly, this proposed procedure computes the seed for frames subsequent to the initial one, considering the replicated packets dispatched by an IoT device. It is expressed as:

$$S_{\text{next}} = f_2(K_d, S_{\text{initial}}, l, n), \tag{4.9}$$

where the function f_2 is represented as:

$$f_2 = (K_d \oplus S_{\text{initial}} \oplus l) \times n. \tag{4.10}$$

Herein, S_{initial} is initially determined by f_1, but for subsequent iterations, its value is derived from S_{next}. The term l enumerates the replicated packets transmitted by an IoT device. The function f_2 updates the seed considering the AKMA key K_d, the present slot S, and the number of data packets. An XOR operation is executed on these entities to compute the next seed.

4.3.1.3 AKMA Key Update Mechanism

The IoT devices waking up after a prolonged sleep schedule are required to refresh the AKMA key K_d. Accordingly, this phase is crucial for refreshing the AKMA key by

considering the replicated packet count and the preceding slot value ascertained from f_2. This is mathematically represented as:

$$K'_d = f_3(K_d, l, S_{\text{last}}), \tag{4.11}$$

where f_3 is given by:

$$f_3(K_d, l, S_{\text{last}}) = \mathfrak{H}(K_d \oplus l \oplus S_{\text{last}}). \tag{4.12}$$

Herein, K_d represents the initial AKMA key, which in subsequent iterations is replaced by its latest version K'_d. The term l is the number of replicated packets from the IoT device, while S_{last} denotes the final value from f_2. The function f_3 delineates a strategy for AKMA key updating. This is achieved by executing an XOR operation on K_d, the packet count, and the last slot number S_{last}. Subsequently, the outcome of this operation undergoes hashing to yield the refreshed key K'_d.

4.3.2 Transmission Pattern

The time slots for transmitting data from IoT devices are divided into fixed-length recurring intervals [87]. These transmission patterns are agreed upon in advance between the IoT devices and the satellite. Consequently, the satellite can quickly identify unauthorised devices based on their assigned transmission patterns. Transmission patterns eliminate the need for a shared secret key for every transmission, as the transmission schedules are followed by the IoT devices and verified by the satellite.

The transmission patterns are generated using non-linear feedback shift registers (NFSRs), which offer an enhanced level of unpredictability that is especially suited for IoT devices communicating with LEO satellites in uplink scenarios. The patterns produced by NFSRs exhibit non-linear dependencies on previous states, which makes them inherently more chaotic and challenging to predict than sequences produced by linear feedback shift registers [125]. This non-linear behaviour avoids computational burdens and presents an efficient framework for IoT applications. For an NFSR of length μ, a non-linear feedback function determines the next state of the NFSR. The generating non-linear monic polynomial for a generic variable $\varkappa$ is given as

$$\begin{aligned} \gamma(\varkappa) =& C_0 + C_1\varkappa + C_2\varkappa^2 + \ldots + C_\mu\varkappa^\mu + D\sin(\beta\varkappa^\alpha) \\ =& \sum_{i=1}^{i=\mu} C_i\varkappa^i + D\sin(\beta\varkappa^\alpha), \quad (C_0, C_\mu = 1,\ D, \beta, \alpha \in \mathbb{R}) \end{aligned} \tag{4.13}$$

where $C_0, C_\mu, D, \beta, \alpha$ are various constants with C_0, C_μ set to 1 to ensure the polynomial

is monic, and the $\sin(\cdot)$ function adds a non-linear component to the polynomial. The precise nature of this function defines the complexity and unpredictability of the sequence produced by the NFSR.

For unauthorised entities, predicting the non-linear feedback function f is challenging. This complexity is accentuated when the satellite and IoT devices have the flexibility to modify the function periodically, thus enhancing the security of transmissions. The creation of transmission patterns using the NFSR is synchronised and decentralised at the satellite and the k-th IoT device, ensuring the consistent and secure relay of information. By leveraging shared identical seeds, both entities can avoid repeated hash function operations for seed security for every transmission, underlining the efficiency and simplicity of the proposed authentication framework.

4.3.3 Authentication Decision

In summary, the proposed authentication framework does not rely on centralised key refreshment for IoT device authentication. Instead, the IoT devices and satellites independently generate a seed from the system network parameters. Subsequently, the seed generates transmission patterns for IoT device transmission and authentication. This authentication process is summarised in Algorithm 1, and the main procedure is presented as follows.

- *Initialisation Requirements:* The algorithm initialises with the requisite inputs: a device-specific secret AKMA key K_d, the length of slots within a frame ω, the size of the spreading sequences codebook matrix n, and a counter l for enumerating the replicated packets transmitted by an IoT device.

- *Objective:* The aim is to ensure secure and authenticated data transmission between the IoT devices and the LEO satellite by mitigating risks associated with centralised key management systems, thereby enhancing communications security in LEO satellite IoT networks.

- *Function f_1:* Generates the initial seed S_{initial} for data transmission through the XOR operation between K_d and the hash of ω, followed by multiplication with n. This seed is crucial in determining the starting slot of the data transmission, particularly after an IoT device becomes active.

- *Function f_2:* Produces the subsequent seed

Algorithm 3 The proposed authentication framework.

Input K_d, ω, n, l.
Output Secure and authenticated data transmission for IoT devices.

function $f_1(K_d, \omega, n)$
    $S_{\text{initial}} \leftarrow (K_d \oplus \mathfrak{H}(\omega)) \cdot n$
    return S_{initial}
function $f_2(K_d, S_{\text{initial}}, l, n)$
    $S_{\text{next}} \leftarrow (K_d \oplus S_{\text{initial}} \oplus l) \cdot n$
    return S_{next}
function $f_3(K_d, S_{\text{last}}, l)$
    $K'_d \leftarrow \mathfrak{H}(K_d \oplus l \oplus S_{\text{last}})$
    return K'_d

procedure TransmitData($\mathbf{x}$)
    if IoT device wakes up with data $\mathbf{x}$ **then**
        if First transmission after wake-up **then**
            $S_{\text{seed}} \leftarrow f_1(K_d, \omega, n)$
        else
            $S_{\text{seed}} \leftarrow f_2(K_d, S_{\text{initial}}, l, n)$
        $\gamma_k(\varkappa) = \sum_{i=1}^{i=\mu} C_i \varkappa^i + D \sin(\beta \varkappa^\alpha)$
        Transmit data $\mathbf{x}$ using transmission pattern γ_k
        if Prolonged sleep detected **then**
            $K'_d \leftarrow f_3(\mathfrak{H}(K_d \oplus l \oplus S_{\text{last}}))$
        else
            $K'_d \leftarrow K_d$

procedure ReceiveData($\mathbf{y}$)
    for $j = 1$ to J
        if $\gamma_k^{[j]}$[satellite] == $\gamma_k^{[j]}$[device] **then**
            $\mathbf{\Gamma}_k^{[j]} = 1.$
        else
            $\mathbf{\Gamma}_k^{[j]} = 0.$

S_{next} utilising K_d, S_{initial}, the counter l, and the nonce n. The incorporation of l ensures the uniqueness of each subsequent seed by refreshing it independently and bolstering security against replay attacks.

- *Function f_3:* Updates the device-specific secret key to K'_d through hashing the XOR combination of K_d, l, and the last seed S_{last}. This dynamic independent AKMA key upgrade further secures the transmission process against potential man-in-the-middle attacks.

- *Procedure* `TransmitData`*:* Details the process for an IoT device to transmit data $\mathbf{x}$, determining the appropriate seed S_{seed} for use—either through f_1 or f_2. The IoT device employs a specific transmission pattern $\gamma_k(\varkappa)$ for transmitting $\mathbf{x}$ based on the generated seed. Upon detection of prolonged inactivity, the AKMA key K_d is updated using f_3; otherwise, it remains unchanged.

