BLE-based Indoor localization and Contact Tracing Approaches

Mohammad Salimibeni

A Thesis
In The Department
of
Concordia Institute for Information Systems Engineering (CIISE)

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy(Electrical and Computer Engineering) at
Concordia University
Montréal, Québec, Canada

June 2023

Concordia University
School of Graduate Studies

This is to certify that the thesis prepared

By: **Mohammad Salimibeni**

Entitled: **BLE-based Indoor localization and Contact Tracing Approaches**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Information and Systems Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

_____ Chair

Dr. Wei-Ping Zhu _____ External Examiner

Dr. Walter Lucia _____ Examiner

Dr. Mohsen Ghafouri _____ Examiner

Dr. Arash Mohammadi _____ Supervisor

Approved by Jun Yan _____
Graduate Program Director

05/23/2023 _____
Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

BLE-based Indoor localization and Contact Tracing Approaches

Mohammad Salimibeni, Ph.D.
Concordia University, 2023

Internet of Things (IoT) has penetrated different aspects of modern life with smart sensors being prevalent within our surrounding indoor environments. Furthermore, dependence on IoT-based Contact Tracing (CT) models has significantly increased mainly due to the COVID-19 pandemic. There is, therefore, an urgent quest to develop/design efficient, autonomous, trustworthy, and secure indoor CT solutions leveraging accurate indoor localization/tracking approaches. In this context, the first objective of this Ph.D. thesis is to enhance accuracy of Bluetooth Low Energy (BLE)-based indoor localization. BLE-based localization is typically performed based on the Received Signal Strength Indicator (RSSI). Extreme fluctuations of the RSSI occurring due to different factors such as multi-path effects and noise, however, prevent the BLE technology to be a reliable solution with acceptable accuracy for dynamic tracking/localization in indoor environments. In this regard, first, an IoT dataset is constructed based on multiple thoroughly separated indoor environments to incorporate the effects of various interferences faced in different spaces. The constructed dataset is then used to develop a Reinforcement Learning (RL)-based information fusion strategy to form a multiple-model implementation consisting of RSSI, Pedestrian dead reckoning (PDR), and Angle-of-Arrival (AoA)-based models. In the second part of the thesis, the focus is devoted to application of multi-agent Deep Neural Networks (DNN) models for indoor tracking. DNN-based approaches are, however, prone to overfitting and high sensitivity to parameter selection, which results in sample inefficiency. Moreover, data labelling is a time-consuming and costly procedure. To address these issues, we leverage Successor Representations (SR)-based techniques, which can learn the expected discounted future state occupancy, and the immediate

reward of each state. A Deep Multi-Agent Successor Representation framework is proposed that can adapt quickly to the changes in a multi-agent environment faster than the Model-Free (MF) RL methods and with a lower computational cost compared to Model-Based (MB) RL algorithms. In the third part of the thesis, the developed indoor localization techniques are utilized to design a novel indoor CT solution, referred to as the Trustworthy Blockchain-enabled system for Indoor Contact Tracing (TB-ICT) framework. The TB-ICT is a fully distributed and innovative blockchain platform exploiting the proposed dynamic Proof of Work (dPoW) approach coupled with a Randomized Hash Window (W-Hash) and dynamic Proof of Credit (dPoC) mechanisms.

Acknowledgments

First and foremost, I want to extend my heartfelt appreciation to my supervisor, Dr. Arash Mohammadi, for his unwavering support throughout my Ph.D. journey. His patience, motivation, enthusiasm, and extensive knowledge have been invaluable to my research and academic growth. I am also deeply grateful to the members of my committee, Dr. Walter Lucia, Dr. Wei-Ping Zhu, Dr. Mohsen Ghafouri, and Xiao-Ping Zhang, for their valuable insights and feedback on my dissertation.

To my parents, I want to express my profound gratitude for their incredible support, inspiration, and understanding throughout this entire process. Their unwavering belief in me has been instrumental in my success. And last but certainly not least, I want to dedicate a special thank you to my beloved wife, Farnoush. Your unwavering support, patience, and love have been my guiding light. I am forever grateful for your presence in my life and the strength you have given me to believe in myself.

Contents

List of Figures

List of Tables

Abbreviations

BLE Bluetooth Low Energy
RSSI Received Signal Strength Indicator
LBSs Location-Based Services
AI Artificial Intelligence
IoT Internet of Things
ML Machine Learning
CNN Convolutional Neural Network
DNN Deep Neural Networks
RL Reinforcement Learning
MDP Markov Decision Process
SR Successor Representations
DQN Deep Q-Networks
DDPG Deep Deterministic Policy Gradient
MADDPG Multi-Agent Deep Deterministic Policy Gradient
RGD Restricted Gradient Descent
CMAC Cerebellar Model Articulation Controllers
RBF Radial Basis Functions
FPKF Fixed-Point Kalman Filte
KTD Kalman Temporal Difference
GPTD Gaussian Process Temporal Difference
LSTD Least Square Temporal Difference
MSE Mean Square Error
MMSE Minimum Mean Square Error
MMAE Multiple Model Adaptive Estimation
MF Model Free
MB Model Based

MARL Multi-Agent Reinforcement Learning
IoT-TD Internet of Things Tracking Dataset
MAK-TD Multi-Agent Adaptive Kalman Temporal Difference
MAK-SR Multi-Agent Adaptive Kalman Successor Representation
RL-IFF RL-based Fusion Framework
CT Contact Tracing
ICT Intelligent Contact Tracing
PDR Pedestrian Dead Reckoning
I/Q In-phase and Quadrature
AoA Angle of Arrival
ToA Time of Arrival
AoD Angle of Departure
ToF Time of Flight
TDoA Time Difference of Arrival
LAA Linear Antenna Array
MUSIC MUltiple SIgnal Classification
PF Particle Filtering
KNN K-nearest neighbour
GMM Gaussian Mixture Models
SVM Support Vector Machines
KFPF Cascaded Kalman Filter-Particle Filter
CSI Channel Impulse Response
DApps Decentralized Applications
GPS Global Positioning Systems
UWB Ultra Wide Band
PoC Proof of Credit
WDPoC Randomized Hash Window Dynamic Proof of Credit
AWGN Additive White Gaussian Noise
SNRs Signal to Noise Ratios
UKF Unscented Kalman Filter
ECC Elliptic Curve Cryptography
ECDSA Elliptic Curve Digital Signature Algorithm
ECDF Empirical Cumulative Distribution Function

LoS Line of Sight

NLoS None Line of Sight

CTE Constant Tone Extension

EKFA Extended Kalman Filter Angle of Arrival

EKFT EKF Time Difference of Arrival

PKI Public Key Infrastructure

UKFA UKF Angle of Arrival

P2P Peer-to-Peer

UKFT UKF Time of Arrival

DH Diffie-Hellman

UKFTA UKF Time Difference of AoA

LTE Long-Term Evolution

GF Gaussian Filter

DL Difficulty Level

DL_h Difficulty level 2 - hard

DL_e Difficulty level 1 - easy

CIR Channel Impulse Response

IUP Infected Users Pool

PoW Proof of Work

dPoW Dynamic Proof of Work

ST Submission Transaction

TT Trace Transaction

QT Query Transaction

RT Alarm Transaction

W-Hash Randomized Hash Window

DPoC Dynamic Proof of Credit

UTXO Unspent Transaction Outputs

NHS National Health Service

HCS Health Code System

RLP Recursive Length Prefix

IoD Internet of Drones

Chapter 1

Overview of the Thesis

1.1 Introduction

Indoor localization is a critical area of research as the Internet of Things (IoT) concept continues to gain popularity and its potential applications are explored. Bluetooth Low Energy (BLE) is widely recognized as one of the most effective technologies for localization in IoT [1]. The integration of different BLE-based indoor localization solutions through fusion strategies has the potential to significantly enhance indoor location estimation [1–9]. BLE offers several benefits over alternative technologies such as WiFi and Ultra Wide-Band (UWB), including its low power consumption [10, 11]. In addition, single-model RSSI-based indoor localization solutions are relatively simple to implement, cost-effective, and can be utilized in a range of environments. However, a major challenge in BLE-based localization is the accurate calculation of dynamic distance estimates between sensing devices and an intended object or user, considering the fluctuating nature of the Received Signal Strength Indicator (RSSI) [12, 13].

RSSI-based solutions are prone to multi-path fading and drastic fluctuations in the indoor environment, which can lead to inaccurate indoor positioning. To deal with these issues, researchers have developed advanced signal processing solutions and Machine Learning (ML) schemes such as Artificial Neural Networks (ANNs) and Reinforcement Learning [10]. However, these methods still have limitations, such as the instability caused by fading effects, and the multi-path effects and heterogeneity of BLE-enabled devices [14, 15].

One of the main solutions to address the problems related to the single-model RSSI-based indoor localization solutions is incorporating multi-model localization/tracking solutions. These solutions involve using a combination of different technologies and models to enhance the accuracy of the localization/tracking services. By combining different models, the system can benefit from the strengths of each individual model, resulting in a more robust and accurate localization system.

Different methods [16–22] have been proposed to localize a user within an indoor environment. Aside from RSSI-based indoor localization solutions, there are other methods that have been proposed to achieve the same goal. Some of these approaches use additional hardware such as single or array of antennas, or complementary board chips to estimate Angle of Arrival (AoA) or Angle of Departure (AoD) parameters. These methods, such as Time of Flight (ToF) or Time Difference of Arrival (TDoA) [23,24], increase the complexity of the system due to the need for sensor synchronization. Another approach to provide indoor localization solutions is Pedestrian dead reckoning (PDR) [7]. This method uses sensor data from a user's device to estimate their location by tracking their movements. Unlike the previous methods, PDR does not require any additional hardware and can be used in a standalone mode. However, the accuracy of PDR relies on the quality of sensor data, which can be affected by various factors such as device orientation or environmental conditions.

By providing indoor localization services and estimating the user's location, a wide range of services such as Smart Contact Tracing (SCT), context-aware solutions, targeted advertisements, tenant assistance, and automated access can be provided. In the context of a global crisis such as COVID-19, SCT is an effective solution in preventing the spread of the virus, especially in indoor environments where the spreading probability is higher [13]. Therefore, there is an urgent need to develop efficient, autonomous, and trustworthy indoor SCT solutions with the highest privacy-preserving capabilities. BLE-based tracking/localization is a viable technology for the development of SCT solutions for indoor environments. To achieve trustworthiness and high indoor localization accuracy, researchers have proposed several methods such as centralized and decentralized solutions. Blockchain-enabled design, coupled with indoor localization via Bluetooth Low Energy (BLE) beacons [21] are among the top solutions provided in this regard. By using blockchain technology, the system can

be made tamper-proof, ensuring the privacy and security of the users' data. AoA-based localization is one of the top indoor localization solutions proposed in SCT frameworks. AoA-based localization [22] can provide more accurate results compared to RSSI-based methods as it takes into account the orientation of the transmitter and receiver. Although the AoA can result in an acceptable indoor localization/tracking solution, there are some drawbacks that come with single model localization solutions. The main disadvantage of using AoA as a single localization model is that it is heavily dependent on the accuracy of the antenna array and the surrounding environment. Factors such as multipath, interference, and noise can greatly affect the accuracy of the AoA measurements. Additionally, AoA is not suitable for situations where the signal strength is weak, as the measurements may become unreliable [3].

Using a multi-model indoor localization approach by fusing the information of different indoor localization solutions can result in a more robust and accurate localization solution. By combining different localization techniques, the system can take advantage of the strengths of each method and overcome the limitations of any individual approach. However, the difficulty lies in effectively integrating and processing the various data sources, as well as ensuring the compatibility and consistency of the different solutions. Additionally, calibrating and fine-tuning the system to the specific environment can also present challenges.

In multi-model indoor localization and tracking, RL-based information fusion strategy [3] has been proposed as a solution to overcome the limitations of single-technology or single-processing model solutions. By integrating different technologies and processing solutions simultaneously, RL-based information fusion techniques can improve the robustness and accuracy of indoor localization. Moreover, while in most localization scenarios, indoor localization should be provided for multi-agent environments where multiple users are in the indoor venue, in these multi-agent scenarios, RL-based solutions can face challenges due to the complexity of the environment and the need to coordinate the actions of multiple agents. To address these challenges, researchers have proposed several approaches, one of the main ones being Multi-Agent Successor Representation (SR). In this context, SR can be proposed to improve the performance of RL-based solutions in multi-agent environments. SR can benefit us in multi-agent scenarios by providing a way for agents to share information and coordinate their actions in a decentralized manner. In a multi-agent system, each agent

may have its own sensors and data, and SR can be used to combine this information in a way that allows the agents to make more informed decisions.

In multi-model indoor localization and tracking, RL-based information fusion strategy has been proposed as a solution to overcome the limitations of single-technology or single-processing model solutions. By integrating different technologies and processing solutions simultaneously, RL-based information fusion techniques can improve the robustness and accuracy of indoor localization. This thesis focuses on using RL-based information fusion techniques to improve the performance of AoA-based localization in indoor environments.

In this regard, multiple-model RL-based information fusion has attracted a great deal of interest from many researchers as a means of providing indoor localization solutions. This is due to its ability to fuse information from multiple sources, which improves the accuracy and robustness of localization results. Additionally, Multi-agent SR solutions can significantly enhance the capability of the localization platform to provide the information fusion result by utilizing the collective intelligence of multiple agents to make more informed decisions. Furthermore, by leveraging the distributed nature of the system, it can overcome challenges such as occlusions and noisy sensor data.

The principal goal of this thesis is to develop a privacy-preserving, robust, and highly precise indoor localization solution. Initially, our investigation was centered on Bluetooth Low Energy (BLE) technology, owing to its broad adoption. We conducted numerous data collection and assessment sessions to thoroughly explore the various facets of BLE-based indoor localization solutions. Our assessments revealed that single-model BLE-based localization approaches, which focus primarily on proximity for robustness, fail to meet the expected accuracy levels. Consequently, for indoor tracking, our main emphasis pivoted towards multi-model and hybrid solutions. These solutions amalgamate the benefits of individual localization techniques to enhance accuracy and provide dynamic localization over time. As part of our multi-model strategy, we employed adaptive filters to mitigate the factors that compromise the accuracy of RSSI. Challenges posed by diverse factors such as device orientation and environmental obstacles were addressed through these adaptive weights. We endeavored to minimize the influence of environmental variables as much as possible, thereby facilitating the deployment of the proposed multi-model indoor localization

solution in a variety of unforeseen settings. In a bid to attain robustness and accuracy, we explored adaptive Reinforcement Learning (RL) agents. In this thesis, the term 'agent' is used to denote a local processing unit equipped with communication, computation, and sensing capabilities. Depending on the context, these agents can represent users carrying handheld devices, cooperating to improve localization accuracy. Alternatively, in certain scenarios, we designate some anchor nodes as agents. These nodes transmit and receive signals, process data, and participate in cooperation schemes to enhance accuracy. Our primary focus in this thesis is on the theory of multi-agent RL, particularly cooperative scenarios, to address uncertainties in measurements. The future work application of this theory would be its deployment for indoor localization.

1.2 Research Objectives

The primary objective of this Ph.D. thesis is to address the challenges associated with indoor localization and its crucial application, Smart Contact Tracing (SCT), by devising an innovative reinforcement learning (RL)-based information fusion strategy. This approach aims to enhance the accuracy of a multi-model indoor localization and tracking solution. Additionally, a novel Blockchain-based system for an Indoor CT framework is proposed to ensure the trustworthiness and security of the platform. In this context, the thesis targets the accomplishment of the following four main research objectives:

- **Investigate different BLE-based indoor localization solutions**: The first research objective seeks to probe into and analyze the existing Bluetooth Low Energy (BLE) based indoor localization solutions and compare their performance, accuracy, and reliability. The intent is to study the diverse methodologies and techniques deployed for indoor localization using BLE signals, such as fingerprinting, triangulation, and others. The research will concentrate on evaluating the merits and demerits of different solutions and identifying the most effective approach for optimal indoor localization accuracy. The findings of this research will be instrumental in providing insights and suggestions for future indoor localization systems based on BLE signals.

- **Study the theory of Multi-agent RL and Successor Representation**

(SR): This research objective focuses on a comprehensive study of Multi-agent Reinforcement Learning (RL) theories and Successor Representation (SR). The primary goal is to understand the underlying principles and the application of these techniques in the domain of indoor localization. This includes understanding how multiple agents can cooperate and share information for improved performance, and how SR can be utilized to handle complexities in multi-agent environments. The result of this study will form the basis for the design and implementation of the RL-based information fusion system for indoor localization.

- **Developing a new RL-based information fusion multiple-model indoor localization**: This research objective involves the development of a new indoor localization system that leverages reinforcement learning (RL) and multiple-model information fusion techniques. The goal is to devise an indoor localization system capable of integrating multiple models and algorithms, thereby yielding a more accurate and reliable indoor localization solution. The RL-based system will be trained on a dataset to learn and optimize the fusion of different models. The system will be evaluated based on its accuracy, reliability, and robustness, and benchmarked against existing indoor localization solutions. The findings of this research will shed light on the potential advantages and limitations of RL-based indoor localization systems.

- **Developing a trustworthy SCT framework**: The central focus of this research objective is to establish a secure, trustworthy, and privacy-oriented Smart Contact Tracing (SCT) framework that plays a crucial role in mitigating the transmission of contagious diseases like COVID-19. The SCT framework must safeguard the confidentiality and privacy of users' personal and health information, while ensuring that the contact tracing data is accurate, current, and protected from tampering. To achieve this, the SCT framework will be designed with privacy as a priority. This can be accomplished by integrating blockchain technology, Internet of Things (IoT) networks, and a state-of-the-art indoor localization solution. This combination will ensure secure data sharing, a robust authentication process, and a data verification system, thereby ensuring the efficacy of the SCT framework in controlling the spread of diseases in indoor settings.

1.3 Targeted Challenges

Despite recent advancements in the field of indoor localization and increase of its potential applications, there are still several open problems and challanges, which require extensive investigations, including:

C1. **Lack of Reliable Indoor User Tracking Data** : The challenge in this situation is the lack of a trustworthy dataset that provides the ground truth of users in indoor environments. This dataset should contain information on the Radio Signal Strength Indicator (RSSI) data received by Bluetooth Low Energy (BLE) modules, the Angle of Arrival (AoA), and the data from the Inertial Measurement Unit (IMU) sensor all simultaneously for multiple environments. This information is crucial in accurately tracking the movement of users in indoor spaces. The absence of such a dataset makes it difficult to validate and improve upon current indoor tracking systems, hindering advancements in this field. Therefore, there is a need for a comprehensive and reliable dataset that can provide accurate and consistent information on user tracking in indoor environments.

C2. **Data-Driven BLE-Based Indoor Localization: Overcoming Challenges in IoT Environments**: In this challenge, the focus is on developing a data-driven localization model that utilizes Bluetooth Low Energy (BLE) sensor measurements to achieve high indoor localization accuracy. The goal is to leverage the availability of IoT indoor localization infrastructures and capitalize on the potential of BLE sensors to provide precise and efficient localization. A lot of challenges are associated with this task. In indoor environments, Line of Sight (LoS) links may not always be available, and the signals may be affected by Additive White Gaussian Noise (AWGN) with varying Signal to Noise Ratios (SNRs). Therefore, the data-driven localization model needs to be able to overcome these obstacles and still provide accurate results. This challenge aims to tackle these difficulties and develop a robust and reliable indoor localization system that leverages the power of BLE and IoT.