- *Procedure* `ReceiveData`*:* Details the process for a LEO satellite to authenticate the data from IoT devices by comparing the received and expected transmission patterns. A match signals successful authentication ($\mathbf{\Gamma}_k^{[j]} = 1$), while a mismatch indicates failure ($\mathbf{\Gamma}_k^{[j]} = 0$).

At the end of the iteration, the authenticated devices data, corresponded by $\mathbf{\Gamma}_k^{[j]} = 1$, in the j-th time slot is transformed into a sparse vector, where the data of the illegitimate devices is replaced with 0's, whereas the authenticated devices data is recovered.

4.4 Security Analysis

The proposed authentication scheme, a novel adaptation of the AKMA framework for LEO satellite-based IoT networks, showcases significant advancements in authentication security. This section examines the scheme's resilience against prevalent threats, such as MITM attacks and unauthorized access.

4.4.1 Mitigation of MITM Attacks

The MITM attack can be modelled as a search space Ω problem for the insertion of an unauthorized node. In this formulation, the adversary's success probability corresponds directly to the size of the search space they must traverse. By carefully designing the proposed authentication framework, we ensure that the search space size remains computationally infeasible. To maintain this level of security, the uniformity of the search

space is paramount, ensuring that no part of the space is more likely to contain a solution than another.

To quantify the security offered by our proposed authentication scheme, we analyze the search space for each component: Initial Slot Selection, Slot Update Mechanism, AKMA Key Update Mechanism, and Transmission Pattern Generation. Each mechanism provides additional cryptographic complexity that helps secure the authentication process.

4.4.1.1 Initial Slot Selection:

In this stage, each IoT device computes an initial slot value S_{initial} using the function f_1 that depends on the AKMA key K_d, slot length ω, and spreading sequence $\mathbf{C}$. For an adversary to determine S_{initial} without prior knowledge of K_d and $H(\omega)$, the search space size is given by

$$\Omega_{\text{Initial}} = 2^{\kappa+b}, \tag{4.14}$$

where κ is the bit-length of K_d, and b is the bit-length of the hash $H(\omega)$. Assuming $\kappa = 128$ and $b = 128$, the effective search space size becomes

$$\Omega_{\text{Initial}} = 2^{256}. \tag{4.15}$$

4.4.1.2 Slot Update Mechanism:

The slot is periodically updated with the function f_2, which depends on K_d, S_{initial}, the number of packet replicas l, and codebook matrix size n. The search space for determining S_{next} is given by

$$\Omega_{\text{Slot Update}} = 2^{\kappa+l+n}. \tag{4.16}$$

For typical values of $\kappa = 128$, $l = 3$, and $n = 200$, the search space size becomes:

$$\Omega_{\text{Slot Update}} = 2^{331}. \tag{4.17}$$

4.4.1.3 AKMA Key Update Mechanism

The AKMA key K_d is periodically updated to prevent long-term exploitation by adversaries. The adversary's task of predicting K_d' involves traversing a search space size given as

$$\Omega_{\text{AKMA Update}} = 2^{\kappa+l}. \tag{4.18}$$

Substituting typical values of $\kappa = 128$ and $l = 3$, the search space size becomes:

$$\Omega_{\text{AKMA Update}} = 2^{131}. \tag{4.19}$$

4.4.1.4 Transmission Pattern Generation

The transmission pattern, generated using an NFSR polynomial, creates another independent security layer. The size of the search space for replicating the transmission pattern is given by

$$\Omega_{\text{Pattern}} = 2^{\mu}, \tag{4.20}$$

where μ represents the bit-length of the polynomial. For $\mu = 128$, the resulting search space size is given as

$$\Omega_{\text{Pattern}} = 2^{128}. \tag{4.21}$$

To successfully execute an MITM attack, an adversary must traverse the combined search space of all four components: Initial Slot Selection, Slot Update, AKMA Key Update, and Transmission Pattern replication. The total search space size is thus the product of the search space of the individual components, given as

$$\Omega_{\text{MITM}} = \Omega_{\text{Initial}} \cdot \Omega_{\text{Slot Update}} \cdot \Omega_{\text{AKMA Update}} \cdot \Omega_{\text{Pattern}}. \tag{4.22}$$

Substituting the values gives

$$\Omega_{\text{MITM}} = 2^{256} \cdot 2^{331} \cdot 2^{131} \cdot 2^{128} = 2^{846}. \tag{4.23}$$

Given the enormity of this search space, the attack is computationally infeasible for classical adversaries. Even with Grover's quantum search algorithm [**?**], which reduces the effective search space to its square root, the size remains prohibitively large, given as

$$\Omega_{\text{MITM}} = \sqrt{2^{846}} = 2^{423}. \tag{4.24}$$

The multi-layered approach of the proposed authentication scheme introduces a series of independent probabilistic barriers, effectively transforming the MITM attack into a computationally infeasible search problem. The overall complexity of the search space, coupled with the uniformity of its distribution, ensures robust security against classical and quantum adversaries alike.

4.4.2 Prevention of Unauthorized Access

The framework's ability to detect discrepancies in transmission patterns between IoT devices and satellites is a robust mechanism to prevent unauthorized access. Any deviation in the pre-arranged transmission slots, independently verified at both ends, immediately flags the device as illegitimate. This approach ensures that only devices that can correctly generate and follow the agreed-upon transmission patterns are authenticated and allowed access to the network.

Moreover, the periodic refreshment of the AKMA key K_d based on the number of packets transmitted l and the last seed value S_{last} ensures that the keys remain current and are difficult for unauthorized entities to predict. This continuous update mechanism ensures that even if an attacker gains temporary access, their ability to maintain unauthorized access over time is severely limited.

The decentralized and continuous authentication framework tailored for LEO satellite-based IoT networks effectively tackles several critical security challenges. By leveraging localized key refreshment, independent seed generation, and unique transmission pattern generation, the proposed authentication scheme fortifies the security of IoT device authentication. It ensures scalable and efficient authentication ideally suited for LEO satellite networks' dynamic and high-velocity environment.

4.5 Results and Discussion

In this section, we assess the performance of the proposed authentication framework. We compare the performance of our proposed authentication framework with physical channel-based authentication using binary hypothesis testing [84], support vector machine-aided detection [142], and with Diffie-Hellman (DH) key exchange protocol [143]. Together, these benchmarks provide a balanced and varied basis for evaluating the effectiveness of our proposed authentication scheme. While the core working of these benchmark solutions is adopted from their respective works, we have adjusted their configurations to suit our system model for a fair comparison. For the physical channel-based benchmark schemes, we utilize the estimates of the received signal strength indicator (RSSI) and DFS as attributes extracted from the physical channel to authenticate the IoT devices. Furthermore, we assume that the physical channels of the illegitimate UAV device are correlated with those of the legitimate IoT devices [2]. This is because the UAV is deployed at a height of 120 meters, which is much smaller than the height of the satellite.

Table 4.2: Access time slots generation using seed.

State	Observations
Spreading pool utilised between a transceiver pair	$\gamma = \{-104 - 104i, 8i, 39 + 39i, \ldots, -2 + 2i, 4\}$
Seed extracted by the satellite	1101010110010
Seed extracted by the IoT	1101010110010
Access time slots at the satellite and IoT	10000000111111110101001010101110111111110 01000100111111111010101011011101000000010 10110000010101111101111110111011011000 00010001101010100011010001000111111011100 01010101000001101111110001010000000000010

Assuming initial authentication between a transceiver pair in the j-th time slot, their observation characteristics are shown in Table 4.2. As detailed in section III, the serving satellite and IoT device independently extract the seed using the AKMA key, the length of slots within a frame, and the size of the spreading sequences codebook matrix. Since the seed source is extracted from these parameters independently and available with the transceiver pair locally, there is no requirement for seed verification. Therefore, once the seed is acquired, the serving satellite and IoT independently generate the access time slots required for transmission. In this work, we utilise the following non-linear monic polynomial for the access time slots generation

$$\gamma(\varkappa) = 1 + \varkappa^1 + \varkappa^3 + \sin(\varkappa^5). \tag{4.25}$$

4.5.1 Experimental Setup

In the 3GPP framework, six reference scenarios have been specified [144–146]. In this work, we consider scenario D2 in [145], where the serving satellite has a speed of 7.56 km/s with an altitude of 600 km from the IoT devices [147]. The serving satellites operate at 2.4 GHz frequency with a max receiver gain of 30 dBi, whereas the IoT devices utilise 0 dBi antenna gain. In the simulations, unless otherwise specified, we consider a scenario where $K = 1000$ potential devices share $N = 400$ resources simultaneously. For each time frame, we randomly select $M = 200$ active devices from the set $\{1, 2, \ldots, K\}$. Every active device transmits $l = 3$ replicas of the same packet. The total number of time slots is fixed at $J = 10$. To transmit the signals, the IoT devices use QPSK modulation. The SNR is varied within the range of 0 to 20 dB. For device detection, we employ the orthogonal matching pursuit-based detection [148]. The results are averaged over 1000

Monte Carlo trials.

The simulations are carried out on MATLAB 2021b running on the Gadi supercomputer of the National Computational Infrastructure (NCI), Australia, utilising 48 cores of Intel Xeon Platinum 8274 (Cascade Lake) processors and 250GB of random access memory. The results are averaged over 1000 Monte Carlo trials.

4.5.2 Performance Metrics

To appropriately evaluate the authentication performance, we use the following metrics: the authentication rate (ρ_{au}) and the misdetection rate (ρ_{md}) as performance metrics. Given the transmit signal $\mathbf{x}$ and the authenticated devices indicator $\mathbf{\Gamma}$ for the k-th IoT device in the j-th time slot, the performance metrics are defined as follows.

- Authentication rate: This metric evaluates the rate of legitimate IoT devices being accurately detected as legitimate devices, given as

$$\rho_{\text{au}} = \frac{1}{K} \sum_{k \in \mathbf{x}^{[j]}} P\left\{ \mathbf{\Gamma}_k^{[j]} = 1 \mid \mathbf{x}_k^{[j]} = 1 \right\}. \tag{4.26}$$

- Misdetection rate: This metric evaluates the rate of illegitimate IoT devices being misdetected, given as

$$\rho_{\text{md}} = \frac{1}{K} \sum_{k \in \mathbf{x}^{[j]}} P\left\{ \mathbf{\Gamma}_k^{[j]} = 1 \mid \mathbf{x}_k^{[j]} = 0 \right\}. \tag{4.27}$$

4.5.3 Authentication Performance:

Fig. 4.3 plots the authentication rate vs the SNR (dB) for a configuration employing parameters $K = 1000$, $N = 400$, and $M = 200$. We can see that the benchmark schemes relying on physical channel attributes perform poorly at the lower SNR range, mainly due to the correlation of physical channels between IoT devices and unauthorised devices. Additionally, the proposed authentication scheme also outperforms the DH method by achieving higher authentication rates due to DH's susceptibility to man-in-the-middle attacks and its reliance on the integrity of the physical communication channel. The proposed framework has the best performance for the entire SNR range. This is attributed to the decentralised seed generation process implemented both at the IoT devices and the serving satellite, resulting in the independent generation of transmission patterns. These transmission patterns enhance the robustness of our authentication framework,

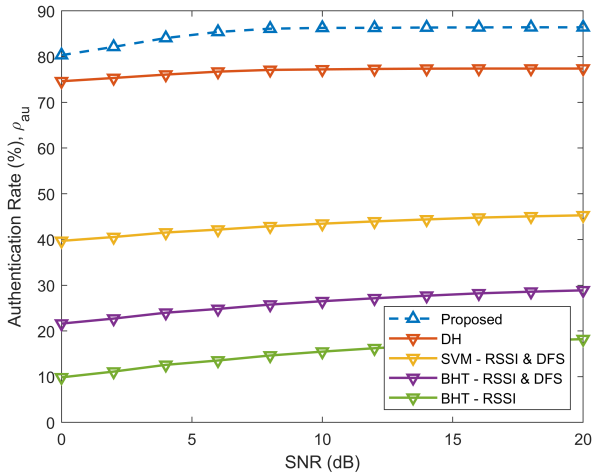


Figure 4.3: Authentication rate versus SNR (dB), with the total number of potential devices $K = 1000$, the number of resources $N = 400$, and the number of active devices $M = 200$.

particularly in the challenging noise-prone conditions typical of satellite wireless communications. Note that the authentication rate of satellite IoT networks is generally poorer than terrestrial networks (where the authentication rate can reach 100%) due to the greater susceptibility to the physical channel attenuation effects [149].

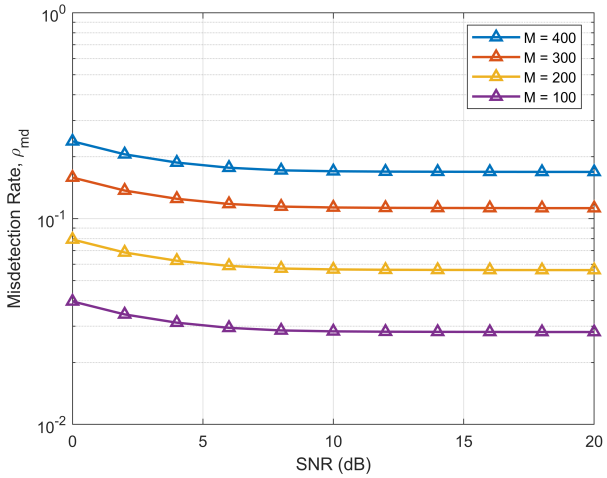


Figure 4.4: Misdetection rate versus SNR (dB) for the varying number of active devices M, with the total number of potential devices $K = 1000$, and the number of resources $N = 400$.

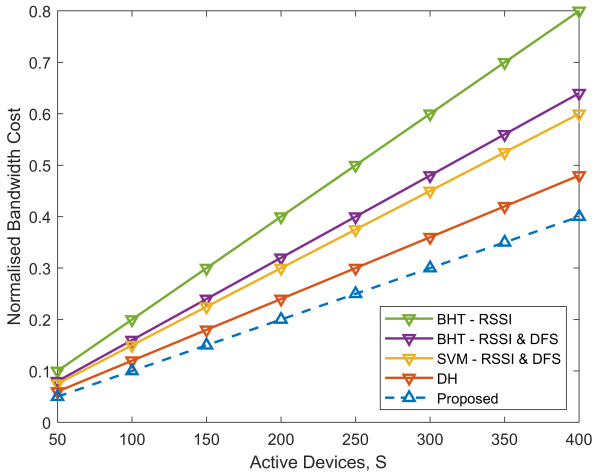


Figure 4.5: Bandwidth cost versus the increasing number of active devices M, with the total number of potential devices $K = 1000$, and the number of resources $N = 400$.

Fig. 4.4 plots the misdetection rate versus SNR (dB) for the varying number of active devices M, with $K = 1000$, and $N = 400$. The figure shows that the misdetection rate improves as the number of active devices M reduces. This is to be expected due to device estimation errors. Further, the results indicate the robustness of our proposed authentication framework in accommodating a diverse range of active devices M. This resilience is primarily attributed to the framework's independence from the traditional reliance on multiple attribute estimates from physical channels, which conventionally necessitates a recalibration of the decision boundary with any fluctuation in the number M of active devices. Instead, our framework harnesses transmission patterns derived from a decentralised seed generation process, endowing it with a scalable adaptability to any number of active transmitting devices M. However, it is crucial to recognise that the observed reduction in misdetection rate with increasing number of active devices M is primarily due to device estimation errors. Such errors are symptomatic of the systemic stresses encountered in scenarios of network overloading.