C3. **Overcoming Limitations in Multi-Agent Reinforcement Learning** : The challenge in Multi-Agent Reinforcement Learning (MARL) lies in the limitations of traditional Model-Based (MB) and Model-Free (MF) algorithms in

handling the complexities of multiple agents and their interactions. The fixed reward model used in these algorithms can lead to suboptimal results and the Deep Neural Network (DNN) solutions, while effective, are susceptible to overfitting, sensitivity to parameter selection and poor sample efficiency. This presents a unique set of challenges in the development of MARL algorithms, requiring innovative approaches to overcome these limitations.

C4. **SR in MARL: Navigating Uncertainty and Balancing Exploration and Exploitation** : SR is a promising approach to be used in MARL. However, its implementation in MARL faces several challenges that need to be addressed. One of the challenges is capturing the uncertainty associated with the SR, which is crucial for decision making in MARL. Another challenge is the calculation of the value function based on the learned SR values and the reward function. This requires a careful balance between exploration and exploitation, as the agents must decide whether to select actions with known rewards or explore new actions with unknown rewards. These challenges make the implementation of SR in MARL a complex and challenging task, but also provide opportunities for research and innovation in this field.

C5. **Bridging the Gap between Stand-alone Models and Reliable Information Fusion** : Indoor localization has been an ongoing research challenge in recent years. However, existing works mainly focus on developing stand-alone models for this purpose, neglecting the importance of information fusion strategies in enhancing the accuracy of the results. Different types of data such as Bluetooth beacons, and inertial sensors have been used in indoor localization, but a comprehensive approach to integrate these data sources is missing. This challenge requires researchers to find ways to combine the different types of data and develop effective information fusion strategies to enhance the accuracy of indoor localization. The goal is to bridge the gap between the stand-alone models and reliable information fusion for improved indoor localization results.

C6. **Complexities in Information Fusion in Hybrid Frameworks** : The use of hybrid frameworks in information fusion is a common approach in modern technology. However, this approach presents a number of challenges that need to be addressed to ensure accurate and effective information fusion. One of

the main challenges is the complexity of data labeling, which requires a lot of effort to properly label data for use in mathematical models. Additionally, the mathematical modeling of indoor environments can also present difficulties, as the data received from different sensors can vary over time, making it challenging to accurately model the environment. These complexities require effective solutions to ensure the reliability and accuracy of information fusion in hybrid frameworks.

C7. **Trustworthy Indoor Smart CT Solution for Pandemic Control**: The ongoing COVID-19 pandemic has brought to light the importance of controlling the spread of the virus in enclosed spaces, where the probability of transmission is much higher. In light of this, there is an immediate need for the development and design of efficient, autonomous, trustworthy and secure indoor CT solutions. The solutions need to be designed in such a way that they can effectively control the spread of the virus within indoor environments, making it safer for people to interact and work in these spaces. The challenge is to design a trusworthy indoor CT solutions that are not only effective but also practical and easily deployable, helping to create a safer environment for people in their daily lives.

C8. **The Integration of Blockchain in Smart Contact Tracing Framework**: The integration of blockchain technology into the smart contact tracing (SCT) framework is a complex task that involves creating a decentralized network that is both secure and trustworthy. This is made more complicated by the use of indoor localization solutions, which require the integration of different systems and technologies to accurately track user movements. However, the main challenge lies in ensuring the privacy and security of user data while still providing the necessary information to track the spread of diseases. A successful integration of blockchain technology will result in a decentralized, secure, and user-friendly network that will play a crucial role in preventing the spread of pandemics and other public health crises.

1.4 Thesis Contributions

Below, the contributions of the thesis are briefly outlined:

- Chapter 3 [7, 9]: Builidng a trustworthy localization dataset and applying IoT-based indoor localization solutions

  - IoT-TD [1, 4, 7, 9] is a real-world dataset to evaluate the performance of different localization methods. The Internet of Things Tracking Dataset (IoT-TD), contains synchronized data from Angle of Arrival (AoA), Received Signal Strength Indicator (RSSI), Inertial Measurement Unit (IMU), and the ground truth of the user's trajectories. The IoT-TD covers three different indoor environments and provides millimeter-level accuracy. Consequently, it is possible to develop a hybrid localization framework that combines RSSI, Pedestrain Dead Reckoning(PDR), and AoA to enhance the performance of indoor localization. The proposed hybrid framework considers the fluctuations in the received signal (RSSI), cumulative error in trajectory estimation (PDR), and frequency/phase shifting between the transmitter and receiver oscillators (AoA) as the most significant issues in BLE-based and IMU-based indoor localization. This contribution targets challenges C1 and C2 in Chapter 3, Section 3.1.
  - To tackle the challenge of evaluating indoor localization approaches, challenge $C2$, especially BLE-based solutions such as AoA and RSSI-based localization and IMU-based solutions such as PDR-based localization, Chapter 3 provides a comprehensive evaluation using the IoT-TD dataset in three different environments. By examining the strengths and weaknesses of BLE technology and the factors that affect the RSSI values, this Chapter highlights the tremendous potential of the proposed indoor localization solutions for precise indoor localization. To enhance the accuracy of RSSI-based localization, the Chapter also presents the development of an Indoor Localization SDK that employs an ML-based model to minimize the impact of certain parameters. The SDK is designed as a REST API, which enables seamless integration with various sensors in an indoor environment to provide real-time user proximity estimation. This contribution targets challenges C2 in Chapter 3, Sections 3.2, 3.3, and 3.4.

- Chapter 4 [2–4, 6, 8]: Multi-Agent Reinforcement Learning Successor Representation and Reinforcement Learning-based Information Fusion Indoor Localization

  - Motivated by the potentials that would be generated by integrating Kalman filter into the Q-learning algorithm, to enhances the accuracy of predictions and improves the sample efficiency of the learning process, a novel Multi-Agent Adaptive Kalman Temporal Difference (MAK-TD) framework [2, 8] is introduced in Section 4.2 targeting challenge $C3$. MAK-TD is a solution for a common problem in MARL where the measurement noise covariance is an unknown parameter. This parameter affects the accuracy of the predictions made by the learning algorithm, and if it is not estimated correctly, it can lead to suboptimal policies. The MAK-TD framework addresses this issue by using the Kalman filter to estimate the unknown parameter and incorporating it into the learning process. The proposed MAK-TD framework implements an off-policy Q-learning approach. Q-learning is a well-established RL algorithm that learns the optimal policy by estimating the expected future rewards for taking specific actions in a given state. In the off-policy approach, the learning algorithm does not follow the current policy but instead uses a behavior policy to gather experiences. This approach allows for more efficient exploration and faster learning, as the algorithm is not restricted to the current policy and can explore a wider range of actions. The MAK-TD framework uses the Kalman filter to estimate the measurement noise covariance, which is then incorporated into the Q-learning algorithm. This allows the algorithm to make more accurate predictions and learn the optimal policy faster. The combination of the Kalman filter and Q-learning provides a solution to the information inadequacy problem, while also improving sample efficiency.

  - The second solution referred to as Multi-Agent Adaptive Kalman Successor Representation (MAK-SR), is designed best on the capabilities we can gain using Successor Representation (SR) solutions. The MAK-SR is an extension of the MAK-TD algorithm. This extension involves the integration of the SR learning process into the filtering problem using the KTD formulation. SR is a better alternative to both model-free and model-based

RL solutions due to its improved sample efficiency, better generalization, handling of model uncertainty, increased flexibility, and reduced computational costs.The KTD formulation is used to approximate the uncertainty of the learned SR, which is a critical factor in the performance of the algorithm. By incorporating the SR learning process into the filtering problem, the MAK-SR algorithm is able to learn and update the successor representation in real-time. This provides a more accurate representation of the environment, which in turn improves the decision-making capabilities of the agent. One of the key benefits of adopting the KTD formulation is the reduction in the required memory and computation time to learn the SR. This makes the MAK-SR algorithm more scalable and computationally efficient compared to other algorithms, such as DNN-based algorithms. Another advantage of using KTD is that it reduces the sensitivity of the model to parameters selection. This results in a more reliable and robust algorithm, as the model is not easily influenced by small changes in the parameters. This improved reliability makes the MAK-SR algorithm more suitable for real-world applications where reliability is of utmost importance. This contribution targets challenges $C3$ and $C4$ in Section 4.3.

– To tackle the challenges of enhancing the accuracy of indoor localization, Chapter 4 presents a cutting-edge hybrid indoor localization solution. This solution combines the strengths of various technologies, including RSSI-based KF and PF, IMU-based PDR, and AoA module. These techniques are integrated via a RL-based Fusion Framework (RL-IFF) [1, 4]. This framework is designed to address the time-varying and non-stationary nature of the information obtained from each localization method, as well as the non-stationary Markov Decision Process in the proposed RL definition. By fusing the results from different sources, the hybrid framework can improve the overall accuracy of the localization and tracking system and reduce the impact of errors caused by any single method. This approach targets the challenges of reducing the impact of errors caused by any single method and enhancing the accuracy of the indoor localization system, as presented in Chapter 4, Section 4.4.This contribution targets challenges $C5$ and $C6$.

- Chapter 5 [1, 5, 8]: Blockchain-Enabled System for Indoor Contact Tracing in Epidemic Control

  - To address the challenges with trustworthy indoor CT an innovative framework called TB-ICT [1, 5] is designed to ensure that CT information is secure and protected from unauthorized access. This is achieved through two main capabilities. Firstly, the framework uses BLE beacons to accurately track close contact information between users in the case of infection. This information is then shared securely within the network. Secondly, the framework uses a secure and privacy-preserving communication protocol based on a blockchain-based distributed ledger. This protocol ensures that users' identities are kept anonymous by using a privacy-preserving address generation scheme. This scheme exploits the non-connectable advertising channels of BLE signals to transfer information between contacts, and unique temporary addresses are generated based on ambient environmental features. This ensures that the users' privacy is maintained, as the identities are not revealed through the transaction process in the blockchain platform.This contribution targets challenges $C7$ and $C8$ in Section 5.1.
  - To tackle challenge $C7$, we introduce a cutting-edge solution in Chapter 5, Section 5.3 - the Randomized Hash Window Dynamic Proof of Credit (WDPoC) [1, 5]. This solution addresses the security risks in blockchain platforms by implementing a randomized hash window and credit score system to monitor and regulate node behavior. The credit score is determined through indoor localization and dynamic difficulty levels are set through the randomized hash window. Nodes that follow physical distancing regulations receive a high credit score and unrestricted network access, while those that do not receive limited access. This ensures a secure and stable network environment while promoting compliance with physical distancing guidelines. Nodes have the opportunity to improve their credit score by following protocols and providing proof of compliance.
  - The inner inddor localiztion method in TB-ICT [1] model is designed to increase the precision and accuracy of indoor localization by using a Convolutional Neural Network (CNN) approach. This CNN-based framework

is designed to be efficient and reliable even in challenging indoor environments where the Line of Sight (LoS) links are not available, or the signals are affected by Additive White Gaussian Noise (AWGN) with varying Signal to Noise Ratios (SNRs). To ensure robustness in the worst-case scenario, the framework takes into account the impact of the elevation angle of the incident signal on the localization process. This is a critical aspect, as the elevation angle has a significant effect on the signal strength and quality, which can impact the accuracy of the localization process.

1.5 Organization of the Thesis

Chapter 1 (this chapter) provided an overview and a summary of important contributions made in the thesis. The reminder of the thesis is organized as follows:

- **Chapter 2** delves into a comprehensive literature review of indoor localization techniques, RL-based approaches and SR solutions. Additionally, it explores the literatures on integration of CT and blockchain technology with IoT frameworks and examines RL-based information fusion strategies.

- **Chapter 3** provides an evaluation of indoor localization techniques. It introduces the IoT-TD dataset and evaluates the performance of different methods, such as BLE-based AoA, RSSI-based localization, and IMU-based PDR localization in Section 3.3 and 3.3. In Section 3.5, the chapter presents an Indoor Localization SDK that uses ML-based models to improve the accuracy of RSSI-based localization.

- In **Chapter 4**, we present our Multi-Agent Adaptive Kalman Temporal Difference (MAK-TD) solution for MARL. Our proposed framework is detailed in Section 4.2. We also introduce a SR-based variant of MAK-TD, MAK-SR, in Section 4.3. In Section 4.4, we present a novel RL-based information fusion strategy for indoor localization to improve accuracy by fusing results from different sources.

- In **Chapter 5**, our cutting-edge, reliable blockchain-based CT solution is presented in detail. You willl find an overview of the integrated blockchain and

IoT technology in Sections 5.1 and 5.3. Section 5.2 focuses on the indoor localization aspect of the solution, while the reliability and security of the proposed framework is discussed in Section 5.4.

- In **Chapter 6**, we conclude the thesis and the remaining work will be discussed.

1.6 Publications

1.6.1 Journal Publications

J3. M. Salimibeni, Z. Hajiakhondi-Meybodi, A. Mohammadi and Y. Wang, "TB-ICT: A Trustworthy Blockchain-Enabled System for Indoor Contact Tracing in Epidemic Control", IEEE Internet of Things Journal, 2022, doi: 10.1109/JIOT.2022.3223329. In Press.

J2. M. Salimibeni, P. Malekzadeh, A. Mohammadi and K.N. Plataniotis, "MAAKF-SR: Multi-Agent Adaptive Kalman Filtering-based Successor Representation", Sensors, vol. 22, no. 4, 2022.

J1. M. Salimibeni, A. Mohammadi, Z. Hajiakhondi, M. Atashi, P. Malekzadeh, and K.N. Plataniotis, "Reinforcement Learning-based Information Fusion for Multiple Model BLE-based Indoor Localization/Tracking," Submitted to APSIPA Trans. on Signal and Information Processing (ATSIP), 2022.

1.6.2 Conference Publications

C6. M. Salimibeni, A. Mohammadi, N. Plataniotis, "RL-IFF: Reinforcement Learning based Information Fusion for Indoor Localization", Submitted to IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

C5. M. Salimibeni, Z. HajiAkhondi-Meybodi, A. Mohammadi, "TB-ICT: A Trustworthy Blockchain-Enabled System for Indoor Contact Tracing in Epidemic Control," IEEE 8th World Forum on Internet of Things (WF-IoT), 2022.

C4. M. Salimibeni, P. Malekzadeh, A. Mohammadi, A. Assa and K. N. Plataniotis, "MAKF-SR: Multi-Agent Adaptive Kalman Filtering-based Successor Representations", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 8037-8041.

C3. M. Salimibeni, M. Atashi, P. Malekzadeh, Z. HajiAkhondi-Meybodi, K. N. Plataniotis, A.Mohammadi, "IoT-TD: IoT Dataset for Multiple Model BLE-based Indoor Localization/Tracking", 28thEuropean Signal Processing Conference (EUSIPCO), 2020, pp. 1697-1701.

C2. M. Salimibeni, P. Malekzadeh, A. Mohammadi and K. N. Plataniotis, "Distributed Hybrid Kalman Temporal Differences for Reinforcement Learning", IEEE International Asilomar Conference on Signals, Systems, and Computers, 2020, pp. 579-583.

C1. M. Salimibeni, P. Malekzadeh, M. Atashi, M. Barbulescu, K. N. Plataniotis and A. Mohammadi, "Event-Triggered Monitoring/Communication of Inertial Measurement Unit for IoT Applications", IEEE SENSORS, 2019, pp. 1-4.

Chapter 2

Literature Review

The focus of this chapter is on the different approaches to BLE-based indoor localization, including fusion strategies and, in particular, Reinforcement Learning (RL)-based information fusion techniques to enhance indoor localization accuracy. This includes the analysis of different Multi-Agent Reinforcement Learning (MARL) techniques, specifically Successor Representation (SR)-based strategies, which can be used in multi-agent information fusion approaches in indoor environments. Additionally, the chapter covers one of the main applications of indoor localization, Smart Contact Tracing (SCT), and its integration with blockchain to address security concerns.The discussion in this chapter can be classified to the following main categories, namely:

- BLE-based indoor localization: BLE technology is widely used in indoor localization as it offers several advantages such as low power consumption, affordability, and ubiquity of BLE sensors. However, the main challenge in BLE-based localization is the computation of accurate dynamic estimates of the distance between sensing devices and an intended object/user via fluctuating RSSI. Later in Section 2.1 the focus would be on discussing the different BLE-based indoor localization solutions and their limitations.

- CT solutions and different SCT approaches: With the current global crisis, the need for Smart Contact Tracing (SCT) solutions has become more crucial. In Section 2.2 different CT solutions and SCT approaches and their limitations are discussed. CT solutions can be divided into centralized, decentralized, and hybrid approaches, each with its own pros and cons. Centralized solutions store user data and tracking information on third-party servers, but are at risk of

data breaches and fraud. Decentralized and hybrid solutions use a honest-but-curious server model, but still face challenges with data privacy and scalability. Additionally, there are CT platforms that integrate blockchain and IoT networks for healthcare and pandemic control, but blockchain solutions also have their own challenges to overcome.

- Blockchain-based SCT applications: In an effort to ensure privacy and security of personal information in SCT, the integration of blockchain technology is explored in Section 2.3. This section highlights the challenges and limitations of existing blockchain-based CT solutions. These solutions often struggle with scalability, computational complexity, and security issues associated with blockchain technology. Furthermore, the CT applications themselves face challenges in accurate and timely reporting, dependence on third parties, reliance on honest users, and incorporating large amounts of data.

- MARL and SR-based MARL: The field of MARL has seen a surge of interest in recent years, with its applications in various use cases, including information fusion, being increasingly recognized. In Section 2.4, we delve into the different MARL algorithms and the recent advancements in SR-based MARL algorithms.

- Multi-model RL-based information fusion indoor localization: Information fusion is an integral part of indoor localization, which involves integrating multiple sources of information to produce a more comprehensive and accurate representation of the underlying system. This can include fusing data from multiple sensors such as BLE beacons, IMUs, and GPS, and multiple processing techniques such as AoA. RL-based information fusion is a specialized approach that uses RL algorithms to optimally combine information from multiple sources. This approach allows the system to learn the optimal policy for information fusion through trial and error and adapt to changes in the environment over time. The application of RL-based information fusion in indoor localization has the potential to significantly improve the accuracy and robustness of indoor localization systems by reducing the impact of errors and fluctuations in individual sources, leading to more reliable and accurate results.The recent techniques and researches about this topic is discussed in Section 2.5.

2.1 Indoor Localization Overview

Indoor localization has become a critical research area in the Internet of Things (IoT) field due to the increasing demand for advanced and innovative indoor Location-Based Services (LBSs). The methodologies used to localize a person or device within an indoor environment differ from outdoor methodologies in terms of the deployed sensor technologies and network design. BLE is one of the most widely used technologies for indoor localization in the context of IoT [16, 19–21]. BLE is considered as a key enabling technology for IoT due to its low power consumption, affordability and ubiquity of BLE sensors [25–27].