Fig. 4.5 plots the normalised bandwidth cost versus the increasing number of active devices M, with $K = 1000$ and $N = 400$. The proposed authentication scheme exhibits significantly lower bandwidth costs than the other benchmark schemes, which is attributed to its decentralised authentication strategy. This effectively minimises reconciliation bits by reducing seed mismatch scenarios, thus achieving more efficient bandwidth utilisation. In contrast, the DH method shows a higher bandwidth cost. This increase is due to the necessity for continuous key exchanges between IoT devices and

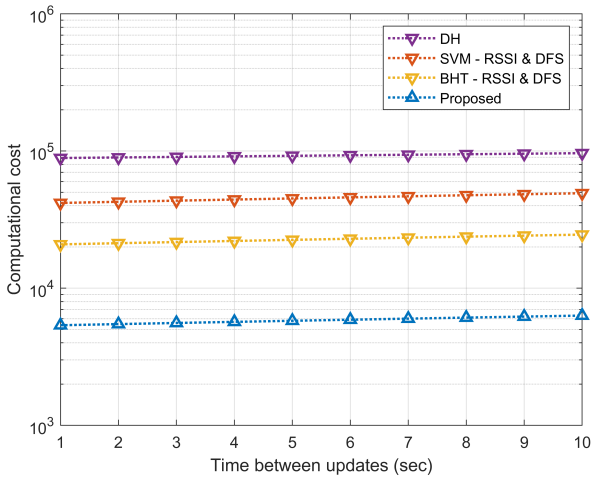


Figure 4.6: Computational cost versus the time between updates (sec), with the total number of potential devices $K = 1000$, the number of resources $N = 400$, and the number of active devices $M = 200$.

the LEO satellite prior to establishing a session, leading to a progressive rise in bandwidth consumption as the number of active devices increases. Furthermore, the physical channel-based authentication schemes incur the highest bandwidth costs. These methods are hampered by frequent seed mismatches resulting from the rapidly changing channel conditions between IoT devices and the LEO satellite. Consequently, substantial reconciliation bits are needed before authentication can be successfully completed, thereby elevating the overall bandwidth cost.

4.5.4 Reduced Computational Overhead:

Fig. 4.6 depicts the computational cost as a function of the update interval (seconds) for a configuration with parameters $K = 1000$, $N = 400$, and $M = 200$. Our proposed authentication framework significantly reduces computational overhead compared to the benchmark schemes. Our framework leverages transmission patterns for IoT device authentication and circumvents the necessity for estimating physical channel attributes—a process typically required in benchmark schemes. This strategy yields a notable decrease in computational demands. In contrast, our proposed authentication framework obviates the need to generate such decision thresholds, facilitating a more computationally efficient process. The reduction in computational cost underscores the practical advantages of our proposed authentication framework in scenarios where computational resources are limited or when rapid authentication is paramount.

4.5.5 Emulation

To further evaluate the performance of our proposed authentication scheme against benchmark schemes, we performed detailed emulations of a LEO satellite-based IoT communication scenario, incorporating realistic parameters to mirror practical deployment conditions based on parameters from the 3GPP technical reports [150–153]. Specifically, our emulated LEO satellite IoT testbed is developed using Matlab 2024b and consists of IoT devices communicating via a constellation of 600 LEO satellites, emulating an environment similar to contemporary satellite networks. Communication latency is anchored in a base link delay of 70 milliseconds, representative of the typical delay experienced with LEO satellites, and a link bandwidth of 10 Mbps, reflecting the data transmission rates expected in such settings. Data packets are set at a size of 256 bytes, aligning with common IoT communication protocols. Each satellite orbits at an altitude of 60 km with a velocity of 7.56 km/s, parameters chosen to approximate real LEO satellite dynamics. The coverage area per satellite is managed with a 20-degree beam width, offering practical coverage for the system. We further account for noise and interference factors, setting the noise power at -120 dBm and applying an interference factor of 0.1, which collectively emulates environmental and inter-device interference. The IoT devices maintain an initial power level of 150 mW, while communication incurs an energy cost of 10 mW, embodying the energy consumption per transmission attempt. To realistically model handoff scenarios, we set a signal strength threshold for satellite handoff at -100 dBm. IoT device processing power and memory constraints were also incorporated, with each device allocated a processing power limit of 100 million instructions per second (MIPS) and a memory limit of 512 KB, emulating the computational and memory demands during authentication processes, which consume 10 MIPS of processing power and 50 KB of memory per authentication. A weather attenuation factor of 3 dB accounts for signal degradation due to adverse weather. A slot transmission probability of 0.3 is chosen to reflect typical access patterns, with an additional penalty factor of 0.5 applied for IoT devices located in remote areas, where 30% of devices were emulated as situated in such regions. To model the Doppler effect, we used a carrier frequency of 2.4 GHz and the speed of light to assess signal shifts due to satellite motion, ensuring comprehensive alignment with real-world LEO satellite communication dynamics.

Fig. 4.7 illustrates the emulation for average authentication latency (ms) as a function of the time between updates (seconds) for a configuration with parameters $K = 1000$, $N = 400$, and $M = 200$. The proposed authentication framework exhibits significantly lower average authentication latency than the benchmark schemes. For example, at

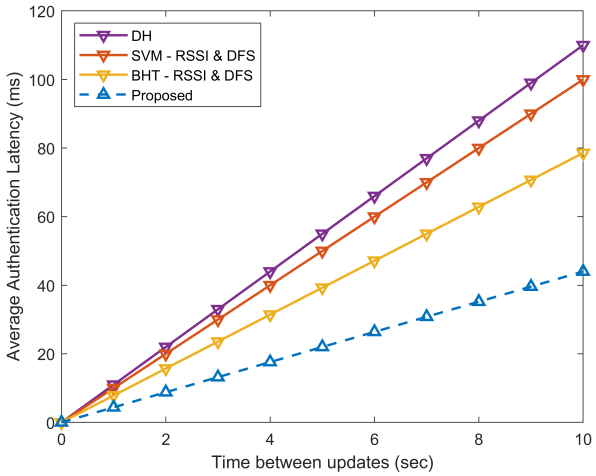


Figure 4.7: Average authentication latency versus the time between updates (sec), with the total number of potential devices $K = 1000$, the number of resources $N = 400$, and the number of active devices $M = 200$.

time between updates = 7 seconds, our proposed authentication scheme exhibits an 85% reduction in average authentication latency compared to DH. This reduced latency is attributed to the inherent efficiency of the proposed framework, which leverages independent seed generation and eliminates the need for estimating or reconciling physical channel attributes. On the other hand, the benchmark schemes require additional transmissions for channel reconciliation due to fast fading channel, leading to increased latency, particularly at longer update intervals. In contrast, our proposed framework avoids these reconciliation requirements, instead relying on independent seed generation, simplifying the authentication process and mitigating the communication overhead associated with multiple back-and-forth exchanges. Consequently, the authentication latency for the proposed scheme increases at a significantly lower rate as the update interval extends, highlighting its robustness and practicality for time-sensitive IoT applications.

4.6 Summary

This work proposed a modified AKMA framework for authentication in LEO satellite-based IoT networks. Our framework encompasses seed generation, seed update, and seed refreshment, all executed in a decentralised fashion. This facilitates tailored transmission patterns for IoT devices, significantly reducing the frequency of authentication interactions with the satellite. As a result, our method effectively counters the threats as-

sociated with the LEO satellite-based IoT networks. Through simulations and emulation, we have demonstrated a marked improvement in both authentication and misdetection rates, underscoring the potential of our framework to contribute to secure authentication measures in LEO satellite-based IoT networks.

Chapter 5

Conclusions and Future Research Directions

In this chapter, we first summarise the general conclusions drawn from the thesis. We then outline some of the research directions directly related to the thesis.

5.1 Summary of Key Findings of Thesis

This thesis focused on the design of providing security to IoT devices using the detect, identify and authentication framework. Specifically, this thesis investigated several challenges faced by IoT devices suffering from constrained resources and operating both in terrestrial and non-terrestrial networks. The detailed contributions are given as follows:

5.1.1 IoT Device Detection and Identification

In Chapter 1, we proposed an attention-based BiLSTM network for AUD in an uplink grant-free NOMA system by exploiting the temporal correlation of active user support sets. First, a BiLSTM network is used to create a pattern of the device activation history in its hidden layers, whereas the attention mechanism provides essential context to the device activation history pattern. Then, the complex spreading sequences are utilised for blind data detection without explicit channel estimation from the estimated active user support set. Thus, the proposed mechanism is efficient and does not depend on impractical assumptions, such as prior knowledge of active user sparsity or channel conditions. Through simulations, we demonstrated that the proposed mechanism outperforms several existing benchmark MUD algorithms and maintains lower computational complexity.

5.1.2 IoT Device Authentication in Terrestrial Network

In Chapter 2, we proposed a secure and efficient continuous authentication scheme for IoT devices. Our scheme utilised the grant-free NOMA protocol's transmission characteristics as a source for seed generation and device authentication. By utilising pre-arranged access time slots and spreading sequences of IoT devices at the AP, the proposed scheme eliminated the need for channel probing, seed reconciliation, and authentication. Simulation results demonstrated the effectiveness of the proposed scheme, with a near three-fold reduction in misdetection rate and close to zero false alarm rate in various system configurations. Additionally, our proposed scheme offered computational efficiency compared to benchmark schemes based on SVM and BHT utilising physical channel information, with at least half the computational cost.

5.1.3 IoT Device Authentication in Non-Terrestrial Network

In Chapter 3, we proposed a modified AKMA framework for authentication in LEO satellite-based IoT networks. Our framework encompasses seed generation, seed update, and seed refreshment, all executed in a decentralised fashion. This facilitates tailored transmission patterns for IoT devices, significantly reducing the frequency of authentication interactions with the satellite. As a result, our method effectively counters the threats associated with the LEO satellite-based IoT networks. Through simulations and emulation, we have demonstrated a marked improvement in both authentication and misdetection rates, underscoring the potential of our framework to contribute to secure authentication measures in LEO satellite-based IoT networks.

5.2 Future Research Directions

This section highlights potential research directions inspired by the findings of each research work presented in this thesis, which may serve as focal points for future research endeavours.

5.2.1 Research Work 1

In this work, we have applied the proposed framework to a spreading-based grant-free NOMA scheme. Future work can investigate whether the proposed framework can be generalised to other signature-based grant-free NOMA schemes. Additionally, incorporating advanced learning mechanisms, such as reinforcement or federated learning, can enhance the AUD process. Extending the framework to non-terrestrial networks, like

LEO satellite networks, and integrating it with physical layer security mechanisms can address unique challenges and enhance overall system security.

5.2.2 Research Work 2

Future research should explore the extended application of the proposed authentication scheme beyond its current context in signature-based grant-free NOMA schemes, examining its adaptability in various scenarios to understand its effectiveness in diverse wireless communication environments. Additionally, investigating the authentication scheme's implementation in satellite-IoT networks presents an exciting opportunity to address unique challenges related to vast coverage and long-distance communication, potentially unlocking secure and efficient communication in satellite-based IoT applications. To ensure real-world viability, a comprehensive security analysis is crucial, covering a wide range of potential attacks, including adversarial and resource exhaustion attacks, to identify weaknesses and develop robust authentication solutions for IoT devices. Additionally, a formal security analysis of the authentication scheme can be carried out to further understand its workability in different scenarios. Furthermore, scalability should be investigated to ensure efficient authentication, even in massive-scale deployments. By optimising the scheme without compromising security and addressing these research areas, the groundwork can be laid for secure, adaptive authentication solutions that bolster IoT device security and seamless integration into our interconnected world.

5.2.3 Research Work 3

Future work can investigate the real-world hardware implementation and performance testing of the proposed authentication scheme on appropriate LEO satellite IoT testbeds. Further, future work can expand this framework to operate seamlessly across satellite constellations. This will necessitate addressing challenges associated with inter-satellite handovers and beam switching to ensure continuous and secure authentication. Additionally, further research can explore integrating advanced cryptographic methods to enhance resilience against evolving security threats and implementing machine learning algorithms to predict and mitigate potential authentication failures in dynamic network conditions. Furthermore, the role and characteristic of the UAV to be able to adapt its position to impact the security performance as an illegitimate device can be considered as future work.

Bibliography

[1] S. Khan, S. Durrani, M. B. Shahab, S. J. Johnson, and S. Camtepe, "Joint user and data detection in grant-free NOMA with attention-based BiLSTM network," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1499–1515, Jul. 2023.

[2] S. Khan, C. Thapa, S. Durrani, and S. Camtepe, "Access-based lightweight physical layer authentication for the internet of things devices," *IEEE Internet Things J.*, vol. 11, no. 7, pp. 11 312–11 326, Apr. 2024.

[3] S. Khan, S. Durrani, C. Thapa, and S. Camtepe, "Modified AKMA for decentralized authentication in LEO satellite-based IoT networks," *submitted in IEEE Internet Things J.*, Jun. 2024.

[4] S. Ahmad, S. Khan, K. S. Khan, F. Naeem, and M. Tariq, "Resource allocation for IRS-assisted networks: A deep reinforcement learning approach," *IEEE Comms. Stand. Mag.*, vol. 7, no. 3, pp. 48–55, Sep. 2023.

[5] F. Naeem, G. Kaddoum, S. Khan, K. S. Khan, and N. Adam, "IRS-empowered 6G networks: Deployment strategies, performance optimization, and future research directions," *IEEE Access*, vol. 10, pp. 118 676–118 696, Nov. 2022.

[6] S. J. Siddiqi, F. Naeem, S. Khan, K. S. Khan, and M. Tariq, "Towards AI-enabled traffic management in multipath TCP: A survey," *Comput. Commun.*, vol. 181, pp. 412–427, Jan. 2022.

[7] S. Khan, C. Thapa, S. Durrani, and S. Camtepe, "Beyond key-based authentication: A novel continuous authentication paradigm for IoTs," in *Proc. IEEE GLOBECOM Wkshps*, Dec. 2023.

[8] S. Idrees, X. Jia, S. Khan, S. Durrani, and X. Zhou, "Deep learning based passive beamforming for IRS-assisted monostatic backscatter systems," in *Proc. IEEE (ICASSP)*, Apr. 2022, pp. 8652–8656.

[9] S. Khan, S. Durrani, and X. Zhou, "Transfer learning based detection for intelligent reflecting surface aided communications," in *Proc. IEEE (PIMRC)*, Sep. 2021, pp. 1–6.

[10] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[11] S. Li, L. D. Xu, and S. Zhao, "The internet of things: a survey," *Inf. Syst. Front.*, vol. 17, no. 2, pp. 243–259, Apr. 2015.

[12] E. Borgia, "The internet of things vision: Key features, applications and open issues," *Comput. Commun.*, vol. 54, pp. 1–31, Dec 2014.

[13] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[14] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, "Opportunistic IoT: exploring the social side of the internet of things," in *Proc. IEEE INFOCOM Wkshps.* IEEE, Apr. 2013, pp. 312–317.

[15] "Number of internet of things (IoT) connections worldwide from 2022 to 2023, with forecasts from 2024 to 2033," Jun. 2024. [Online]. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[16] "Worldwide spending on the internet of things is forecast to surpass $1 trillion in 2026," Jun. 2023. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prUS50936423

[17] "Internet of things - australia," Jun. 2024. [Online]. Available: https://www.statista.com/outlook/tmo/internet-of-things/australia

[18] S. Sicari, A. Rizzardi, L. Grieco, and A. Coen-Porisini, "Security, privacy and trust in internet of things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan 2015.

[19] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in internet of things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan. 2015.

[20] R. Ratasuk, N. Mangalvedhe, R. Bhatoolaul, and A. Ghosh, "LTE-M Evolution Towards 5G Massive IoT," in *Proc. IEEE GLOBECOM*, 2017, pp. 1–6.