In BLE-based localization, the RSSI is used as the key measurement to compute the distance of a user to a BLE module in an indoor environment [3,25–27]. However, the drastic fluctuations of the RSSI value lead to random errors and inaccuracies in the underlying positioning system [14,15]. The instability caused by fading effects and the multipath effects and heterogeneity of BLE-enabled devices are two main drawbacks of the RSSI-based approaches [18]. The noisy RSSI values need to be pre-processed (smoothed) to mitigate these errors [20]. Moreover, there are other challenges yet to be resolved, such as the orientation effect of the BLE packet transmitter, distinct range of RSSI in different mobile phones, and the effect of noise and interference on the BLE sensors in a noisy environment [21].

Despite the recent surge of interest and development efforts in LBSs for indoor environments, low accuracy in dynamic BLE-based distance tracking/estimation remains a major problem in creating reliable and robust indoor tracking solutions [27]. Stand-alone BLE-based localization cannot provide the level of accuracy required for indoor localization/tracking solutions. Therefore, there is an unmet need to compute accurate dynamic estimates of the distance between sensing devices and an intended object/user for LBSs in indoor environments. As stated earlier, this is due to factors such as multipath fading, drastic RSSI fluctuations, nonlinear fluctuation of RSSI values and orientation of the device [11, 29–31].

To overcome these drawbacks, recent advancements in BLE technology with the introduction of BLE v5.1 protocol have promised a prosperous future for BLE-based dynamic tracking via utilization of In-phase and Quadrature signals (I/Q samples) for angular array-processing. However, this is still yet to be materialized. As a result, there has been a surge of interest in developing hybrid (sometimes referred to as

"Multiple-Model" or "Mixture of Experts") BLE-based estimation solutions achieved by combining more than one tracking model. Examples of such hybrid solutions include coupled fingerprinting and Particle Filtering (PF) [20], Integrated K-nearest neighbour (K-NN) and Support Vector Machines (SVM) [32], Combined K-NN and trilateration [33], and Cascaded Kalman Filter-Particle Filter (KFPF) [14].

Aside from BLE, other IoT sensor technologies can be coupled with BLE to improve localization performance. One such solution is Pedestrian Dead Reckoning (PDR), which is based on data from an Inertial Measurement Unit (IMU) [9,15,34–36]. The IMU, which is available in most smartphones, consists of a 3-axis accelerometer, 3-axis magnetometer, and 3-axis gyroscope and reports variations in the movement and angular acceleration of the device [9, 36, 37]. While PDR-based solutions are prone to cumulative error in trajectory estimation, the pervasiveness of IMUs in smartphones makes PDR an attractive modality to be coupled with other indoor localization units such as BLE.

Antenna-based techniques such as Angle of Arrival (AoA), Time of Flight (ToF), and Time Difference of Flight (TDoF) have also been recognized as effective techniques to deal with indoor position estimation error. The AoA method uses triangulation to determine the two-dimensional coordinates of the transmitter with the help of at least two BLE beacons with known positions [38–41]. To calculate the angle of the incident signal, an antenna array such as the Linear Antenna Array (LAA) is required. Among all the different AoA approaches, measuring the beam-forming or spectral density of the received signal provides higher accuracy [42–44]. However, critical challenges arise, especially with frequency/phase shiftings and switching error. In recent works [45, 46], new algorithms are leveraged to improve the accuracy of indoor position estimation.

The ToF method measures the time it takes for the signal to travel from the transmitter to the receiver. It is a distance-based approach, and the estimated distance can be used to determine the location of the user [23, 24]. The ToF method is widely used for precise indoor localization, but it requires a high level of accuracy in synchronizing the clocks of the beacons. TDoF, on the other hand, is a time-based approach that measures the time difference between the signals received by multiple beacons to determine the location of the user [16, 17]. However, it requires a large number of beacons to achieve high accuracy, and the complexity of the system increases as the

number of beacons increases.

These different indoor localization approaches based on IoT sensor technologies results in different estimation accuracies. To gain a reliable localization service, one major approach is to leverage the information from multi-sensors systems to improve the system robustness, accuracy of the prediction and enhance the detection range. Information fusion is one major approach in multi-sensors IoT networks.

2.1.1 Angle of Arrival (AoA)-based Localization

Wireless technology for IoT applications faces several challenges, including high power consumption. However, BLE technology has addressed this issue by providing low-energy communications. BLE operates in the same frequency spectrum as WiFi, with 37 data channels and 3 advertisement channels, each with a bandwidth of 2 MHz. There are various indoor localization frameworks based on BLE, including RSS-based methods, ToA, and AoA [47]. AoA, also known as Direction of Arrival (DoA) estimation, is a triangulation indoor localization technique that requires at least two BLE beacons with known positions to determine the two-dimensional coordinates of the transmitter [48, 49]. An antenna array, such as the Linear Antenna Array (LAA), is required to receive the same signal with different phases and calculate the angle of the incident signal. Measuring the beam-forming or spectral density of the received signal provides higher accuracy. However, frequency/phase shiftings and switching error pose critical challenges in AoA-based frameworks. Authors in [44] determined the angle of the incident signal based on the phase difference between I/Q samples and investigated the effects of unbalanced carrier frequency.

AoA-based localization has gained attention in the research field as a reliable method for indoor tracking. Subspace-based algorithms such as MUltiple Signal Classification (MUSIC) and its extensions are among the most accurate methods, but are vulnerable to the multi-path effect, which is a common problem in indoor environments [50–52]. There have been efforts to mitigate the multi-path effect through channel classification [53], Kalman filter-based techniques [54], and subsample interpolation methods [55]. However, these methods are challenging due to their computational complexity and the requirement for a strong LoS path between the transmitter and receiver.

Recently, the focus has shifted towards data-driven approaches using Artificial

Intelligence (AI) [56, 57]. In this context, a CNN-based AoA localization method was proposed in a 2-D indoor environment, where the transmitted signal is only affected by noise [58]. However, the elevation angle of the incident signal was not considered in the localization, leading to location error. To address this issue, the authors in [59] investigated the effect of noise on the estimated angle in a 3-D indoor environment measured by DNNs. Additionally, DNN-based localization frameworks have been proposed in [56, 57, 60], where the Channel Impulse Response (CSI) is used as the input of the DNN. However, CSI is prone to noise, shadowing, and small-scale fading, leading to significant localization error.

In some recent works [45, 46], a fusion processing method is proposed to eliminate the effect of noise, frequency error, and switching error and a processing method to compensate for the lack of Non Line of Sight (NLoS) link. However, existing approaches for AoA localization via BLE sensors are stand-alone and have not yet been extended to hybrid/multi-modal settings.

2.2 CT solutions and different SCT approaches

In leveraging IoT devices and services to reduce the adverse effects of infectious diseases such as COVID-19, there should be a level of guarantee that no security or safety issues would threaten the users. Mobile application solutions developed to alleviate the negative effects of COVID-19 can be classified into the following major categories [61], i.e., CT, social distancing, symptoms/health monitoring, and telemedicine applications. With regards to CT, Alert in Canada, COVIDSafe in Australia, StopCovid in France, Corona-Warn-App in Germany, LeaveHomeSafe in China, and TraceTogether in Singapore [62] are the official government-based CT applications. For example, the latter (TraceTogether) is a centralized service that holds users' real identity/data and uses Bluetooth technology to discover and store users' proximity locally. Holding personalized data in a centralized fashion poses critical security and privacy issues. In addition, power consumption due to active BLE advertisement utilization could be problematic. There are also other CT-based solutions and social distancing applications such as Social Distancing Web Survey [63], Google/Apple joint project on a CT platform [64], COVID-19 application proposed by National Health Service (NHS) [65], and a Chinese application named Health

Code System (HCS) [66]. Unlike TraceTogether, Apple/Google platform does not hold the user's real identity/data, but a central server is used for contact profiling and sending necessary notifications. The NHS COVID-19 application has a similar vulnerability, which could result in user information leakage. The HCS, on the other hand, uses QR code-based relational cross-match, which has lower power consumption. However, the centralization of the platform and violating anonymity of users are key security issues of this platform. Considering the level of access to the users' data and management of the core services of the application, smart CT platforms can be classified into centralized, decentralized, and hybrid solutions [67]. In centralized CT platforms such as [62], users' personal data and tracking information will be stored on third-party servers, which are assumed to be safe, secure, and trusted. The level of privacy provided via such applications is user-level and default security solutions on the cloud. However, the trustworthiness of third parties managing these platforms is one major problem of such applications. Moreover, such solutions are vulnerable to data breaches and fraudulent operations on the server-side. On the other hand, decentralized and hybrid applications employ an honest-but-curious [68] server model, however, they are still prone to the available sensitive data leakage, security, and scalability issues. Moreover, although anonymous IDs are being shared in tracing close contacts [69], the real identity of the infected users can be exposed on both platforms. Hybrid solutions are also prone to third-party interventions and data leakage vulnerabilities. For example, hybrid applications such as [13, 70] manage risk analysis and notifying solutions at the server-side, which are prone to data leakage and are still suffering from high communication costs and high volume of the message exchange. There are other CT platforms developed by integration of blockchain and specified IoT networks such as Internet of Drones (IoDs) [71] to provide different services in healthcare [72] and for pandemic control [73]. In particular, Islam *et al.* [73] proposed a blockchain-enabled solution, which is an integration of IoT, AI, and blockchain, leveraging IoDs to automate a supervision scheme to monitor the crises. To address the aforementioned issues related to decentralized and hybrid solutions, there are few CT approaches designed based on blockchain technology.

2.3 Blockchain-based SCT Applications

Blockchain is a distributed ledger that uses cryptography, Public Key Infrastructure (PKI), economic modeling, and a shared consensus mechanism to synchronize a distributed database. Blockchain can be considered as a Peer-to-Peer (P2P) network, operating on a large number of distributed devices to securely store/manage information, and transfer immutable append-only transaction logs, which are signed cryptographically [74]. In a blockchain system, transactional data is stored in blocks that are linked together in a hierarchical chain [75]. Each block contains a header with several fields, including the hash value of the block, the hash value of the previous block, the nonce field for consensus algorithms, and the main body field for transactional records. The hash values of each block provide a secure chain and ensure the integrity of the transactions. Each block must be validated and mined by all the nodes in the network, after which it is added to the chain and cannot be altered [75]. The decentralized structure, fault tolerance, persistency, anonymity, transparency, and immutability of blockchain make it a suitable technology for IoT-based applications, particularly for securely storing CT information in a trustworthy and distributed manner without relying on a central processing unit [76].

As discussed earlier, several security and privacy preservation benefits can be gained via leveraging blockchain, however, these advantageous come with some extra costs [61, 77–79].

As shown in Table 2.1, while different blockchain platforms [80–84] are proposed to address the problems related to CT solutions, they suffer from the following drawbacks and limitations [85]:

• ***Integration of Proximity Approximation Algorithms with the Blockchain***: Several recent researches [81,84], mainly focused on data security and location privacy in smart CT services. For instance, Martinez *et al.* [86] used blockchain to store digitally signed pairwise encounters, which are transferred between users' handheld devices. This platform, however, only focused the Sybil attack without deployment of any localization approach. COVID-19 Risk Framework [84], is a permissionless blockchain solution that is designed to estimate the risk of being infected by COVID-19 based on mathematical calculations and probabilities. While this approach is not dependent on the third-party servers and provides trace and risk notification services, its main feature is location-based services via GPS, which is not suitable for indoor

localization. This ambiguity about indoor localization can be found in some other works [87, 88] as well.

• ***Scalability Issues***: Scalability is still a major issue in many proposed blockchain-based CT solutions, specifically in public blockchains [79]. In most related works [88–91], there is no discussion about scalability and throughput evaluation of the platform.

• ***Hybrid Solutions***: Many of the proposed blockchain-based CT solutions have hybrid designs [87,92] which keep them still dependent on the third party. Xu *et al.* [87], propose the BeepTrace, which is one of these hybrid solutions using two blockchain networks for tracing and notifications. BeepTrace utilizes different communication techniques, including BLE, GPS, and WiFi, and proposes a public key infrastructure but exploits third-party servers for geo-matching. The same issue is observed with [84] and [92]. In [84] the authors represent a unified blockchain system for proximity-based CT and location-based CT considering Bluetooth technology. In [92], a permissioned blockchain named DIMY is represented. DIMY leverages chain codes and stores the data on user devices, servers, and blockchain networks. This hybrid platform is based on BLE and uses the Deffi-Hellman key generation scheme and Bloom filters. In this approach, a server is used along with the blockchain network to provide privacy services. Sending notifications and risk analysis are the services provided in this platform, but as it is mentioned, this platform is dependent on third-party servers.

• ***Smart Contract-based Solutions***: Some blockchain-based CT solutions are mainly designed to leverage approaches based on Smart Contract [93]. These solutions are mainly costly and dependent on the other services to perform. ByChain [90], is a permissionless blockchain that is a zero-knowledge distributed database using a location-based consensus mechanism. In this approach, smart contracts are used to prove location service. This approach is highly dependent on an artificial allocation scheme exploiting IoT devices to provide claims and proofs. In [88], authors proposed a permissioned blockchain on Ethereum smart contracts based on zero-knowledge proof algorithms, which are dependent on a decentralized on-chain oracle platform to apply CT solutions. This platform provides tracing and alerting solutions; however, the cost is high. PriLok [94], is a costly permissioned blockchain that should be handled by collaboration of various entities, e.g., health and public authorities and telecommunication firms.

• ***Large Data Incorporation and Optimistic Assumptions***: One major issue

Table 2.1: Different blockchain-based CT solutions.

Solution	Communication	Cryptographic Technique	Blockchain Type	Smart Contract	Description	Limitations
BeepTrace [87]	BLE, GPS, WiFi, Cellular	Public key Scheme	Permissioned	No	A hybrid system using a certificate authority and geo-solver	Hybrid solution, leverage third party servers for geomatching
DIMY [92]	BLE	Deffi-Hellman key generation scheme and Bloom filters	Permissioned	Chain Codes	A hybrid system using blockchain for risk analysis and notification, server for privacy features	Hybrid platform, leverage a separate server to provide privacy
Algorand's approach [89]	BLE	Cryptographic Sortition	Permissionless	No	Crowd gathering monitoring and epidemiological surveillance Aggregated Data Board	Large Data Incorporation and Optimistic Assumptions, BLE-based privacy concerns
Blockchain Meets COVID-19 [84]	BLE, GPS	Distributed blockchain database	Permissionless	Permissioned	GPS Location-based using formulas to calculate infection probability	A hybrid system, an privacy preserving issues relying on randomized mac addresses
ByChain [90]	BLE, GPS, WiFi, LTE	Zero Knowledge, Distributed Blockchain	Permissionless	Yes	A location consensus based on virtual electric field	Needs incentivize users to share resources (e.g., bandwidth) for PoL service
COVID-19 CT Using Blockchain [88]	BLE, GPS	Proof of Locations	Permissioned	Yes	On Ethereum SC, zero-knowledge proof, decentralized on-chain oracle	unclear localization approach, high cost of solution
Pronto-C2 [91]	BLE	Deffi-Hellman, Blind Signature	Permissioned	No	A blockchain network as Bulletin Board	DH protocol least 256 bits required elements pass BLE limitation, needs lighter version
TB-ICT [1]	BLE	Dynamic PoW, Proof of Credit	Permissionless	No	Randomized W-Hash blockchain dynamic PoW	-

with many proposed CT schemes is the massive traffic caused by a large amount of data that should be processed and transferred [87,91]. Moreover, many solutions are designed based on optimistic assumptions that a user would be honest in uploading infection data. In [89], a public blockchain is designed to enable each user to upload information, i.e., the hash of the encounters list, about the interaction with others at certain distances and for a certain period of time as qualified encounters. However, similar to [84], this approach suffers from several privacy threats, such as deanonymization of an infected user, and the main assumption of these solutions is that the users are always honest and will upload their infection status on the blockchain.

In summary, the common problems related to blockchain and associated CT platforms can be classified into the following categories:

• ***Problems related to Blockchain-based Solutions***: Despite their benefits, IoT-based networks generally suffer from the following challenges:

(i) *Network Scalability*: Most of the designed blockchain-based solutions [95] have

limited network capability to handle large amounts of transaction data on their platform in a short span of time.

(ii) *Computational Complexity*: Another major issue of the blockchain platforms is their computational complexity [85]. This complexity represents how difficult it is to mine and add a new block to the blockchain.

(iii) *Security Issues*: Such issues are the concerns related to keeping records submitted in the blockchain immutable and the ledger tamper-resistant [77, 85, 95].

- ***Problems related to blockchain-based autonomous CT***:

(i) *Accurate and Timely Reporting*: In this context, miss-detection and miss-classification of users is a critical issue as most of the existing blockchain solutions [87, 92] fail to focus properly on providing reliable localization/proximity services.

(ii) *Dependence on Third Party*: Many of the proposed solutions [87,92] have hybrid structures, making them dependent on third parties.

(iii) *Assuming Honest Users and Deanonymization Problems*: Many solutions [84, 89] assume that the users are always honest and will correctly update their infection status on the blockchain. Another issue is suffering from privacy threats due to deanonymization of an infected user.

(iv) *Large Data Incorporation*: It is common [79] to analyze, process and add several unnecessary data to the blockchain transactions to be mined. This, in turn, decreases the throughput of the existing solutions.

2.4 MARL and SR-based MARL

Traditionally, RL algorithms are classified as (i) Model-Free (MF) approaches [28, 96, 97] where sample trajectories are exploited for learning the value function, and (ii) Model-Based (MB) techniques [98] where reward functions are estimated by leveraging search trees or dynamic programming [99]. MF methods, generally, do not adapt quickly to local changes in the reward function. On the other hand, MB techniques can adapt quickly to changes in the environment, but this comes with a high computational cost [100–102]. To address the above adaptation problems, SR

approaches [103, 104] are proposed as an alternative RL category. The SR method provides the flexibility of the MB algorithm and has computational efficiency comparable to that of the MF algorithms. In SR-based methods, both the immediate reward expected to be received after each action and the discounted expected future state occupancy (which is called the SR) are learned. Afterwards, in each of the successor states, the value function is factorized into the SR and the immediate reward. This factorization only needs learning of the reward function for new tasks, allowing rapid policy evaluation when reward conditions are changed. In scenarios with a limited number of states, the SR and the reward function (thus, the value function) associated with each state can be readily computed. Computation of the value function, however, is infeasible for MARL problems, as in such scenarios we deal with a large number of continuous states [105]. In other words, conventional approaches developed for single agent scenarios such as single-agent SR, Q-Learning, or policy gradient cannot be directly adopted to MARL to compute the value function. The main problem here is that, typically, from a single agent's perspective, the environment tends to become unstable as each agent's policies change during the training process. In the context of deep Q-learning [106], this leads to stabilization issues as it is difficult to properly use the previous localized experiences. From the perspective of policy gradient, typically, observations demonstrate high variance in coordinating multiple agents.

To leverage SR-based solutions for MARL, value function approximation is unavoidable, and one can use either linear or non-linear estimation approaches [107,108]. In both categories, a set of adjustable parameters define the value of the approximated function. Non-linear function approximators, such as Deep Neural Networks (DNNs) [108–111], have enabled application of RL methods to complex multi-agent scenarios. While DNN approaches like Deep Q-Networks (DQN) [112] and Deep Deterministic Policy Gradient (DDPG) [113] achieved superior results, they suffer from some major disadvantages including the overfitting problem, high sensitivity in choosing parameters, sample inefficiency, and high number of episodes required for training the models. The linear function approximators, on the other hand, transform the approximation problem into a weight calculation problem in order to fuse several local estimators. Convergence can be examined when linear function approximators are utilized, as they are better understood than their non-linear counterparts [114, 115].

Cerebellar Model Articulation Controllers (CMACs) [116] and Radial Basis Functions (RBFs) [117] are usually used as linear estimators in this context. It has been shown, however, that the function approximation process can be better represented via gradual-continuous transitions [118]. Albeit the computation of the RBFs' parameters is usually based on prior knowledge of the problem at hand, these parameters can also be adapted leveraging observed transitions in order to improve the autonomy of the approach. In this context, cross entropy and gradient descent methods [119] can be utilized for the adaptation task. Stability of the gradient descent-based approach was later improved by exploiting a restrictive method in [118], which is adopted in this manuscript.

After verifying the value function's structure, to train the value function approximator, the following methodologies can be used: (i) Bootstrapping methods, e.g., Fixed-Point Kalman Filter (FPKF) [120]; (ii) Residual techniques such as Kalman Temporal Difference (KTD) and Gaussian Process Temporal Difference (GPTD) [121], which is a special form of the KTD; and (iii) Projected fixed-point methods such as Least Square Temporal Difference (LSTD) [122]. Among these methodologies, KTD [123, 124] is a prominent technique as, based on the selected structure, it provides both uncertainty and Minimum Mean Square Error (MMSE) approximation of the value function. In particular, uncertainty is beneficial for achieving higher sample efficiency. The KTD approach, however, requires prior knowledge of the filter's parameters (e.g., noise covariance of the process and measurement models), which are not readily available in realistic circumstances. Parameter estimation is a well-studied problem within the context of Kalman Filtering (KF), where several adaptive schemes are developed over the years including but not limited to Multiple Model Adaptive Estimation (MMAE) methods [125–127] and, innovation-based adaptive schemes [128]. When the system's mode is changing, the latter has the superiority to adapt faster and its efficiency was shown in [129], where different suggested averaging and weighting patterns were compared. MMAE methods were already utilized in the RL problems, for instance, Reference [130] proposed a multiple model KTD coupled with a model selection mechanism to address issues related to the parameter uncertainty. Existing multiple model methodologies are, however, not easily generalizable to the MARL problem.

In methods proposed in [101, 131–133], while the classical TD learning is coupled

with DNNs, uncertainty of the value function and that of the SR is not studied. To deal with uncertainty, a good combination of exploitation and exploration should be used to prevent the agent's overconfidence about its knowledge to fully rely on exploitation. Alternatively, an agent can perform exploration over other possible actions, which might lead to improved results and a reduction in the uncertainty. Although, from computation points of view, it is intractable to find an optimal trade-off between exploitation and exploration, it has been represented that exploration can benefit from the uncertainty in two separate ways, i.e., through added randomness to the value function, and via shifting towards uncertain action selection [134]. Consequently, the approximated value function's uncertainty, is a beneficial information for resolving the available conflict between exploration and exploitation [123, 134]. It was shown in [123] that the sensitivity of the framework to the parameters of the model can be diminished via uncertainty incorporation within the KTD method. Therefore, the required time and memory to find/learn the best model will be reduced compared to DNN-based methods [101, 131–133]. The reduced sensitivity in setting the parameters enhances the reproducibility feature of a reliable approach, which leads to regeneration of more consistent outputs while running multiple learning epochs. Consequently, the risk of getting unacceptable results in real scenarios will decrease [135]. Geerts et al. [103] leveraged the KTD framework to estimate the SR for problems with discrete state-spaces, however information related to uncertainty of the estimated SR is not considered in the action selection procedure. We have started our research on signal processing-based RL solutions by introducing the MM-KTD [8, 28], which is a multiple model Kalman temporal difference approach for single-agent environments with continuous state-space. The AKF-SR is then proposed in [136], which is an adaptive KF-based SR approach developed for single-agent scenarios.

2.4.1 Reinforcement Learning (RL)-based Signal Processing

Increasingly tendency of the researchers to imitate human behavior let the astonishing research area flourishing to meet this goal. One of the most prosperous fields aiming to learn from the history of the interaction between the agent and the environment is Reinforcement Learning (RL) [134]. Generally speaking, RL is a class of Machine Learning (ML) algorithms enabling autonomous agents to learn the optimal control (action) policy by using trial and error based on the feedback received from the

environment after each action [6, 28]. Unlike supervised learning, RL approaches do not need labeled data, which is challenging to acquire in most practical scenarios. The RL methods are utterly beneficial as they leverage the feedback [137–140] provided by the reward received by each agents' movement in the environment. This procedure will lead to policy optimality to choose the best action to increase the reward gained by the nodes in the system, which inherently minimizes the error. Traditionally, RL algorithms are categorized into two main classes: (i) Model-Free (MF) methods [6,28, 141, 142], which learn the value function using sample trajectories, and; (ii) Model-Based (MB) methods [143] that estimate transition and reward functions through search trees or dynamic programming [144]. MB approaches focus on learning the model of the environment, leveraging the past status of the system. However, in MF-based algorithms, the goal is to find the optimal policy in the system without any attempts to learn the system's dynamics. Algorithms belonging to the former category (MF algorithms) typically fail to rapidly adjust an agent to localized changes in the reward function. The MB algorithms can quickly adjust to the environmental changes but with a high computational cost [100, 101].

2.4.2 Successor Representations Approaches

As a remedy to the aforementioned adaptation problems, SR methods [6, 103, 145] are proposed recently as an alternative class of RL algorithms. The SR methods provide computational efficiency comparable to that of MF algorithms, concurrently with the flexibility of MB algorithms. The SR-based algorithms learn the expected immediate reward received after each action together with the expected discounted future state occupancy (i.e., the SR). SR-based approaches are mainly developed for single-agent scenarios and have not yet been extended to multi-agent scenarios. This proposal addresses this gap for potential applications to predictive analytic and demand response in SCs. This SR-based solution is also compared with an innovative multiple model adaptive Kalman filtering methods employed as another alternative to deep RL solution for the multi-agent scenarios.

Considering the number of agents in an environment, there are two kinds of environments, i.e., single-agent and multi-agent environments. In single-agent systems with a relatively limited number of status and actions, Monte-Carlo [146] is one common approach to calculate the state action or value function for each state using some

sample runs to find the optimal strategy. On the other hand, the temporal difference(TD) [147] approaches emerge, making use of bootstrapping to update the value during an episode in an online fashion optimization. Due to the curse of dimensionality [148], all these approaches become useless in a more complicated environment. Combining the Q-learning [149] and the basic RL approaches with the artificial neural networks proposes an alternative solution to approximate the state action/value function over the entire state-space [108–111]. Most early attempts in the Deep RL did not achieve acceptable results due to overfitting. However, this problem was mostly overcome in the further attempts [150] and also was extended to the environments with the continuous action space [113]. Albeit there are many signs of progress to fit the Deep RL model, e.g., Deep Q-Network (DQN) [151] to the single-agent and multi-agent scenarios, these models are still prone to overfitting and high sensitivity in parameter selection and sample inefficiency. Converting the value function approximation problem to a weight estimation approach by applying a set of weighted local estimators is another well-practiced approach to address the previously mentioned problems. Various local estimators were proposed in the literature, among which Radial Basis Functions (RBFs) [152], and Cerebellar Model Articulation Controllers (CMACs) [116] are most popular. It was shown that RBFs are more suitable than CMACs in systems with continuous states due to their continuous nature [153]. More recently, the Fourier basis was proposed as the local estimator function; however, the performance of the system was shown to be comparable to those using RBFs [154]. The parameters of the RBFs are usually computed based on the knowledge of the problem. However, it is possible to adapt these parameters using the observed transitions to enhance the autonomy of the method. Cross entropy and gradient descent methods were proposed by [155] for that matter. The stability of the latter was later enhanced using a restrictive technique in [118]. By determining the structure of the value function, to gradually minimize the error between the approximation and real values received in all status, different methods are proposed in the literature. These algorithms are categorized as bootstrapping [120], residual [121], and projected fixed-point [122] approaches, and are compared thoroughly in [123]. Kalman Temporal difference (KTD) [124] stands out to be a better solution regarding its Minimum Mean Square Error (MMSE) in value function estimation and the uncertainty in [156] terms of their error covariance that can be used to gain higher sample efficiency [124].

To estimate the parameters of the Kalman filters, two main approaches can be categorized as innovation-based adaptive methods [128] and multiple model adaptive schemes [157]. The latter approach is a fast and adaptable solution, and its different averaging and weighting schemes are mainly discussed in [129]. Additionally, in [156] the system's uncertainty of the proposed model-based multiple model approach was challenged exploiting different models.

2.5 Multi-model RL-based Information Fusion Indoor Localization

Technically, information fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making [158]. Information fusion techniques can be categorized into traditional methods and Machine Learning (ML)-based approaches. Probabilistic fusion, evidential belief reasoning fusion, fuzzy theory, and tensor fusion are among the main traditional information fusion strategies [159]. Many progresses in different fields including data gathering techniques, processing hardware and different data steaming and processing solutions pave the path to leverage new information fusion approaches. In waht follows we will learn about common solutions in traditional information fusion and ML-based approaches.

2.5.1 Traditional Information Fusion Approaches

Generally, information fusion is widely used in diverse applications and fields including but not limited to IoT-based services, wireless sensor networks, localization/tracking services, image processing, navigation and radar systems, different assessment techniques and, etc [160]. Among the traditional information fusion techniques, four typical categories are (i) probabilistic fusion e.g., Bayesian solutions, (ii) evidential belief reasoning fusion e.g., Dempster/Shafer (D-S) evidence theory, (iii) fuzzy theory, and (iv) tensor fusion methods [159]. The main focus of the probabilistic fusion is to get the inference results by combining the observed data and prior probability [161].

In [162], a Bayesian fusion method is proposed with a RL-based multislot double-threshold spectrum sensing to gather industrial big data, capable of fast recognition of required idle channels while guaranteeing spectrum sensing performance. A fuzzy multi-entity Bayesian network is proposed in [163] to fuse data generated from human-based sources with those generated by physical sensors. Information fusion using D-S evidence theory are more flexible than Bayesian solutions since uncertain information are described by leveraging interval estimation rather than point approximation. In [164] a D-S evidence theory is applied as the fusion algorithm and an enhanced belief divergence measurement solution is proposed in [165] to address the high conflict issue in D-S theory-based information fusion. Solving this highly conflicting evidence issue is also targeted in [166] by proposing a new divergence measurement to quantify the differences between basic probability assignments (BPAs). Some undefined problems in information fusion is addressed by fuzzy theory [167–169] by modeling the fusion approach in a loose way. In [167], a fusion solution is proposed based on fuzzy and D-S evidence theory to solve the multi-dimensional data fusion at the decision layer. Another fuzzy-theory based information fusion approach is represented in [168], where the fusion of visual data, signals, and multidimensional statistical data is used to build intelligent systems for diagnostics and decision support. There are also many approaches that utilize tensor fusion methods regarding the tensors' powerful capabilities in information representation. This method is proposed in [170] to address the spatial data processing issue by exploiting a cyber-physical-social transition tensor (CPST2) in a multi-step transition fusion model. Although the aforementioned traditional information fusion techniques are proposed to address specific demands in many concrete applications, there are still important challenges that should be tackled in data fusion solutions to maximize its benefits [158]. These challenges are mostly stem from from application environments' complexity where sensors are located, the variety of data with different types that should be combined, and so on [160]. These complexities can be classified as (i) Data imperfection, (ii) Data inconsistency, (iii) Data confliction, (iv) Data alignment/registration and correlation, (v) Data type heterogeneity, (vi) Fusion location and (vii) Dynamic fusion [160]. ML-based information fusion approaches can be a solution to address these drawbacks. In section 2.5.2, we breifly discuss the common approaches in RL-based information fusion.

2.5.2 RL-based Information Fusion Approaches

Machine Learning (ML) is a powerful technique that allows systems to learn from the provided data without the need for programming the problem or the dynamics of the environment. ML is designed to create a knowledge base based on the robust relationships between input data and to use the learning procedure's output to estimate, predict, or classify data. This makes ML an ideal technique for information fusion approaches [160, 171].

However, the implementation of ML-based information fusion approaches presents several challenges, such as data labeling and the time-consuming and costly process of training the system to fuse data from multiple sources [10, 172]. These challenges are addressed by incorporating Reinforcement Learning (RL) techniques into the information fusion process [163, 173, 174]. RL is considered to be one of the best solutions for information fusion due to its ability to alleviate the computational and cost-wise processing of information fusion [161, 163, 174].

RL is a class of ML techniques that aims to provide human-level adaptive behavior by constructing an optimal control policy [134]. The objective is to learn from previous interactions between the autonomous agent and its environment through trial and error. The optimal control policy is obtained through RL algorithms and the feedback provided by the environment after each action taken by the agent [2, 6, 8, 28, 137, 138].

There have been various studies proposing the use of Reinforcement Learning (RL) in information fusion. For example, Guo et al. proposed a deep RL multi-modal decision-making fusion weight allocation method in [163] to overcome the limitations of traditional weighted fusion methods with fixed weight allocation in decision level fusion. Liu et al. proposed a deep RL multi-type data fusion framework for algorithmic trading services in [175]. This approach defines a static Markov Decision Process (MDP) for the fusion strategy and utilizes RL mainly for making trading decisions based on temporal features of stock data and other key performance indicators (KPIs). In [174], a priori knowledge RL-based information fusion method for multi-sensors in air combat data is proposed, where RL is used to find the coefficients for the fusion of sensory data inputs.

While some of these works may lack timeliness and are not applied to localization purposes, the actual indoor localization task requires a time-varying system.

Although several sensor fusion localization approaches have been proposed in literature [176–178], they mostly rely on traditional sensor fusion techniques or supervised ML techniques. To address the challenges related to data labeling, the complexity of mathematical modeling of indoor environments, and the time-varying nature of information from different sensors, an RL-based information fusion method was employed to fuse the localization results from different localization solutions in different indoor venues [176].

Considering the difficulties in data labeling, complexity of mathematical modeling of indoor environments, and time-varying nature of the information received from different sensors, we propose the use of an RL-based information fusion method to fuse the localization results obtained from different localization solutions in different indoor environments. This approach allows for a more flexible and adaptive solution to handle the complex and dynamic nature of indoor localization tasks.

2.6 Conclusion

In conclusion, this chapter focuses on the different approaches to BLE-based indoor localization, including the use of RL-based information fusion techniques. The chapter also covers the integration of blockchain technology with Smart Contact Tracing (SCT) to address privacy and security concerns. The different sections of this chapter cover the limitations of BLE-based indoor localization (Section 2.1), different CT solutions and SCT approaches (Section 2.2), blockchain-based SCT applications(discussed in Section 2.3), Multi-Agent Reinforcement Learning (MARL) and SR-based MARL algorithms (Section 2.4), and the recent advancements in RL-based information fusion in indoor localization(Section 2.5). Overall, the chapter provides a comprehensive overview of the current state of indoor localization and the challenges and limitations associated with it.

Chapter 3

IoT-based Indoor Localization

This chapter tackles the first objective of the thesis, providing an investigation into various BLE-based indoor localization solutions. As an extension of the overarching objective, we delve into the world of IoT-based indoor localization, focusing particularly on the potential of BLE technology. Our analysis includes an examination of BLE-based indoor localization techniques using a localization dataset, the IoT-TD Dataset, presented in Section 3.1. This chapter sets the stage for exploring multi-model solutions, as indicated in the introduction, by first understanding the strengths and weaknesses of existing single-model techniques.

In Section 3.2, we offer a detailed examination of BLE technology, its features, and the concept of BLE beacons. Furthermore, in Section 3.3, we delve into different BLE-based indoor localization solutions, analyzing their strengths and limitations. Additionally, we explore the role of IMU in indoor positioning and their integration with information fusion technology in Chapter 3.4 to enhance indoor localization accuracy.

This chapter serves as the foundation for our study, as it provides crucial insights into the current state-of-the-art in BLE-based indoor localization and sets the stage for the development of a more advanced and accurate solution in the following chapters.

3.1 The IoT-TD Dataset

The IoT-TD dataset is critical to the development of effective indoor localization solutions. This dataset provides robust and accurate location-based services (LBS) in indoor environments, especially in the context of the COVID-19 pandemic and the need for efficient indoor contact tracing frameworks.

The IoT-TD dataset overcomes the limitations of existing datasets that lack ground truth information, making it challenging to accurately evaluate and compare different algorithms. With real-world data, the IoT-TD dataset enables the development of more accurate and reliable indoor localization solutions by providing the necessary information for training and testing algorithms. Additionally, the use of multiple sensors, including IMU sensor measurements, provides a comprehensive view of the indoor environment, leading to the development of more robust and effective indoor localization solutions.

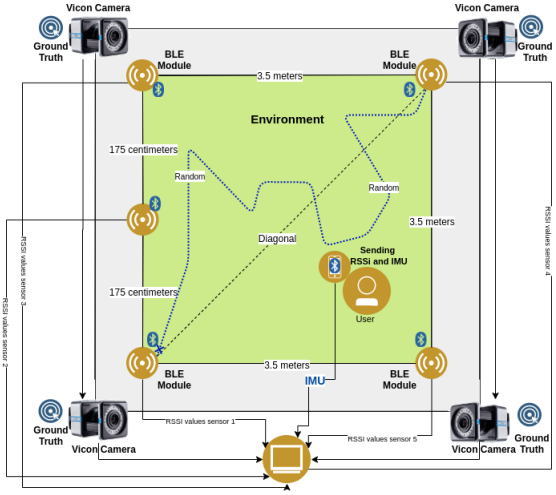


Figure 3.1: Experimental setup for collection of the IoT-TD dataset in one of the 3 environments.

The IoT-TD dataset includes synchronized ground truth trajectories, RSSI values collected in a multi-sensor setting, and IMU sensor measurements obtained from a hand-held device carried by the moving target. All three components of the dataset are time-stamped and pre-processed, ensuring the highest level of accuracy and reliability.

Figure 3.2: Data collection setup in three different environments. Top two figures are from the first environment, while the bottom figures show the other two environments.