[21] S. Chen and J. Zhao, "The Requirements, Challenges, and Technologies for 6G," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 36–42, 2020.

[22] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, 2019.

[23] V. Sharma, I. You, and M. Atiquzzaman, "Satellite Communication for 5G and Beyond: A Survey," *IEEE Access*, vol. 8, pp. 103 999–104 018, 2020.

[24] B. Di, H. Zhang, and L. Song, "Ultra-Dense LEO: Integration of Satellite Access Networks into 5G and Beyond," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 62–69, 2018.

[25] S. Pradhan and N. Sinha, "Security Challenges in Satellite Communication Networks," in *Proc. IoT-SIU*, 2019, pp. 1–6.

[26] L. Zhang, X. Zhang, and H. Guo, "Security and Privacy in Space Information Networks: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 2, pp. 1011–1033, 2019.

[27] L. Xiao, X. Lu, D. Xu, Y. Tang, R. Wang, and G. Zhu, "A survey of physical-layer security techniques for 5G wireless networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 679–695, May. 2018.

[28] Y. Zou, X. Wang, W. Shen, and H. Sun, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727–1765, Sep. 2016.

[29] J. Granjal, E. Monteiro, and J. S. Silva, "Security for the internet of things: A survey of existing protocols and open research issues," *IEEE Commun. Surv. Tutorials*, vol. 17, no. 3, pp. 1294–1312, Jul. 2015.

[30] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[31] M. A. Khan and K. Salah, "Iot security: Review, blockchain solutions, and open challenges," *Future Gener. Comput. Syst.*, vol. 82, pp. 395–411, May. 2018.

[32] R. Roman, J. Zhou, and J. Lopez, "Features and challenges of security and privacy in distributed internet of things," *Comput. Netw.*, vol. 57, no. 10, pp. 2266–2279, Apr. 2013.

[33] Y. Zhou, N. Zhang, and P. Li, "Security and privacy in smart cities: Challenges and opportunities," *IEEE Access*, vol. 7, pp. 2169–3536, Jun. 2019.

[34] M. Tran, A. Le, and H. Cheng, "Lightweight security algorithms for iot devices: A review," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2557–2567, Apr. 2019.

[35] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, "Secure and sustainable load balancing of edge data centers in fog computing," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 60–65, May. 2018.

[36] L.-B. Chen, W.-H. Lee, Y.-J. Chang, and K.-K. R. Choo, "Iot security: ongoing challenges and research opportunities," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 100 450–100 464, Oct. 2019.

[37] D. E. Kouicem, A. Bouabdallah, and H. Lakhlef, "Internet of things security: A top-down survey," *Comput. Netw.*, vol. 141, pp. 199–221, Jun. 2018.

[38] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[39] A. Mosenia and N. K. Jha, "A comprehensive survey of security issues in the internet of things," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1802–1831, Dec. 2017.

[40] N. Kshetri, "Can blockchain strengthen the internet of things?" *IT Prof.*, vol. 19, no. 4, pp. 68–72, Jul. 2017.

[41] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the internet of things: perspectives and challenges," *Wirel. Netw.*, vol. 20, no. 8, pp. 2481–2501, Nov. 2014.

[42] H. Suo, J. Wan, C. Zou, and J. Liu, "Security in the internet of things: A review," *Proc. IEEE*, vol. 3, no. 5, pp. 648–658, May. 2012.

[43] G. A. Fink and D. Zarzhitsky, "Security for the internet of things: A survey of existing protocols and open research issues," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 51–58, Nov. 2010.

[44] Y. Li and Y. Vorobeychik, "Learning from the past: Where cognitive security research has been and where it is going," *Proc. IEEE*, vol. 106, no. 3, pp. 357–370, Mar. 2018.

[45] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, May 2020.

[46] M. B. Shahab, S. J. Johnson, M. Shirvanimoghaddam, and M. Dohler, "Enabling transmission status detection in grant-free power domain non-orthogonal multiple access for massive internet of things," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 9, p. e4565, Jun. 2022.

[47] M. Mohammadkarimi, M. A. Raza, and O. A. Dobre, "Signature-based nonorthogonal massive multiple access for future wireless networks: Uplink massive connectivity for machine-type communications," *IEEE Veh. Technol. Mag.*, vol. 13, no. 4, pp. 40–50, Oct. 2018.

[48] Y. Shan, C. Peng, L. Lin, Z. Jianchi, and S. Xiaoming, "Uplink multiple access schemes for 5G: A survey," *ZTE Communications*, vol. 15, no. S1, pp. 31–40, Jun. 2017.

[49] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *Proc. IEEE (VTC-Spring)*, Jul. 2016, pp. 1–5.

[50] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 629–646, Oct. 2017.

[51] N. Docomo *et al.*, "Uplink multiple access schemes for NR," in *R1-165174, 3GPP TSG-RAN WG1 Meeting*, vol. 85, May 2016, pp. 1–4.

[52] J.-P. Hong, W. Choi, and B. D. Rao, "Sparsity controlled random multiple access with compressed sensing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 998–1010, Oct. 2014.

[53] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-things-based smart cities: Recent advances and challenges," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 16–24, 2017.

[54] A. Salari, M. Shirvanimoghaddam, M. B. Shahab, Y. Li, and S. Johnson, "NOMA joint channel estimation and signal detection using rotational invariant codes and GMM-based clustering," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2485–2489, Oct. 2022.

[55] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory.*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[56] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory.*, vol. 55, no. 5, pp. 2230–2249, Apr. 2009.

[57] A. T. Abebe and C. G. Kang, "Iterative order recursive least square estimation for exploiting frame-wise sparsity in compressive sensing-based MTC," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1018–1021, Mar. 2016.

[58] A. C. Cirik, N. M. Balasubramanya, and L. Lampe, "Multi-user detection using ADMM-based compressive sensing for uplink grant-free NOMA," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 46–49, Sep. 2017.

[59] Y. Du, B. Dong, W. Zhu, P. Gao, Z. Chen, X. Wang, and J. Fang, "Joint channel estimation and multiuser detection for uplink grant-free NOMA," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 682–685, Feb. 2018.

[60] N. Y. Yu, "Multiuser activity and data detection via sparsity-blind greedy recovery for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 2082–2085, Aug. 2019.

[61] L. Wu, P. Sun, Z. Wang, and Y. Yang, "Joint user activity identification and channel estimation for grant-free NOMA: A spatial-temporal structure enhanced approach," *IEEE Internet Things J.*, Mar. 2021.

[62] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2320–2323, Aug. 2016.

[63] Y. Du, B. Dong, Z. Chen, X. Wang, Z. Liu, P. Gao, and S. Li, "Efficient multi-user detection for uplink grant-free NOMA: Prior-information aided adaptive compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812–2828, Dec. 2017.

[64] Y. Cui, W. Xu, Y. Wang, J. Lin, and L. Lu, "Side-information aided compressed multi-user detection for up-link grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7720–7731, Nov. 2020.

[65] T. Li, J. Zhang, Z. Yang, Z. L. Yu, Z. Gu, and Y. Li, "Dynamic user activity and data detection for grant-free noma via weighted $\ell_{2,1}$ minimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1638–1651, Mar. 2022.

[66] W. Kim, Y. Ahn, and B. Shim, "Deep neural network-based active user detection for grant-free NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2143–2155, Apr. 2020.

[67] T. Sivalingam, S. Ali, N. Huda Mahmood, N. Rajatheva, and M. Latva-Aho, "Deep neural network-based blind multiple user detection for grant-free multi-user shared access," in *Proc. IEEE PIMRC.*, Sep. 2021, pp. 1–7.

[68] X. Miao, D. Guo, and X. Li, "Grant-free NOMA with device activity learning using long short-term memory," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 981–984, Feb. 2020.