3.1.1 Experiment Setup

IOTD Dataset

SessionDate	SessionDuration	SessionFrequencv	RecordID	Timestamp	timeIntervalSince1970	GvroX	GvroY	GvroZ	AccX	AccY	AccZ	MaoX	MaqY	MaqZ
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.1370	1573333169.1374	0.030554180964828	-0.116176404058933	0.033809095621109	-0.305419921875	-0.072509765625	-0.949966430664062	250.543334
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.1460	1573333169.14614	0.030554180964828	-0.116176404058933	0.033809095621109	-0.156112670898438	-0.058441162109375	-1.000671386718750	204.931030
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.1730	1573333169.17297	0.030554180964828	-0.116176404058933	0.033809095621109	-0.146179199218750	-0.054122924804688	-0.986892700195312	204.838500
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.1890	1573333169.18923	0.030554180964828	-0.116176404058933	0.033809095621109	-0.145889282226563	-0.046920776367188	-0.988983154296875	205.310424
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.2060	1573333169.20557	0.030554180964828	-0.116176404058933	0.033809095621109	-0.142837524414063	-0.053985595703125	-0.987731933593750	205.252471
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.2070	1573333169.20721	0.030554180964828	-0.116176404058933	0.033809095621109	-0.142837524414063	-0.053985595703125	-0.987731933593750	205.252471
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.2220	1573333169.22219	0.030554180964828	-0.116176404058933	0.033809095621109	-0.134170532226563	-0.064407348632813	-0.984695434570312	204.772033
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.2400	1573333169.24035	0.000564854645473	0.068591840565205	0.150881662964821	-0.161590576171875	-0.037551879882813	-0.993804931640625	204.772033
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.2570	1573333169.25698	-0.009303220547736	0.013185310177505	0.030754717066884	-0.172180175781250	-0.029373168945313	-0.993103027343750	205.500045
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.2720	1573333169.27249	-0.012896090745926	-0.002075934084132	-0.004255997482687	-0.152923583984375	-0.040069580078125	-0.987213134765625	205.81413
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.2890	1573333169.28884	-0.012609801255167	0.045835133641958	0.035332690924406	-0.146072387695313	-0.048141479492188	-0.979156494140625	205.528198
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3080	1573333169.3075	-0.016544355079532	0.087029710412026	0.16211673617363	-0.133224487304688	-0.066299438476563	-0.983139038085938	207.147018
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3230	1573333169.3226	-0.016058595851064	0.107981860637665	0.206176608800888	-0.147933959960938	-0.050033569335938	-0.975036621093750	207.982315
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3390	1573333169.33941	-0.022368954494596	0.076357886195183	0.167073145508766	-0.156616210937500	-0.033798217773438	-0.987976074218750	207.915145
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3550	1573333169.3555	-0.030541663989425	0.031610123813152	0.107769869267941	-0.153076171875	-0.036331176757813	-0.987258911132812	207.915145
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3570	1573333169.35693	-0.035922046750784	-0.010379936546087	0.066749468445778	-0.153076171875	-0.036331176757813	-0.987258911132812	207.915145
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3730	1573333169.37342	-0.040039826184511	-0.025873143225908	0.055858206003904	-0.146957397460938	-0.060684204101563	-0.979598999023438	208.392700
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3760	1573333169.37641	-0.040039826184511	-0.025873143225908	0.055858206003904	-0.146957397460938	-0.060684204101563	-0.979598999023438	208.392700
2019-11-09	15:59:29.1260	00:05	100	0	2019-11-09	15:59:29.3900	1573333169.3905	-0.041357558220625	-0.010879545472562	0.079249545931816	-0.152145385742188	-0.069290161132813	-0.98095703125	208.959777

IMU Raw Data

time_sample	X_Location	Y_Location	sensor_10	sensor_8	sensor_5	sensor_13	sensor_4
0.1second	0	0	-55	-65.115966254987	-67.0726612268033	-65.115966254987	-65.7458813395397
	0.097928342514837	0.100059497316259	-55.5606929286222	-65.021796822513	-66.9787338713368	-65.023869441981	-65.6335249205866
	0.193297100498636	0.200491549819873	-56.1039221309522	-64.927219598964	-66.8842822856075	-64.934448224893	-65.5210544966112
	0.288272523694077	0.302051840085028	-56.6252549825837	-64.831556674653	-66.7878620208248	-64.845860711834	-65.4065187333646
	0.38434961172343 8	0.40294520157055	-57.1158249468777	-64.736754614686	-66.6895480377352	-64.756697667532	-65.2893339773279
	0.479729241380428	0.502918191929442	-57.5740520811148	-64.643136512292	-66.5903217516397	-64.668818198366	-65.1710067438221
	0.573746896116381	0.602402906510663	-58.0025488460996	-64.550323261212	-66.4902878631197	-64.58307934723	-65.0519413824495
	0.669490564515172	0.700874516213274	-58.4053590905676	-64.459357473577	-66.3880681690897	-64.49637862165	-64.9296202251651
	0.766755844128767	0.799689307895586	-58.7884296192 07	-64.369092816067	-66.2829530985517	-64.409182254855	-64.8035533361787
	0.865002460328879	0.900514813142268	-59.1567824826091	-64.277918905776	-66.1741956814146	-64.3225307341 65	-64.6733825920002
	0.963799361729374	1.001109795430 67	-59.5054223155441	-64.188389232488	-66.063121058363	-64.236738112149	-64.5403065409789
	1.06343645285721	1.098732574242 95	-59.8305580641073	-64.103869493868	-65.9510330001249	-64.151008397373	-64.4049618146607
	1.16559710834996	1.194659044164 7	-60.1392065110031	-64.023803182071	-65.8361821043632	-64.063783569335	-64.2649875566947
	1.27081576794693	1.289025334782 48	-60.4338877524968	-63.948729176779	-65.718046489332	-63.974520020027	-64.1194595796793
	1.37645803126017	1.383560405458 5	-60.7163586306843	-63.876420269292	-65.5970310566944	-63.886768903913	-63.9701986474302
	1.47933243301172	1.479926284217 39	-60.9870576248598	-63.804138361774	-65.4739575772116	-63.805027573802	-63.820014195178
	1.58117239844953	1.576892902110 18	-61.2462995005891	-63.733945385203	-65.3484206307551	-63.727367389746	-63.6675737964451
	1.68312002063004	1.676000044247 6	-61.4982688085512	-63.664768968018	-65.218560563433	-63.653544151553	-63.5109944105312
	1.78426705496256	1.777195234976 79	-61.7423455203452	-63.596533614992	-65.0847805522833	-63.585111545384	-63.3513877099423

RSSI Raw Data

		Global		Angle	wand_v2:wand_v2				
Frame	Sub	Frame	RX	RY	RZ	TX	TY	TZ	
		rad		rad	rad	mm	mm	mm	
1	0	-0.00669719	0.0839872	0.690167	-1527.93	682.343	231.249		
2	0	-0.00525841	0.0808313	0.691423	-1527.58	682.107	231.605		
3	0	-0.00544113	0.0769393	0.693239	-1527.35	682.091	232.076		
4	0	-0.00388878	0.0756825	0.694474	-1527.16	681.997	232.486		
5	0	-0.00295042	0.0753683	0.695591	-1527.07	682.024	232.824		
6	0	-0.00303624	0.0761926	0.695443	-1527.03	682.184	233.096		
7	0	-0.00429717	0.0772653	0.694735	-1526.93	682.311	233.339		
8	0	-0.00439835	0.0786323	0.694864	-1526.8	682.333	233.546		
9	0	-0.00525626	0.0791608	0.694961	-1526.68	682.401	233.652		
10	0	-0.00603072	0.0794121	0.694883	-1526.51	682.471	233.757		
11	0	-0.00683004	0.0798764	0.694909	-1526.33	682.495	233.859		
12	0	-0.00741131	0.0799232	0.69506	-1526.19	682.511	233.914		
13	0	-0.00758209	0.0797761	0.695343	-1526.07	682.545	233.911		
14	0	-0.00755282	0.0791645	0.696173	-1525.99	682.431	233.9		
15	0	-0.00750999	0.0788913	0.695922	-1525.82	682.509	233.841		
16	0	-0.00737974	0.0789611	0.695641	-1525.63	682.534	233.733		
17	0	-0.00656902	0.0791613	0.69571	-1525.44	682.499	233.53		
18	0	-0.00567095	0.0793665	0.695762	-1525.26	682.489	233.316		
19	0	-0.00458993	0.0793066	0.69548	-1525.07	682.481	233.104		
20	0	-0.00382384	0.0791414	0.695426	-1524.91	682.447	232.922		

Ground Truth Raw Data

Figure 3.3: Raw data samples from the IoT-TD dataset.

The IoT-TD,[1] is constructed based on three thoroughly separated indoor environments to incorporate effects of various interferences faced in different spaces. The main objective is to have a dataset to potentially develop multiple-model implementations to fuse the three different localization approaches (i.e., RSSI-based, PDR, and AoA-based). The first environment is shown in Fig. 3.1, illustrating the data collection setup. In these data-gathering tasks, data is synchronously collected from five different BLE modules together with I/Q samples for AoA, and IMU sensor data from the target's handheld phone. The central processing unit is responsible for data collection. The IoT-TD dataset was collected through more than 600 data-gathering sessions. Each environment, at least, has 200 data-gathering sessions. 50 different data collection sessions were devoted to each of all the four different moving scenarios, for all three tracking paths including time-stamped RSSI data received from five BLE modules, IMU built-in target's device sensors, and I/Q samples. The related ground truth of the tracked user is also gathered in all experiments. During each tracking epoch, along with the IMU sensor data, which consists of 9 different parameters, the RSSI values received by all BLE modules are captured. Simultaneously 6 parameters consisting of 3-D position data and three rates of angular velocity (pitch, roll, and yaw) are received from Vicon cameras as the target's true flight position. For this massive amount of data gathered in each epoch, I/Q samples related to the signal transmitted from the target's phone to each BLE module is calculated in an offline mode and utilized to extract the AoA of signals at each synchronous moment. Fig. 3.3 shows raw data samples from the IoT-TD dataset.

In order to consider different trajectory movements, three tracking scenarios are considered, i.e., rectangular, diagonal, and random movements of the user. More accurately, the three tracking scenarios are as follows: (i) *Rectangular Walking*, is a movement pattern where the user walks continuously in the sides of the rectangular area of each environment in a pre-defined path; (ii) *Diagonal Walking*, where the user walks in the diagonals of each location. Related data for two different diagonal movements are gathered, and; (iii) *Random Walking*, where the user walks randomly in the surveillance region of all three environments. The rectangular and the diagonal movement trajectories (Scenarios (i) and (ii)) are more predictable and rhythmic, therefore, provide more straightforward tracking scenarios. The situation is different

[1]The constructed dataset, together with its description, can be accessed freely through its webpage: https://github.com/MSBeni/IoT_Journal_Dataset.

in the third scenario (random movement). This trajectory in all the venues is expected to be more unpredictable.

3.1.2 Construction of the Ground Truth

Having access to the reliable and accurate ground truth in real-time is a key feature in evaluating the accuracy of implemented localization/tracking algorithms. As shown in Fig. 3.2, the Vicon Vero cameras (VICON Blade, VICON Motion Systems, UK) are being used to achieve this goal. The specific optical cameras in Vicon leads to tracking/localizing a target with a millimeter accuracy. To be more precise, the Vicon Motion Capture System, utilized in this work, consists of four Vicon Vero infrared 1.3 megapixel cameras with a sampling rate of 100 Hz. Moreover, to manage the data gathering sessions, the Vicon Data Stream (SDK), specifically the Vicon Tracker v.3.7 is used to construct the ground truth trajectories. After calibrating the cameras, the installed onboard sensors of the Vicon System are used to control the cameras' timely performance to ensure the accuracy and reliability of the cameras' data and enables the system to timely monitor the Vicon cameras' position. The cameras' calibration for each data collection session is the vital initial part of the experiments to prevent any possible mistake threatening the accuracy of the collected ground truth data. The user's phone's ground truth consists of the time-stamped position in 3-D, (x, y, z) and the rate of angular velocity (pitch, roll, and yaw). The real-time 3-D location information is essential for evaluating the accuracy of tracking models. Furthermore, to check the validity of the results and the models developed based on IMUs data, the rate of angular velocity data received from the Vicon System can be used for comparison purposes.

3.1.3 AoA Dataset Based on The Ground Truth

In the latest version of the BLE technology, which is the Bluetooth Core Specification v.5.1, the enhancement of the Low Energy (LE) controller layer enables generation and transmission of raw direction-finding data in the BLE-based networks [179]. The Constant Tone Extension (CTE) of the LE Link Layer is added to the transmission packet, which is sent at the carrier frequency of 250 kHz, enabling the I/Q sampling capability. In order to use this new feature, several issues are still in place, which

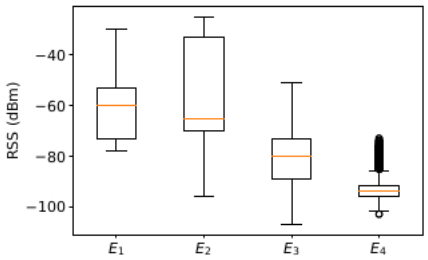
are expected to be addressed and probably solved in the future profile specifications presented by the Bluetooth Special Interest Group (SIG). It is, therefore, essential to start researching this new feature and check its potentials. Toward this goal, we simulate the received signals from the user's phone to BLE modules and calculate raw I/Q samples based on the user's real location, provided by Vicon cameras.

In the first step, the AoA is estimated according to the user's exact location, provided by Vicon cameras, and the BLE modules' installation coordinates. Then, based on the estimated AoA values, I/Q samples are generated. Since five BLE modules are installed in the experimental environments, five sets of raw I/Q samples for each experiment are obtained. Additive White Gaussian Noise (AWGN) is also fed to the constructed signals to consider the fluctuations of the real scenarios. Based on the generated I/Q samples, the user's position, which is the intersection area of lines drawn from each BLE beacon, is calculated. Note that the angle between lines from BLE beacons and X-axis is determined based on the estimated AoA of the received signal. It is assumed that the user's final location is calculated based on two installed BLE beacons in the environment to keep the processes fruitful and straightforward for the other fusion scenarios and other localization approaches.

3.2 BLE Technology

BLE is introduced by Bluetooth Special Interest Group in 2010 and is designed to be used in applications with small power consumption beside no requirement to send large amounts of the data [179]. BLE is an appropriate connection network for periodic transmission of small amounts of data over the 2.4 GHz ISM band [12, 13]. The low power consumption of BLE makes this technology highly available in many IoT devices, specifically in almost every smartphone. Many new devices like BLE beacons have been created to be used based on this technology, make BLE a suitable choice for many smart services, specifically the localization approach. BLE and WiFi are working in the same radio frequency bands, but since the BLE advertisement is just through three channels $(37, 38, and 39)$ which are widely spaced at $2402, 2426, and 2480$ MHz, there would be no interference between BLE and popular WiFi channels. BLE has two modes of communication, connectable and non-connectable. These two modes of communication can be used for different use cases. In the localization and SCT

case, the non-connectable mode is the focal point since it provides the possibility of a more secure connection. In the non-connectable advertisement mode, there is no need for the receiver to accept the incoming connection request, which will prevent any unauthorized access to the users' sensitive data. The periodic propagation of the advertisement packets by BLE modules in non-connectable mode makes it possible for the devices equipped with BLE, e.g., smartphones, to hear and receive these packets within the broadcast range. According to the system-defined advertising interval, these advertisement packets are broadcasted periodically through the BLE advertisement channels. The RSS can also be measured by smartphones while receiving the packets. Regarding this communication paradigm, two major challenges ought to be considered:



E	Environment Type
E_1	Room
E_2	Corridor
E_3	Indoor Open Area
E_4	Outdoor

Figure 3.4: The variation of RSS values in different environments.

1. **Short Length of the Advertisement Packets** that is only limited to 47 bytes, and only 31 bytes of this space can be used to include information that can be a challenge in the context of SCT.

2. **High Fluctuations of the RSSI Values**, while the RSSI value (in dBm) is inversely proportional to the square of the distance, $RSSI \propto \frac{1}{d^n}$, where d is the distance between two devices and n is the path loss exponent and is an environmental factor varies in different environments and ought to be considered. Fig. 3.4 shows the effect of the environment in RSS values. Using filtering capabilities can help smooth RSSI values and gain better final localization results. Two simple, mainly used filters are the Kalman filter and moving average filter.

3.2.1 BLE Beacons Features

As discussed earlier, BLE beacons, known as beacons which can have different protocols, i.g., ibeacon, are highly affordable, have a small size, and very low power consumption [12]. Beacons, utilizing only advertising mode, periodically broadcast packets of data in intervals from $20ms$ up to $10s$, which can be received with any BLE-enabled device like smartphones. Longer intervals provide more Battery life for the beacon. The main characteristics of the beacons that maid them a reliable solution for indoor localization are as follows:

1. **Small Size**, can be placed almost everywhere in complex environments.

2. **Battery Lifetime**, can work over months with single-coin cell batteries or even using USB-powered or solar-powered beacons.

3. **Transmission Power**, can reach up to $60m$ of transmission. In most localization cases, transmission ranges between 2 to 5 meters are good enough for proximity classification. The advertising intervals between the consecutive transmissions are also critical for the lifespan of the beacons. However, for tracking dynamic objects, short advertising intervals are necessary, which makes the signal unstable. Long intervals improve the signals' stability and increase the beacon's battery lifetime but are not appropriate for dynamic use cases.

4. **Measured Power**, is the expected RSS at 1 m distance from the beacon, and its main feature, practically enables the localization. This value will be calibrated and used by the receiver to approximate the distance from the beacon.

3.3 BLE-based Indoor Localization Techniques

BLE-based indoor localization techniques refer to the methods and technologies used to determine the location of a device or a user within an indoor environment using BLE technology. These techniques have been extensively researched due to the increasing demand for indoor location-based services and the widespread use of BLE-enabled devices.

There are several BLE-based indoor localization techniques, including:

- RSSI-Based Localization: This technique uses the strength of the BLE signal received by the device to estimate its proximity to the beacon. The RSSI value is then used to calculate the distance between the device and the beacon using the path-loss model.
- Trilateration: Trilateration is a technique where the distance between the user and each BLE beacon is estimated using the received RSSI values. The estimated distances are then used to determine the location of the user.
- AoA-Based Localization: This technique uses the direction of the incoming BLE signal to determine the location of the device. It requires multiple BLE beacons with directional antennas and a device with an array of antennas to perform the localization.
- ToF-Based Localization: This technique uses the time taken for the BLE signal to travel from the beacon to the device to estimate the distance between the two. This information is then used to determine the location of the device.
- Fingerprinting: This technique involves collecting a database of RSSI values at known locations in the indoor environment. The collected database is then used to compare with the real-time RSSI values to estimate the user's location. The comparison is performed using a statistical method such as k-Nearest Neighbors (k-NN) or Gaussian Mixture Models (GMM).
- Hybrid Localization Techniques: These techniques combine multiple methods to improve the accuracy of the localization. For example, a combination of the RSSI and AoA techniques can be used to provide a more accurate location estimate.

Each of these techniques has its own strengths and weaknesses and the choice of technique to use depends on the requirements and constraints of the particular use case. In the following sections, we will provide a deeper analysis of some of these techniques and challenges associated with their implementation.

3.3.1 Path Loss Model

Due to the inverse-square law, the distance between the receiver and the beacon can be calculated using the RSSI from the beacon if no other errors contribute to the

faulty results. Each beacon transmits its location id along with transmission power (TX) value. In the simplest form, the path loss will be calculated as follows:

$$RSSI = -10n\log_{10}(d) + A \tag{1}$$

where d denotes the distance, A is the RSS at $1m$, and n is a signal propagation constant depending mainly on the environment. In a noisy environment, the formula will be as follows

$$RSSI = RSSI_0 - 10n\log_{10}(\frac{d}{d_0}) + v \tag{2}$$

where the $RSSI_0$ is the RSSI value at a reference distance of d_0 in $1m$, n is the path loss component, and v, as the random effect of shadowing, is a Gaussian random variable with zero mean and standard deviation equal to σ. Euqations 32 and 4 respectivly are the distance with and without considering the noise of the RSSI,

$$d_{noiseless} = d_0 10^{\frac{RSSI_0 - RSSI}{10n}} \tag{3}$$

$$d = d_{noiseless} exp^{-0.5\left(\frac{\sigma_{RSSI}\ln 10}{10n}\right)^2} \tag{4}$$

The n for each environment can be calculated by knowing the $RSSI_0$ and d_0 using the equation 1.

3.3.2 Trilateration

Lateration is a techniques whic is used to approximate the location of the receiver using the distance from a set of points with known location. Trilateration as shown in Fig 3.5, caculates the intersecting point of the three circles with known center points where the beacons are located and their radii. This intersection is the location of the smartphone. Assuming smartphone at (x, y), one beacon at $(0, 0)$ and others will be at (l, m), and $(k, 0)$. Consequently we can estimate the smartphone location based

on Fig 3.5 and following equations:

$$r_1^2 = x^2 + y^2 \tag{5}$$

$$r_2^2 = (x - l)^2 + (y - m)^2 \tag{6}$$

$$r_3^2 = (x - k)^2 + y^2 \tag{7}$$

By combining the above equations, the location of the smartphone can be calculated as follows:

$$x = \frac{r_1^2 - r_3^2 + k^2}{2k} \qquad y = \frac{r_1^2 - r_3^2 + l^2 + m^2}{2m} - \frac{l}{m}x \tag{8}$$

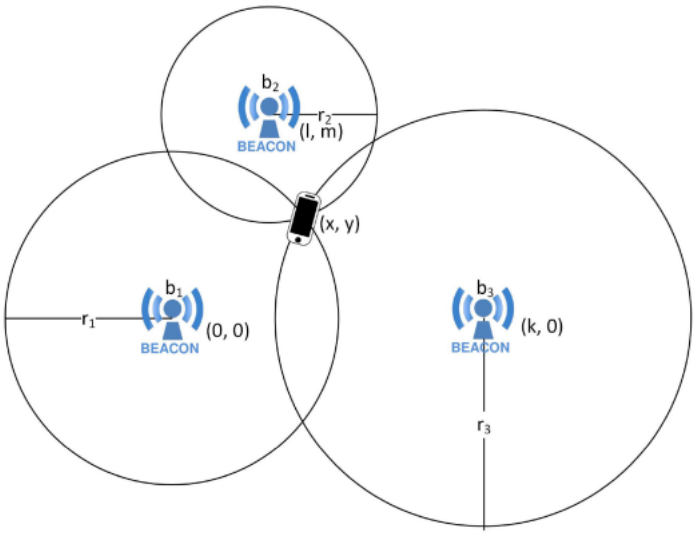


Figure 3.5: Example of trilateration with three beacons, b_1 , b_2 , and b_3 in known locations, $(0, 0)$, (l, m), and $(k, 0)$, respectively, are the transmitters and a smartphone at the intersection, (x, y), as the receiver.

To measure the accuracy of the model and knowing about required calibration, mean square error (MSE) can be used to calculate the error between the estimated and actual location as follows:

$$MSE_{est} = \sqrt{(x_{est} - x_{real})^2 + (y_{est} - y_{real})^2} \tag{9}$$

3.3.3 RSSI-based Coupled Kalman and Particle Filtering

The Kalman Filter is a powerful tool that can improve the accuracy of indoor localization services in noisy and complex environments [12]. By mitigating the noise in an environment, the Kalman Filter can improve the accuracy of RSSI-based indoor localization. In this section, we utilized the Kalman Filter in combination with particle filtering to enhance the accuracy of the results and overcome challenges such as multi-path fading and drastic fluctuations in the indoor environment. We consider tracking a person walking in an indoor venue with N_b number of installed BLE sensors. The following non-linear state-space is considered to model dynamics of the tracking problem

$$\boldsymbol{x}_k = \boldsymbol{f}\big(\boldsymbol{x}_{k-1}\big) + \boldsymbol{w}_k, \tag{10}$$

$$\text{and} \quad \boldsymbol{z}_k = \begin{bmatrix} Z_k^{(1)} \\ \vdots \\ Z_k^{(N_b)} \end{bmatrix} = \underbrace{\begin{bmatrix} h^{(1)}(\boldsymbol{x}_k) + v_k^{(1)} \\ \vdots \\ h^{(N_b)}(\boldsymbol{x}_k) + v_k^{(N_b)} \end{bmatrix}}_{\boldsymbol{h}(\boldsymbol{x}_k)+\boldsymbol{v}_k}, \tag{11}$$

where $\boldsymbol{z}_k \in \mathbb{R}^{N_b}$ denotes the sensor's measurement vector at iteration k; $\boldsymbol{x}_k = [X_k, \Delta X_k, Y_k, \Delta Y_k]^T \in \mathbb{R}^4$ denotes the state vector (2-D location of the target) and their rates of changes $\Delta \boldsymbol{X}_k$ and $\Delta \boldsymbol{Y}_k$; functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$, respectively, are the transition and observation models; terms $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ represent uncertinities in the process and measurement models.

The RSSI value $Z_k^{(j)}$ obtained from the j^{th} active BLE beacon, for $(1 \leq j \leq N_b)$, at time instance k is represneted based on the following observation model

$$Z_k^{(j)} = \underbrace{-10\, N\, \log(\frac{D_k^{(j)}}{D_0}) + C_0}_{h^{(j)}(\boldsymbol{x}_k^{\text{RSSI}})} + v_k^{(j)}, \tag{12}$$

where

$$D_k^{(j)} = \sqrt{(X_k^{\text{RSSI}} - X^{\text{RSSI}}{}_k^{(j)})^2 + (Y_k^{\text{RSSI}} - Y^{\text{RSSI}}{}_k^{(j)})^2} \tag{13}$$

with $\boldsymbol{x}^{\text{RSSI}}{}_k^{(j)} = [X^{\text{RSSI}}{}_k^{(j)}, Y^{\text{RSSI}}{}_k^{(j)}]^T$ denotes 2-D location of j^{th} sensor; D_0 is the reference distance; C_0 is the average RSSI value at reference distance; N is the pathloss

exponent. The RSSI filter consists of the following two main steps:

RSSI - S1. Smoothing: Given the RSSI fluctuations, first, we smooth the RSSI values with a Kalman Filter (KF)-based algorithm. In this regard, we model the smoothed RSSI values with $\boldsymbol{y}_k \in \mathbb{R}^{N_b}$ based on $\boldsymbol{y}_k = \boldsymbol{y}_{k-1} + \boldsymbol{\nu}_k$ as the transition model. The measured RSSI values receiving from all N_b beacons ($\boldsymbol{z}_k = [Z_k^{(1)}, \dots Z_k^{(j)}, \dots Z_k^{(N_b)}]^T$) are used as the input vector to the KF with $\boldsymbol{z}_k = \boldsymbol{y}_k + \boldsymbol{u}_k$ as the observation model. Terms $\boldsymbol{\nu}_k$ and $\boldsymbol{u}_k$ are zero-mean Gaussian additive noises with their second-order statistics ($\boldsymbol{Q}_{\text{RSSI}}$ and $\boldsymbol{R}_{\text{RSSI}}$) being learned through an initial calibration phase. The output of the KF at each iteration is, therefore, denoted by $\hat{\boldsymbol{y}}_{k|k}$, which represents the smoothed RSSI vector [20].

RSSI - S2. Particle Filtering: To estimate the location of a user, denoted by $\hat{X}_k^{\text{RSSI}}, \hat{Y}_k^{\text{RSSI}}$, we apply a Particle Filtering (PF) approach via the following dynamic model

$$\begin{bmatrix} X_k^{\text{RSSI}} \\ \Delta X_k^{\text{RSSI}} \\ Y_k^{\text{RSSI}} \\ \Delta Y_k^{\text{RSSI}} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{F}_k} \begin{bmatrix} X_{k-1}^{\text{RSSI}} \\ \Delta X_{k-1}^{\text{RSSI}} \\ Y_{k-1}^{\text{RSSI}} \\ \Delta Y_{k-1}^{\text{RSSI}} \end{bmatrix} + \boldsymbol{w}_k. \tag{14}$$

The PF approximates the posterior distribution $P(\boldsymbol{x}_k|\boldsymbol{z}_k)$ using a set of N_p particles $\{\mathbb{X}_k^i\}_{i=1}^{N_p}$ and their associated normalized weights W_k^i. More details on PF is available in Reference [180]. Covariance matrix associated with RSSI-based localization data denoted as $\boldsymbol{\Sigma}^{\text{RSSI}}$ can be calculated as follows,

$$\boldsymbol{\Sigma}^{\text{RSSI}} = \frac{1}{N_p - 1} \sum_{i=1}^{N_p} (x_i^{\text{RSSI}} - \overline{x}^{\text{RSSI}})^2 \tag{15}$$

where n is the number of samples, $\overline{x}^{\text{RSSI}}$ is the mean of the x^{RSSI} values.

3.3.4 Angle of Arrival (AoA)

AoA technique is one of the most robust methods to assess users' approximate position with high precision. AoA sensing requires an antenna array to receive the same signal with different phases, which is a combination of N_e number of connected antennas,

called elements. Fig. 3.6 shows a typical Linear Antenna Array (LAA). Among all the AoA methods, the MUltiple SIgnal Classification (MUSIC) algorithm is one of the most promising frameworks to measure radio wave incident direction. The incident signals from N_u users are represented by $s_u(t)$, for $(1 \leq u \leq N_u)$, where $s_u(t)$ is a transmitted narrowband signal, expressed as

$$s_u(t) = s_u^{(b)}(t)e^{j2\pi f_c t}, \tag{16}$$

where $s_u^{(b)}(t)$ is the baseband version of the transmitted signal, and f_c denotes the carrier frequency, which is between 2.4 and 2.48 GHz in the BLE standard. Taking into account that τ_e is the time delay required by the antenna array to receive the incident signal, $s_u(t-\tau_e)$ denotes the received signal by element e in LAA. Based on the narrowband assumption, the frequency response can be considered flat. Therefore, we have $s_u(t-\tau_e) \simeq s_u(t)$. In this regard, $s_u(t-\tau_e)$ can be calculated as follows

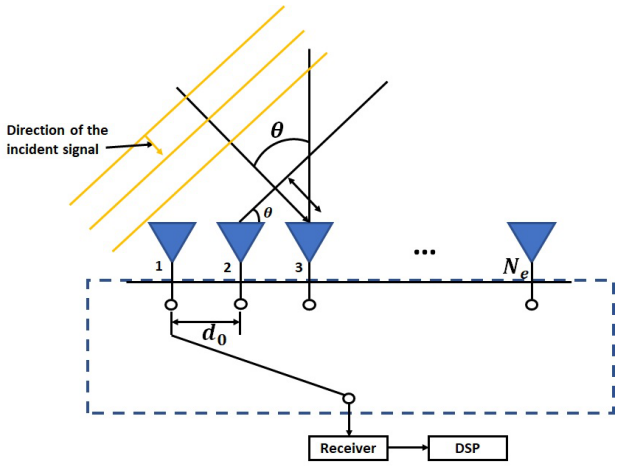


Figure 3.6: A typical structure of uniform Linear Antenna Array.

$$s_u(t-\tau_e) = s_u^{(b)}(t-\tau_e)e^{j2\pi f_c(t-\tau_e)} = s_u^{(b)}(t)e^{j2\pi f_c(t-\tau_e)}. \tag{17}$$

As shown in Fig. 3.6 and by considering the first element in LAA as the reference point, the incident signal received by e^{th} element travels extra distance, denoted by $r = dsin\theta_u$, which leads to receiving same signal with different phase by different elements. Note that d represents the space between the first and e^{th} element of the LAA, where the distance between two consecutive elements denotes by $d_0 = \lambda/2$. Moreover, θ_u is the direction of the incident signal by u^{th} user. Therefore,

the received signal by e^{th} element can be expressed by $s_u^{(b)}(t)e^{-j(e-1)\frac{2\pi dsin\theta_u}{\lambda}}$, where $\lambda = \frac{c}{f_c}$ denotes the wave length of signal. c is the speed of light, about $3 \times 10^8 m/s$. Considering an AWGN channel, and taking into account that LAA receives multiple signals from N_u users at the same time, we have

$$r_e(t) = \sum_{u=1}^{N_u} s_u^{(b)}(t)e^{-j(e-1)\frac{2\pi dsin\theta_u}{\lambda}} + n_e(t), \tag{18}$$

where $n_e(t)$ and $r_e(t)$ denote the noise and the received signal by e^{th} element, respectively. By assuming $a_e(\theta_u) = e^{-j(e-1)\frac{2\pi dsin\theta_u}{\lambda}}$, we have

$$r_e(t) = \sum_{u=1}^{N_u} a_e(\theta_u)s_u^{(b)}(t) + n_e(t). \tag{19}$$

Matrices can define this expression as:

$$\boldsymbol{R} = \boldsymbol{AS} + \boldsymbol{N}, \tag{20}$$

where

$$\boldsymbol{R} = [r_1(t), \ldots, r_{N_e}(t)]^T, \tag{21}$$

$$\boldsymbol{S} = [s_1(t), \ldots, r_{N_u}(t)]^T, \tag{22}$$

$$\boldsymbol{A} = [\boldsymbol{a}(\theta_u), \ldots, \boldsymbol{a}(\theta_{N_u})]^T = \begin{bmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ e^{-j(e-1)\frac{2\pi dsin\theta_1}{\lambda}} & \ldots & e^{-j(e-1)\frac{2\pi dsin\theta_{N_u}}{\lambda}} \end{bmatrix},$$

and

$$\boldsymbol{N} = [n_1(t), \ldots, n_{N_e}(t)]^T. \tag{23}$$

To calculate the angle of incident signals, we need to calculate the correlation matrix as follows

$$\boldsymbol{R}_r = E[\boldsymbol{R}\boldsymbol{R}^H] = E[(\boldsymbol{A}\boldsymbol{S} + \boldsymbol{N})(\boldsymbol{A}\boldsymbol{S} + \boldsymbol{N})^H] = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + \boldsymbol{R}_s, \tag{24}$$

where $\boldsymbol{R}_s = E[\boldsymbol{S}\boldsymbol{S}^H]$ and $\boldsymbol{R}_n = \sigma^2 I$ represent the signal and the noise correlation matrices, respectively. σ^2 is the power of noise and I is a unit matrix of $(N_e \times N_e)$. The array correlation matrix $\boldsymbol{E}$ has N_e eigen values. By sorting the eigen values from the largest to the smallest, the matrix $\boldsymbol{E}$ is divided into two subspaces $[\boldsymbol{E}_N \boldsymbol{E}_S]$. The spatial spectrum function for MUSIC, denoted by $P_{MU}(\theta)$ is defined as

$$\boldsymbol{P}_{MU}(\theta) = \frac{\boldsymbol{a}^H(\theta_u)\boldsymbol{a}(\theta_u)}{\boldsymbol{a}^H(\theta_u)\boldsymbol{E}_N\boldsymbol{E}_N^H\boldsymbol{a}(\theta_u)}. \tag{25}$$

Then, the maximum peak of the spatial spectrum indicates the angle of the incident signal. It is assumed that θ_u^b indicates the angle between x-axis and the line between user u and BLE beacon b. To estimate the location of user u, denoted by $(\boldsymbol{x}_u^{\text{AoA}}, \boldsymbol{y}_u^{\text{AoA}})$, we have

$$\boldsymbol{x}_u^{\text{AoA}} = \frac{D_{k,l}\tan\theta_u^l}{\tan\theta_u^l - \tan\theta_u^k}, \tag{26}$$

$$\boldsymbol{y}_u^{\text{AoA}} = \frac{D_{k,l}\tan\theta_u^l\tan\theta_u^k}{\tan\theta_u^l - \tan\theta_u^k}, \tag{27}$$

where $D_{k,l}$ is the distance between l^{th} and k^{th} BLE beacons. Moreover, (x_l, y_l) and (x_k, y_k) represent the location of l^{th} and and k^{th} BLEs, respectively.

3.3.5 Fingerprinting

Fingerprinting is a well-established indoor localization technique that leverages the RSSI of wireless signals, such as BLE, to determine the position of a device within an indoor environment. It works by creating a map of the environment, called a fingerprint, that associates the RSSI values of the wireless signals at different locations with the corresponding positions. The position of a device is then estimated by comparing the RSSI values of the signals received at that device with the fingerprint.

The basic formula for RSSI-based fingerprinting is as follows:

$$P_{i,j} = P_0 - 10n\log_{10}(d_{i,j}) \tag{28}$$

where $P_{i,j}$ is the measured RSSI value at location i for the jth signal, P_0 is the reference RSSI value, $d_{i,j}$ is the distance between the device and the jth signal source, and n is the path loss exponent that characterizes the signal propagation in the environment.

In the offline phase, the fingerprinting algorithm creates a database of the RSSI values at different locations in the environment. To do this, the algorithm measures the RSSI values of the wireless signals at a set of predefined locations, typically using a mobile device. The locations and the corresponding RSSI values are then stored in the database as a map of the environment.

In the online phase, the fingerprinting algorithm uses the database to estimate the position of the device. To do this, the algorithm measures the current RSSI values of the wireless signals at the device and compares them to the values stored in the database. The position of the device is then estimated based on the closest match between the current RSSI values and the values in the database.

Fingerprinting is a powerful indoor localization technique as it can provide high accuracy if the database is correctly created and maintained. However, it also has some limitations. One of the main limitations is that the accuracy of the localization depends on the size and coverage of the database, as well as the variability of the RSSI values over time. To overcome these limitations, some fingerprinting algorithms use dynamic update mechanisms to adapt the database to changes in the environment and improve the accuracy of the localization.

3.4 The PDR Path, IMU based Indoor Localization

As stated previously, PDR is a widely used method to localize a user in an indoor environment. The functionality of this method is based on the data derived from three portable sensors, including accelerometer, gyroscope, and magnetometer, embedded in smartphones. The accelerometer is the most commonly used inertial sensor reporting the translational acceleration data concerning three orthogonal axes. The accelerometer's data can be analyzed to detect the steps of the user walking in a venue. Moreover, to estimate the user's heading at each timestamp, the accelerometer's data should be merged with the magnetometer's data to gain pitch and roll

angles. Gyroscope, another sensor within the IMU, reports the rate of changes in angular velocity. Gyroscope's data, however, is prone to sudden drifts and a high rate of error at specific points. Therefore, a gyroscope data correction algorithm, such as a KF or a complementary filter, is often applied to the smoothed gyroscope data to mitigate such sudden drifts.

Additionally, a magnetometer is regarded as a compass, i.e., inertial navigation sensor, to estimate path trajectory. As mentioned previously, a combination of the accelerometer and magnetometer's processed data can lead to the heading estimation of the user at each data point/step concerning the earth's magnetic field. The magnetometer, however, is highly prone to both soft and hard iron effects. The soft iron effect occurs when the magnetometer is close to a temporary magnetic field, imposing the magnetometer's distortion data. Analyzing such noisy data will consequently result in mistaken heading estimation. The hard iron effect is the result of the earth's magnetic field, which is hard to be compensated. Therefore, to estimate the actual heading in a path trajectory point, it is essential to eliminate soft and hard iron effects on the magnetometer. Hence, the magnetometer requires an additional processing unit to compensate the errors imposed by soft and hard iron effects. The yaw angle represents the users' heading in an indoor environment at each timestamp while the user holds the smartphone in a fixed texting position.