[69] Y. Zou, Z. Qin, and Y. Liu, "Joint user activity and data detection in grant-free NOMA using generative neural networks," in *Proc. IEEE ICC.*, Aug. 2021, pp. 1–6.

[70] M. H. Rahman, M. A. S. Sejan, S.-G. Yoo, M.-A. Kim, Y.-H. You, and H.-K. Song, "Multi-user joint detection using Bi-directional deep neural network framework in NOMA-OFDM system," *Sensors*, vol. 22, no. 18, p. 6994, Sep. 2022.

[71] A. Emir, F. Kara, H. Kaya, and H. Yanikomeroglu, "Deepmud: Multi-user detection for uplink grant-free NOMA IoT networks via deep learning," *IEEE Wireless Commun. Lett.*, Feb. 2021.

[72] Z. Mao, X. Liu, M. Peng, Z. Chen, and G. Wei, "Joint channel estimation and active-user detection for massive access in internet of things—A deep learning approach," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2870–2881, Feb. 2022.

[73] J. E. Hershey, A. A. Hassan, and R. Yarlagadda, "Unconventional cryptographic keying variable management," *IEEE Trans. Commun.*, vol. 43, no. 1, pp. 3–6, Jan. 1995.

[74] N. Patwari, J. Croft, S. Jana, and S. K. Kasera, "High-rate uncorrelated bit extraction for shared secret key generation from channel measurements," *IEEE Trans. Mobile Computing.*, vol. 9, no. 1, pp. 17–30, Jan. 2009.

[75] Q. Xu, P. Ren, H. Song, and Q. Du, "Security enhancement for IoT communications exposed to eavesdroppers with uncertain locations," *IEEE Access*, vol. 4, pp. 2840–2853, Jun. 2016.

[76] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Mar. 2017.

[77] H. Sharma, N. Kumar, B. K. Panigrahi, and A. Alotaibi, "Deep learning-based authentication framework for secure terrestrial communications in next generation heterogeneous networks," *IEEE Internet Things Mag.*, vol. 5, no. 4, pp. 174–179, Dec. 2022.

[78] M. Abdrabou and T. A. Gulliver, "Adaptive physical layer authentication using machine learning with antenna diversity," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6604–6614, Oct. 2022.

[79] R. H. Weber, "Internet of things - new security and privacy challenges," *Comput. Law Secur. Rev.*, vol. 26, no. 1, pp. 23–30, Jan. 2010.

[80] X. Wang, P. Hao, and L. Hanzo, "Physical-layer authentication for wireless security enhancement: Current challenges and future developments," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 152–158, Jun. 2016.

[81] L. Y. Paul, J. S. Baras, and B. M. Sadler, "Physical-layer authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 1, pp. 38–51, Feb. 2008.

[82] W. Hou, X. Wang, J.-Y. Chouinard, and A. Refaey, "Physical layer authentication for mobile systems with time-varying carrier frequency offsets," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1658–1667, May 2014.

[83] P. Zhang, Y. Shen, X. Jiang, and B. Wu, "Physical layer authentication jointly utilizing channel and phase noise in MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2446–2458, Jan. 2020.

[84] N. Xie, J. Chen, and L. Huang, "Physical-layer authentication using multiple channel-based features," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 2356–2366, Jan. 2021.

[85] Y. Chen, P.-H. Ho, H. Wen, S. Y. Chang, and S. Real, "On physical-layer authentication via online transfer learning," *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1374–1385, Jan. 2022.

[86] H. Fang, X. Wang, and L. Hanzo, "Learning-aided physical layer authentication as an intelligent process," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2260–2273, Nov. 2019.

[87] H. Fang, X. Wang, N. Zhao, and N. Al-Dhahir, "Lightweight continuous authentication via intelligently arranged pseudo-random access in 5G-and-beyond," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4011–4023, Mar. 2021.

[88] M. Mitev, A. Chorti, H. V. Poor, and G. P. Fettweis, "What physical layer security can do for 6G security," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 375–388, Feb. 2023.

[89] P. Tedeschi, S. Sciancalepore, and R. Di Pietro, "Satellite-based communications security: A survey of threats, solutions, and research challenges," *Comput. Netw.*, vol. 216, p. 109246, Aug. 2022.

[90] "Authentication and key management for applications (AKMA) based on 3GPP credentials in the 5g system (5gs)," in *3GPP Technical Specification 33.535, v17.7.0*, Sep. 2022.

[91] J. Li, J. Wang, and C. Huang, "Online distribution method of application key based on AKMA," in *Proc. IEEE (WCCCT)*, Feb. 2023, pp. 135–138.

[92] M. Khan, P. Ginzboorg, and V. Niemi, "AKMA: Delegated authentication system of 5G," *IEEE Commun. Mag.*, vol. 5, no. 3, pp. 56–61, Sep. 2021.

[93] W. Lv, P. Yang, Y. Ding, Z. Wang, C. Lin, and Q. Wang, "Energy-efficient and QoS-aware computation offloading in GEO/LEO hybrid satellite networks," *Remote Sens.*, vol. 15, no. 13, p. 3299, Jun. 2023.

[94] A. K. Dwivedi, H. Chougrani, S. Chaudhari, N. Varshney, and S. Chatzinotas, "Efficient transmission scheme for LEO satellite-based NB-IoT: A data-driven perspective," *arXiv preprint arXiv:2406.14107*, Jun. 2024.

[95] H. Deng, C. Zhang, W. Zhang, J. Liang, L. Wang, and L. Zhu, "IoTAuth: A decentralized cross-chain identity authentication scheme for 6G non-terrestrial IoT networks," *IEEE Netw.*, pp. 1–1, Mar. 2024.

[96] Z. Gao, D. Zhang, J. Zhang, Z. Liu, H. Liu, and M. Zhao, "BC-AKA: Blockchain based asymmetric authentication and key agreement protocol for distributed 5G core network," *China Commun.*, vol. 19, no. 6, pp. 66–76, Jun. 2022.

[97] B. Yang, S. Liu, T. Xu, C. Li, Y. Zhu, Z. Li, and Z. Zhao, "AI-oriented two-phase multifactor authentication in SAGINs: Prospects and challenges," *IEEE CONSUM ELECTR M.*, vol. 13, no. 1, pp. 79–90, Jan. 2024.

[98] G. Elamparithi, "Resilient service authentication for smart city application using IoT." *Intell. Autom. Soft Comput.*, vol. 36, no. 1, Sep. 2023.

[99] W. B. Ameur, P. Mary, M. Dumay, J.-F. Hélard, and J. Schwoerer, "Power allocation for BER minimization in an uplink MUSA scenario," in *Proc. IEEE (VTC-Spring)*, Jun. 2020, pp. 1–5.

[100] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, Feb. 2017.

[101] S. Chen and D. Donoho, "Basis pursuit," in *Proc. IEEE ACSSC.*, Aug. 1994, pp. 41–44.

[102] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[103] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[104] Y. Bin, Y. Yang, F. Shen, N. Xie, H. T. Shen, and X. Li, "Describing video with attention-based bidirectional LSTM," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2631–2641, May. 2019.

[105] N. Kim, D. Kim, B. Shim, and K. B. Lee, "Deep learning-based spreading sequence design and active user detection for massive machine-type communications," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1618–1622, Jun. 2021.

[106] 3GPP, "DMRS design for non-orthogonal UL multiple access user channel estimation," in *R1-1609548, 3GPP TSG-RAN WG1 Meeting*, Oct. 2016.

[107] Z. Yuan, C. Yan, Y. Yuan, and W. Li, "Blind multiple user detection for grant-free MUSA without reference signal," in *Proc. IEEE (VTC-Fall)*, Feb. 2017, pp. 1–5.

[108] H. Palangi, R. Ward, and L. Deng, "Distributed compressive sensing: A deep learning approach," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4504–4518, Apr. 2016.

[109] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972–974, May 2012.

[110] "Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation," in *3GPP Technical Report 36.211, v16.6.0*, Jun. 2021.