Raw data from three-axis IMU sensors of a smartphone are reported through an iOS (SWIFT language-based) application. Data should be processed before making any analysis of the IMU data. The possessing unit includes two significant components, i.e., the smoothing and the calibration units. The smoothing unit is realized as an essential processing block of any PDR-based localization system since it can significantly decrease the noisy and high-frequency data. Therefore, to eliminate the noise and error (mostly present in high frequencies) representing a sudden noise or outlier motion data, the raw data reported by the accelerometer, magnetometer, and gyroscope are fed to a low pass filter (100 Hz) smoothing filter. Moreover, a calibration unit should be designed to mitigate/compensate for noise effect on the magnetometer's data. Ellipsoid Fit method [181] is recognized as one of the most efficient methods to calibrate the magnetometer's data. This calibration unit can estimate all the error model parameters and compensate for errors caused by both soft and hard iron interference. Fig. 3.7 illustrates a sample of magnetometer's Ellipsoid

Fit calibration result.

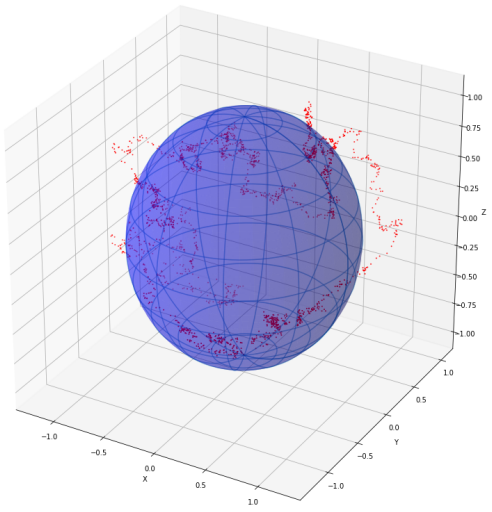


Figure 3.7: Ellipsoid Fit Magnetometer's calibration method.

- ***PDR - S1. Step Detection:*** The smoothed accelerometer data is then used to estimate the number of steps in the path trajectory. The step detection unit consists of two major stages. First, based on the axis that holds the highest value of the accelerometer's variance, the step detection method should count the data values' peaks. The number of peaks can be proportional to the number of steps; however, there are always some peaks in the signal representing a sudden shake or error of accelerometer. Therefore, the step detection algorithm should differentiate the actual steps and the false steps (local peaks) detected by the algorithm. The designed algorithm assigns a threshold value proportional to the signal's variance to mitigate the number of such local peaks. Only two consecutive peaks are allowed to be detected as two steps if the value of those peaks has a reasonable distance (exceeding the threshold) in the signal. After estimating the steps' number, the step-index of the data (the index in which the peak representing a step has occurred) and the number of data sent between every two consecutive peaks are also recorded to be utilized in the step-based heading estimation algorithm. The step indexes are then used to split the whole IMU data into small data subsets, each representing a particular step trajectory.

- ***PDR - S2. Heading Estimation:*** Heading estimation is regarded as the

most challenging and most crucial part of the PDR-based localization technique. In order to predict the heading of the user in a venue, the data reported by IMU sensors are split into small subsets representing the motion data of the user's steps.

Once the pitch and roll angles for each step are calculated, the yaw angle, that represents the 2-D heading of the user in a venue, can be estimated using the magnetometer's calibrated data as follows

$$h_k = \arctan\left(\frac{-\big(M_y \cos(R) + M_z \sin(R)\big)}{\big(M_x \cos(P) + M_y \sin(P) \sin(R)\big)}\right), \tag{29}$$

where pitch and roll angles, denoted by P and R, respectively, are estimated based on the accelerometer's data as follows

$$P = \arctan \frac{A_y}{\sqrt{{A_x}^2 + {A_z}^2}}, \tag{30}$$

$$R = \arctan\left(\frac{-A_x}{A_z}\right), \tag{31}$$

with A_x, A_y, and A_z denoting the accelerometer values, calculated at each time stamp for x, y, and z axes, respectively. Finally, terms M_x, M_y, and M_z denote the magnetometer values.

3.5 Experiments and Results

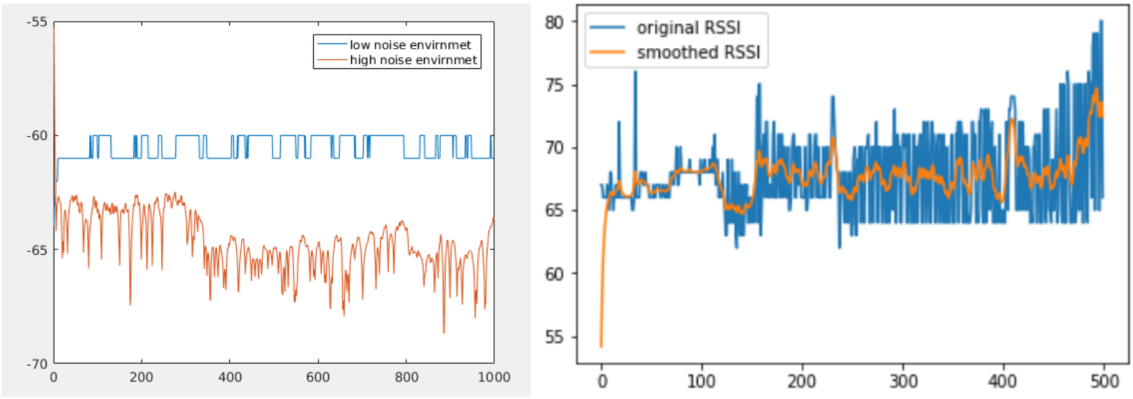
In the current state of IoT dynamic tracking applications, it is commonly assumed that a user's true location within an indoor environment follows a specific pattern and can be estimated without a proper comparison to the ground truth. However, this assumption inherently introduces errors as the actual location of the user is assumed. In this section, we utilize the IoT-TD dataset to analyze the various factors affecting the RSSI values and describe the results of an indoor LBS application named indoor tracking software development kit (SDK) that uses ML-based proximity classification for indoor positioning services. Furthermore, we present the performance of the RSSI-based location estimation, AoA-based localization, and PDR-based tracking estimates, which are evaluated based on the accuracy of the ground truth dataset.

3.5.1 Factors affecting RSSI Values

In this section, we analyze the impact of various factors such as Noisy Environment, Orientation, Obstacle, and Body Shadowing on the RSSI values. To do so, we conduct various test scenarios and examine the effect of these factors on the RSSI readings. This helps us to understand how these factors can affect the accuracy of the RSSI-based location estimation and highlight the need for alternative solutions to overcome these limitations. By exploring these factors, we aim to provide a comprehensive understanding of the limitations of the RSSI-based location estimation and the need for alternative solutions to improve the accuracy of indoor LBS applications.

Effect of Noisy Environment on BLE

In this section, we examine the impact of noisy environments on RSSI values by conducting experiments in two different environments. The RSSI signals collected at a distance of 3 meters between the phone and BLE modules are shown in Fig.3.8(a). To mitigate the effect of noise, we apply Kalman filtering to the RSSI values and compare the smoothed values to the RSSI values measured in a low-noise environment (Fig.3.8(b)). The results show that the RSSI values in high-noise environments are lower than those in low-noise environments, leading to an overestimation of the distance between the phone and BLE module based on the path-loss model. This highlights the importance of considering the effect of noisy environments when estimating the location of a device using RSSI values.



(a) (b)

Figure 3.8: (a) Changes in RSSI values in Noisy and less noisy environments. (b) Smoothed Gathered RSSI Values.

Effect of Orientation on BLE

In this experiment, the impact of orientation on the received RSSI values was studied. It was found that the orientation of a hand-held device relative to the beacon significantly affects the RSSI values. To address this issue and obtain orientation-free RSSI values, a large amount of RSSI data was collected at different distances and in different directions. This experiment was performed at distances of 1 meter and 3 meters from a BLE sensor and involved roughly 12 million RSSI readings collected in 8 different orientations (0, 45, 90, 135, 180, 225, 270, and 315 degrees). The smartphone was fixed in each orientation and 4 sensors simultaneously gathered the RSSI values. Fig. 3.9 shows the RSSI values received from 4 different sensors at the same distance and orientation. It is evident that the RSSI values are different for each orientation.

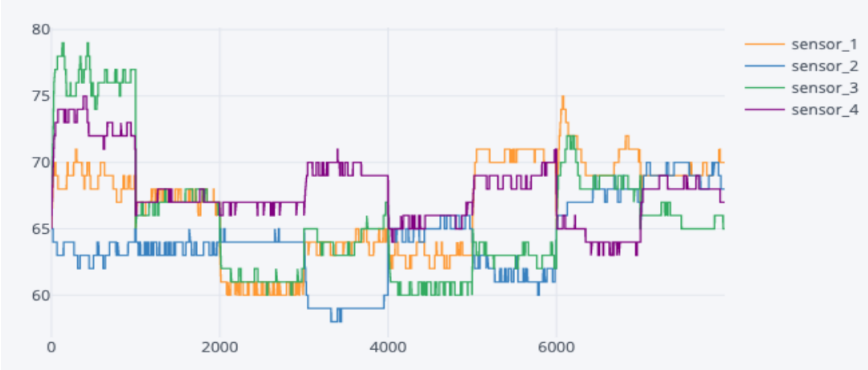


Figure 3.9: RSSI values in different orientation for four different sensors. In this plot, 1,000 RSSI samples are collected from 4 BLE sensors in each orientation $(0, 45, 90, 135, 180, 225, 270, 315$ degrees).

Effect of Obstacle and Body Shadowing on BLE

The presence of obstacles and the body shadowing effect can significantly impact the measured RSSI values. To quantify this effect, experiments were conducted in a controlled environment where the distance between the phone and the BLE module was set to 2 meters. Four different scenarios were considered, including the presence of a wooden, metal, glass obstacle, or no obstacle at all. The results, shown in Fig.3.10, indicate that the presence of obstacles can lead to a decrease in the RSSI values. In addition, the negative impact of another BLE device on the RSSI values is also demonstrated in Fig.3.10(d). The results are summarized in Table 5.1, which highlights the adverse effects of different obstacles and body shadowing on the RSSI

values. These findings highlight the importance of considering the impact of obstacles and body shadowing in the design and deployment of indoor localization systems.

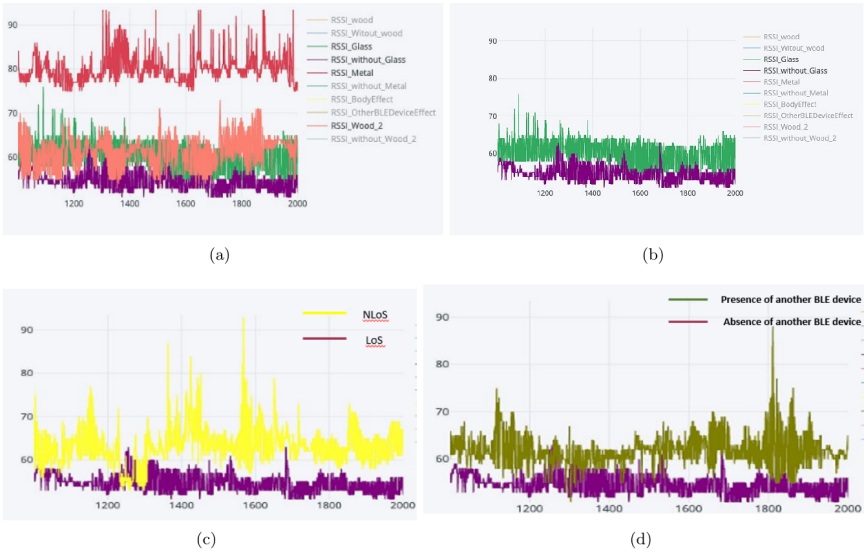


Figure 3.10: Effects of different obstacles on RSSI Values: (a) Effects of 3 different obstacles on RSSI values. (b) Effects of RSSI due to presence or absence of a glass obstacle. (c) The body shadowing effect, LoS vs NLoS. (d) Effect on RSSI values due to presence of another BLE.

Table 3.1: The Propagation Model Parameters for each Situation.

Parameter	**LoS**	**NLoS**	**Metal**	**Wooden**	**Glass**
$RSSI_0$	-52db	-60db	-72db	-57.6db	-57db
n	1.07	1.42	2.36	1.48	1.26

3.5.2 Applied Indoor Localization Algorithms

In this section, the effectiveness of various indoor localization solutions is evaluated using the ground truth IoT-TD data described in Section 3.1.2. The localization techniques used in this study include Angle of Arrival (AoA), Received Signal Strength Indicator (RSSI), and Pedestrian Dead Reckoning (PDR). Additionally, a Reinforcement Learning (RL)-based information fusion approach is proposed and evaluated in the next Chapter 4.

The current indoor tracking applications are limited to predetermined movement scenarios. However, tracking a user's random movements or changes in motion direction or walking patterns is still a major challenge. This is because these predetermined

scenarios are often based on assumptions without a fair comparison to the ground truth data, introducing inherent errors in the estimates of the user's location.

The proposed indoor tracking solutions, including the RSSI path, the PDR path, and the AoA path, are evaluated on a computer with a 3.79 GHz AMD Ryzen 9 processor with 12 cores. The experiments are conducted in all environments and each test is repeated 50 times to ensure accuracy.

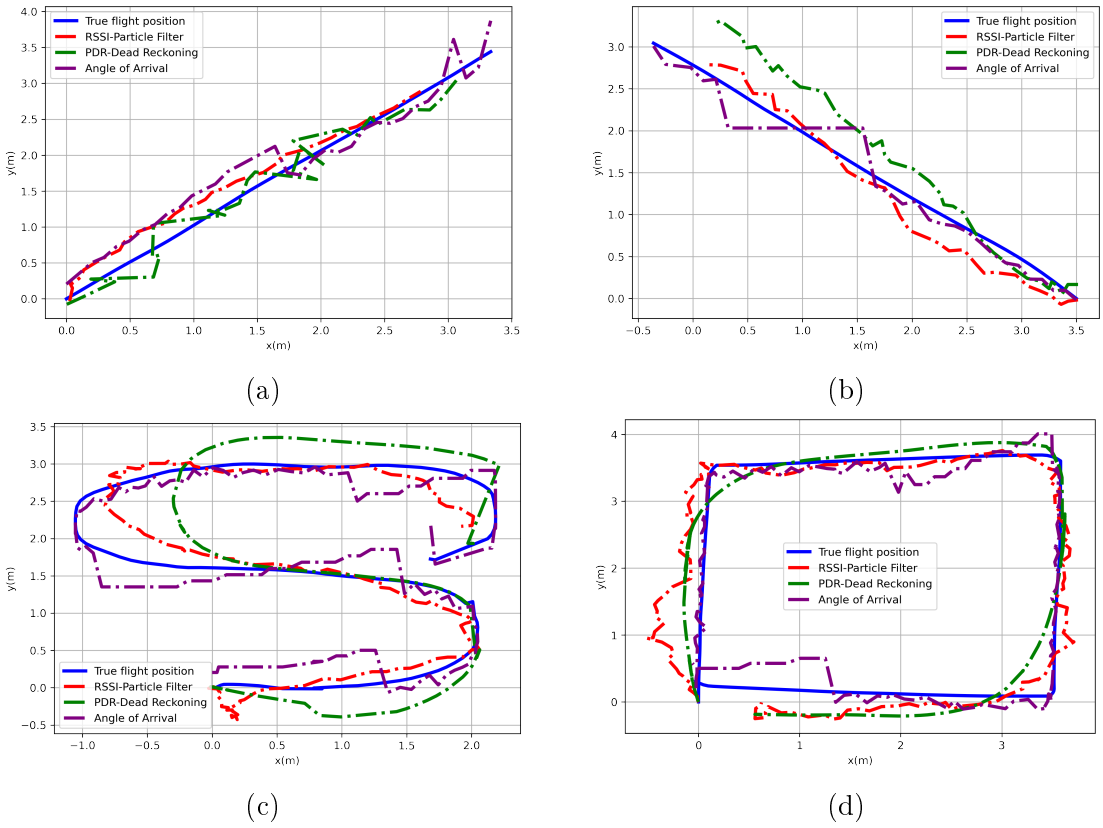


Figure 3.11: Four different movement scenarios in the first environment: (a) Diagonal A; (b) Diagonal B; (c) Random, and (d) Rectangular trajectories.

Figs.4.5,3.12, and 4.6 show the estimated trajectories based on the RSSI-based tracking algorithm coupled with Kalman Filter (KF) and Particle Filter (PF), the PDR-based tracking scheme, the AoA estimation algorithm, and the ground truth data in four different movement scenarios in three environments. The results show that the AoA estimation outperforms the other two scenarios in accurately estimating the target's position, as indicated by the "blue color".

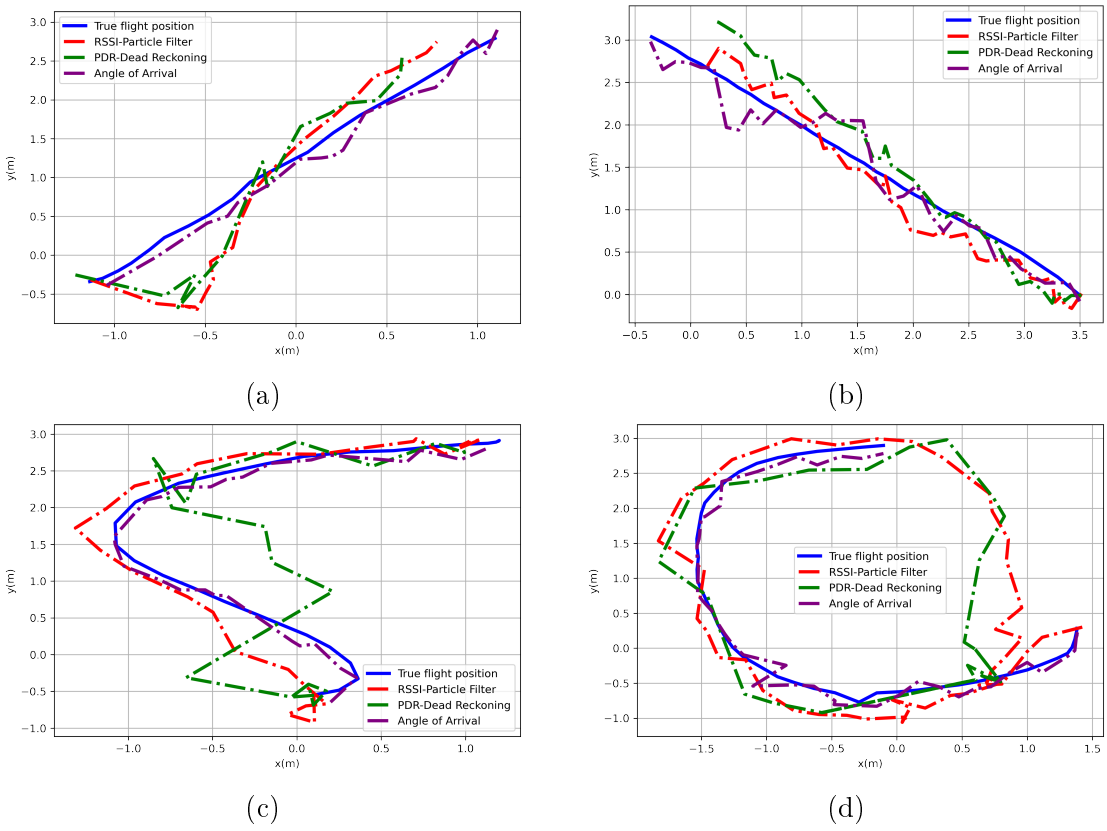


Figure 3.12: Four different movement scenarios in the second environment: (a) Diagonal A; (b) Diagonal B; (c) Random, and (d) Rectangular trajectories.