[111] A. Goldsmith, *Wireless communications.* Cambridge university press, Aug. 2005.

[112] T. S. Rappaport, *Wireless communications: principles and practice.* Cambridge University Press, Feb. 2024.

[113] Y. Zhang, L. T. Yang, and J. Chen, *RFID and sensor networks: architectures, protocols, security, and integrations.* CRC Press, Nov. 2009.

[114] T. Qiu, Y. Zhang, D. Qiao, X. Zhang, M. L. Wymore, and A. K. Sangaiah, "A robust time synchronization scheme for industrial internet of things," *IEEE Trans Industr Inform.*, vol. 14, no. 8, pp. 3570–3580, Aug. 2018.

[115] A. Elsts, X. Fafoutis, S. Duquennoy, G. Oikonomou, R. Piechocki, and I. Craddock, "Temperature-resilient time synchronization for the internet of things," *IEEE Trans Industr Inform.*, vol. 14, no. 5, pp. 2241–2250, May. 2018.

[116] J. A. Stankovic, "Research directions for the internet of things," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 3–9, Feb. 2014.

[117] O. O. Koyluoglu, H. El Gamal, L. Lai, and H. V. Poor, "Interference alignment for secrecy," *IEEE Trans. Inf.*, vol. 57, no. 6, pp. 3323–3332, Jun. 2011.

[118] J. B. Perazzone, L. Y. Paul, B. M. Sadler, and R. S. Blum, "Cryptographic side-channel signaling and authentication via fingerprint embedding," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 9, pp. 2216–2225, Sep. 2018.

[119] B. Wang, L. Dai, T. Mir, and Z. Wang, "Joint user activity and data detection based on structured compressive sensing for NOMA," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1473–1476, Apr. 2016.

[120] F. J. Liu, X. Wang, and S. L. Primak, "A two dimensional quantization algorithm for CIR-based physical layer authentication," in *Proc. IEEE (ICC)*, Jun. 2013, pp. 4724–4728.

[121] J. Liu and X. Wang, "Physical layer authentication enhancement using two-dimensional channel quantization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4171–4182, Feb. 2016.

[122] N. Xie, Z. Li, and H. Tan, "A survey of physical-layer authentication in wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 282–310, Dec. 2021.

[123] M. Edman, A. Kiayias, and B. Yener, "On passive inference attacks against physical-layer key extraction?" in *Proc. 4th Euro. Wksp. System Security*, Apr. 2011, pp. 1–6.

[124] K. Zeng, "Physical layer key generation in wireless networks: challenges and opportunities," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 33–39, Jun. 2015.

[125] P. H. Bardell, W. H. McAnney, and J. Savir, *Built-in test for VLSI: pseudorandom techniques*. Wiley-Interscience, 1987.

[126] J. B. Fraleigh, *A first course in abstract algebra*. Pearson Education India, Dec. 2003.

[127] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. IEEE (VTC-Fall)*, Sep. 2014, pp. 1–5.

[128] B. Wang, K. Wang, Z. Lu, T. Xie, and J. Quan, "Comparison study of non-orthogonal multiple access schemes for 5G," in *Proc. IEEE (BMSB)*, Jun. 2015, pp. 1–5.

[129] A. Alnoman, S. Erkucuk, and A. Anpalagan, "Sparse code multiple access-based edge computing for IoT systems," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7152–7161, May 2019.

[130] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. De Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sep. 2018.

[131] I. Altaf, M. A. Saleem, K. Mahmood, S. Kumari, P. Chaudhary, and C.-M. Chen, "A lightweight key agreement and authentication scheme for satellite-communication systems," *IEEE Access*, vol. 8, pp. 46 278–46 287, Mar. 2020.

[132] T. Wang, Y. Liu, and A. V. Vasilakos, "Survey on channel reciprocity based key establishment techniques for wireless systems," *Wireless Netw.*, vol. 21, no. 6, pp. 1835–1846, Aug. 2015.

[133] M. Wilhelm, I. Martinovic, and J. B. Schmitt, "Secure key generation in sensor networks based on frequency-selective channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1779–1790, Sep. 2013.

[134] M. Edman, A. Kiayias, Q. Tang, and B. Yener, "On the security of key extraction from measuring physical quantities," *IEEE Trans. Inf. Forensics Security,*, vol. 11, no. 8, pp. 1796–1806, Aug. 2016.

[135] J. Zhang, A. Marshall, R. Woods, and T. Q. Duong, "Efficient key generation by exploiting randomness from channel responses of individual ofdm subcarriers," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2578–2588, Jun. 2016.

[136] T. M. Hoang, T. Q. Duong, H. D. Tuan, S. Lambotharan, and L. Hanzo, "Physical layer security: Detection of active eavesdropping attacks by support vector machines," *IEEE Access*, vol. 9, pp. 31 595–31 607, Feb. 2021.

[137] R.-F. Liao, H. Wen, S. Chen, F. Xie, F. Pan, J. Tang, and H. Song, "Multiuser physical layer authentication in internet of things with data augmentation," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 2077–2088, Mar. 2020.

[138] F. Clazzer, A. Munari, G. Liva, F. Lazaro, C. Stefanovic, and P. Popovski, "From 5G to 6G: Has the time for modern random access come?" *arXiv preprint arXiv:1903.03063*, Mar. 2019.

[139] Z. Zhang *et al.*, "User activity detection and channel estimation for grant-free random access in LEO satellite-enabled internet of things," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8811–8825, May. 2020.

[140] B. Di, L. Song, Y. Li, and H. V. Poor, "Ultra-dense LEO: Integration of satellite access networks into 5G and beyond," *IEEE Wirel. Commun.*, vol. 26, no. 2, pp. 62–69, Apr. 2019.

[141] A. Abdi, W. Lau, M.-S. Alouini, and M. Kaveh, "On the second-order statistics of a new simple model for land mobile satellite channels," in *Proc. IEEE VTC-Fall*, vol. 1, Jul. 2001, pp. 301–304 vol.1.

[142] M. Abdrabou and T. A. Gulliver, "Authentication for satellite communication systems using physical characteristics," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 48–60, Nov. 2023.

[143] U. M. Maurer and S. Wolf, "The diffie–hellman protocol," *Des. Codes Cryptogr.*, vol. 19, no. 2, pp. 147–171, Mar. 2000.

[144] "Study on new radio (NR) to support non terrestrial networks," in *3GPP Technical Report 38.811 (Release 15)*, Jun. 2018.

[145] "Technical specification group radio access network; solutions for NR to support nonterrestrial networks (NTN)," in *3GPP Technical Specification 38.821 (Release 16)*, Dec. 2019.

[146] V. Mandawaria, C. Majumdar, S. Park, N. Sharma, A. Nigam, and J. Jung, "Grant-free massive access for LEO-satellite based 6G IoT networks," in *Proc. IEEE GLOBECOM Wkshps*, Dec. 2022, pp. 862–867.

[147] "White paper: 6G satellite and terrestrial network convergence," May. 2023. [Online]. Available: https://www.mediatek.com/blog/white-paper-6g-satellite-and-terrestrial-network-convergence/

[148] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972–974, May. 2012.

[149] J. A. Fraire, O. Iova, and F. Valois, "Space-terrestrial integrated internet of things: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 60, no. 12, pp. 64–70, Sep. 2022.

[150] "Study on new radio (NR) to support non terrestrial networks," in *3GPP Technical Report 38.811 (Release 15)*, Jun. 2018.

[151] "Solutions for NR to support nonterrestrial networks (NTN)," in *3GPP Technical Specification 38.821 (Release 16)*, Dec. 2019.

[152] "RAN improvements for machine-type communications," in *3GPP Technical Specification 37.868 (Release 11)*, Sep. 2011.

[153] "Study on provision of low-cost machine-type communications (MTC) user equipments (UEs) based on LTE," in *3GPP Technical Specification 36.888 (Release 11)*, Jun. 2012.