Table 3.2: The MSE of the location estimation based on the dataset gathered in the FIRST environment by comparing the ground truth (collected with the Vicon system) and the proposed RL-IFF.

Movement Scenario	**AoA**	**RSSI**	**PDR**
Rectangular	0.01979	0.12994	0.21997
Random	0.02508	0.14502	0.17303
Diagonal-A	0.0439	0.1961	0.2925
Diagonal-B	0.01992	0.16002	0.10029

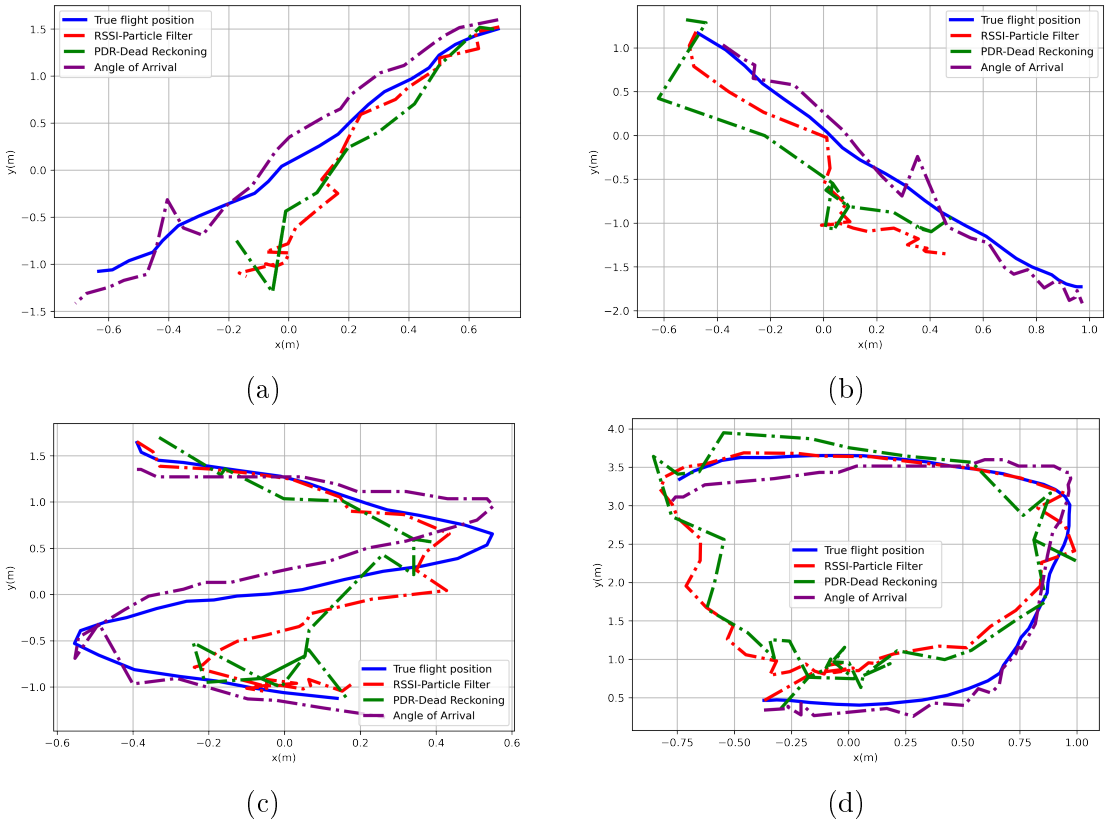


Figure 3.13: Four different movement scenarios in the third environment: (a) Diagonal A; (b) Diagonal B; (c) Random, and (d) Rectangular trajectories.

Table 3.3: The MSE of the location estimation based on the dataset gathered in the SECOND environment by comparing the ground truth (collected with the Vicon system) and the proposed RL-IFF.

Movement Scenario	**AoA**	**RSSI**	**PDR**
Rectangular	0.02193	0.05307	0.09464
Random	0.10133	0.09432	0.16103
Diagonal-A	0.11368	0.12331	0.19331
Diagonal-B	0.06854	0.11268	0.1105

Table 4.5 shows the Mean Squared Error (MSE) (in meter) of the localization estimation in the first environment driven from the primary localization approaches including AoA, RSSI coupled with KF-PF and PDR. According to the results, the AoA path outperforms other tracking paths representing the high potentials of the proposed RL-based solution. The same results can be seen in Table 4.6, which compares the MSE of basic localization approaches in the second environment. As expected, the AoA shows the minimum MSE in location estimation compared to the

Table 3.4: The MSE of the location estimation based on the dataset gathered in the THIRD environment by comparing the ground truth (collected with the Vicon system) and the proposed RL-IFF.

Movement Scenario	AoA	RSSI	PDR
Rectangular	0.12379	1.63892	1.5869
Random	0.02923	0.1123	0.14699
Diagonal-A	0.01468	0.22397	0.29195
Diagonal-B	0.01698	0.06556	0.0961

other baseline algorithms. Table 4.7, similar to the two previous tables, represents the MSE of the tracking approachesin the third environment. The proposed AoA path outperforms other algorithms in this environment as well. As expected, the AoA, outperforms the other underlying models and represents the high potentials of the proposed fusion strategy.

3.5.3 Indoor Localization SDK

The design of the application for indoor localization is based on the RSSI-based solutions. The aim of the application is to reduce the negative impact of the parameters discussed in Section 3.5.1 on the RSSI values. To achieve this, the application integrates an ML-based model to enhance the accuracy of the RSSI-based localization.

The application is designed as a REST API, allowing different sensors in an indoor environment to connect and send real-time data to it. The underlying ML engine uses the position of the sensors and the received RSSI values to estimate the proximity of a user in the indoor environment.

To train the ML model, a large number of RSSI values were measured at different locations and distances in the offline phase. The reference RSSI values were obtained by fixing BLE beacons at a distance of 1 meter from a smartphone. Fig. 3.14 illustrates the use of reference RSSI values in the path-loss model formula and the fingerprinting method for different zones. The model was trained based on the collected RSSI values.

In the online phase, the trained model is embedded in a real-time indoor localization application, as shown in Figs.3.15 and3.16. The real-time stream of the received RSSI values is used to find the proximity of the user. The zone of the user is recognized based on the proximity to the BLE beacons. Proximity is classified into three categories: immediate for distances less than 1 meter, near for distances between 1 to 2 meters, and far for distances greater than 2 meters.

By using this application, it is possible to provide accurate proximity information in an indoor environment, enhancing the overall performance of indoor localization.

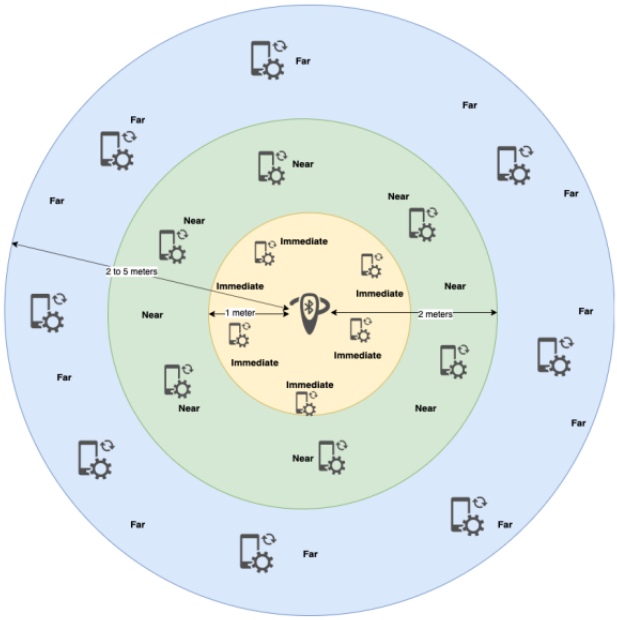


Figure 3.14: Proximity to the BLE beacons, Immediate, Near and Far.

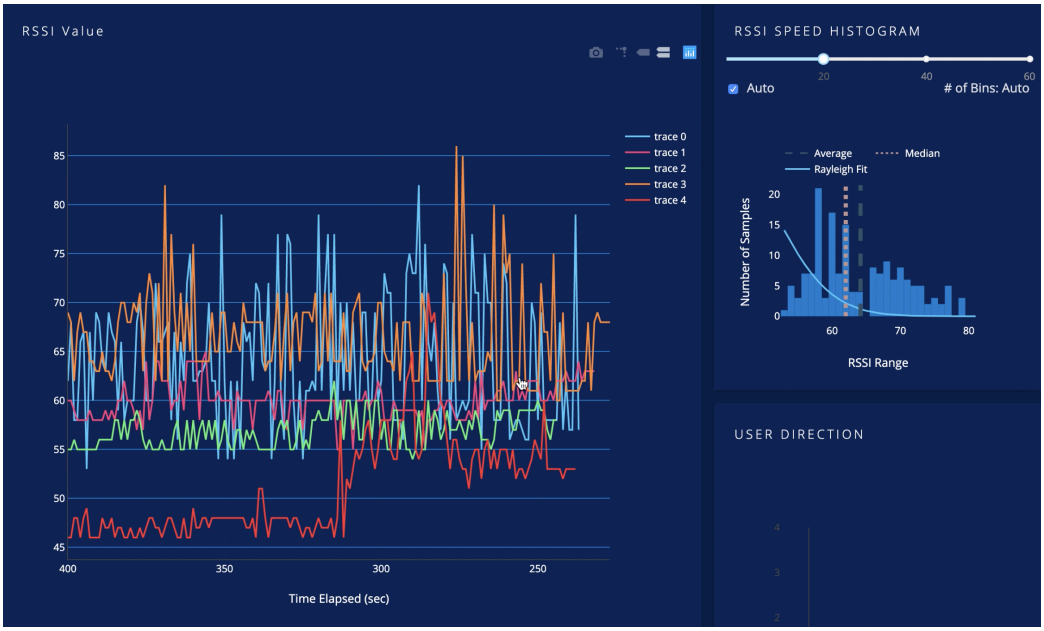


Figure 3.15: Designed SDK for indoor localization services, Received signals from 5 different sensors.

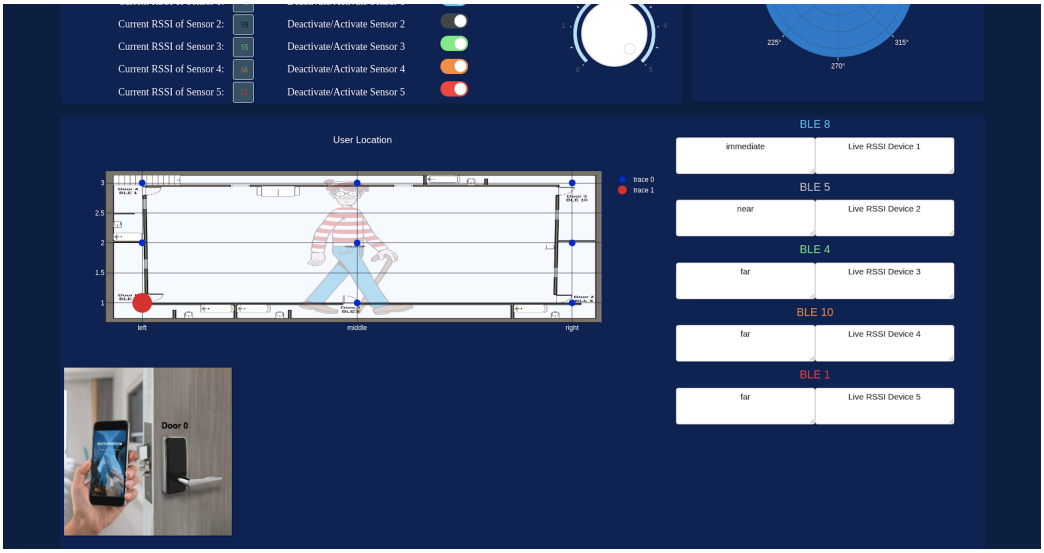


Figure 3.16: Designed SDK for indoor localization services, Proximity results of the agent in the environment.

3.6 Conclusion

In this chapter, we aimed to tackle the challenge of evaluating different BLE-based indoor localization services by introducing a reliable and accurate dataset, the IoT-TD dataset in Section 3.1. The current limitations of existing datasets, which lack precise ground truth information, make it difficult to evaluate and compare different algorithms effectively. However, our IoT-TD dataset overcomes these limitations by providing precise ground truth information with millimeter accuracy. This was achieved by using a specific set of four optical cameras during the data collection sessions.

To consider the effects of the environment, the data collection sessions were held in three different environments. The BLE technology, the factors affecting the RSSI values, and different BLE-based indoor localization solutions are discussed in detail in Sections 3.2 and 3.3. Additionally, the problems related to using these algorithms to provide indoor localization services are addressed in 3.3.

In Section 3.4, we also discuss the PDR path as a solution to indoor localization using IMU sensors. This approach is aimed to be used in the information fusion strategy, which will be presented in the next chapter, to improve the overall indoor localization accuracy. To the best of our knowledge, this is the first time a localization database with ground truth and different localization approaches has been represented

and different indoor localization solutions have been tested and fused to provide the best localization accuracy.

Chapter 4

Multi-Agent Reinforcement Learning Successor Representation and Reinforcement Learning-based Information Fusion Indoor Localization

Building on the findings of Chapter 3, this chapter progresses towards achieving the second objective of this thesis: studying the theory of Multi-agent RL and Successor Representation (SR). We explore how these theories can augment our approach to indoor localization. This chapter pivots towards the exploration of RL agents as discussed in the introduction, specifically focusing on addressing the uncertainties in measurements.

We confront the challenges faced by traditional Model-Based (MB) or Model-Free (MF) RL algorithms when addressing Multi-Agent Reinforcement Learning (MARL) problems, which include overfitting, high sensitivity to parameter selection, and sample inefficiency. Our aim is to develop a novel RL-based information fusion strategy to enhance the accuracy of indoor localization services.

The chapter also introduces an innovative RL-based information fusion strategy, the development of which forms our third objective. This strategy aims to improve the precision of indoor localization services, a challenge discussed in Chapter 3. The

strategy fuses localization data from different indoor localization paths such as RSSI, AoA, and PDR.

The problem formulation for RL algorithms including MARL is first discussed in Section 4.1. Then, the MAK-TD approach is proposed in Section 4.2. The SR-based variant of MAK-TD, referred to as the MAK-SR, is introduced in Section 4.3. Finally, the innovative RL-based information fusion strategy is presented in Section 4.4.

4.1 Problem Formulation

To provide the background required for development of the proposed MAK-TD/SR frameworks, in this section, we present an overview of single agent and MARL techniques.

4.1.1 Single-Agent Reinforcement Learning (RL)

In conventional RL scenarios, typically, a single agent is placed in an unknown environment performing autonomous actions with the goal of maximizing its accumulated reward. In such scenarios, the agent starts its interactions with the environment in an initial state denoted by $\boldsymbol{s}_0$ and continues to interact with the environment until reaching a pre-defined terminal state $\boldsymbol{s}_T$. Action set $\mathcal{A}$ is defined from which the agent can select potential actions following a constructed optimal policy. In other words, given its current state $\boldsymbol{s}_k \in \mathcal{S}$, the single agent follows a policy denoted by π_k and performs action $a_k \in \mathcal{A}$ at time k. Following the agent's action, based on transition probability of $P(\boldsymbol{s}_{k+1}|\boldsymbol{s}_k, a_k) \in \mathcal{P}_a$, it moves to a new state $\boldsymbol{s}_{k+1} \in \mathcal{S}$ receiving reward of $r_k \in \mathcal{R}$. A discount factor $\gamma \in (0, 1)$ is utilized to incorporate future rewards as such balancing the immediate rewards and future ones. In summary, a Markov Decision Process (MDP), denoted by 5-tuple $\{\mathcal{S}, \mathcal{A}, \mathcal{P}_a, \mathcal{R}, \gamma\}$, is typically used as the underlying mathematical model that governs the RL process. Therefore, the main objective is learning an optimal policy to map states into actions by maximizing the expected sum of discounted rewards, which is referred to as the optimal policy π^* [134]. The optimal policy π^* is typically obtained based on the following

state-action value function:

$$Q_{\pi}(\boldsymbol{s}, a) = \mathbb{E}\left\{\sum_{k=0}^{T} \gamma^k r_k | \boldsymbol{s}_0 = \boldsymbol{s}, a_0 = a, a_k = \pi(\boldsymbol{s}_k)\right\}. \tag{32}$$

Note that in Equation (32), $\mathbb{E}\{\cdot\}$ denotes the expectation operator. To perform an action at the learning stage, the current policy is utilized. Once convergence is reached, $a_k = \arg\max_{a \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_k, a)$, which is the optimal policy, can be used by the agent to perform the required tasks. This completes a brief introduction to RL, next, the TD learning is reviewed as a building block of the proposed MAK-TD/SR frameworks.

4.1.2 Off-Policy Temporal Difference (TD) Learning

By taking an action and moving from one state to another, based on the Bellman equation and Bellman update scheme [182], the value function is gradually updated using sample transitions. This procedure is referred to as Temporal Difference (TD) update [182]. There are two approaches to update policy: "on-policy learning" or "off-policy learning". The former techniques use the current policy for action selection. For example, SARSA [183, 184] is an on-policy approach that optimizes the network as

$$Q_{\pi}(\boldsymbol{s}_k, a_k) = Q_{\pi}(\boldsymbol{s}_k, a_k) + \alpha\Big(r_k + \gamma\, Q_{\pi}(\boldsymbol{s}_{k+1}, a_{k+1}) - Q_{\pi}(\boldsymbol{s}_k, a_k)\Big), \tag{33}$$

where α denotes the learning rate and $Q_{\pi}(\boldsymbol{s}_k, a_k)$ is the state-action value function. In on-policy methods, by following a defined policy, selecting a new state becomes a non-optimal task. Additionally, this approach seems to be inefficient in sample selection since the value function is updated through the current policy instead of using the optimized one. In "off-policy" solutions, such as Q-learning [149, 184–186], the information received from previous policies is exploited to update the policy and reach a new one (exploitation). On the other hand, to properly explore new states, a stochastic policy is usually chosen as the behavior policy (exploration). In brief, Q-learning is formed based on the Bellman optimal equation as follows:

$$Q_{\pi^*}(\boldsymbol{s}_k, a_k) = Q_{\pi^*}(\boldsymbol{s}_k, a_k) + \alpha\Big(r_k + \gamma \max_{a \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_{k+1}, a) - Q_{\pi^*}(\boldsymbol{s}_k, a_k)\Big), \tag{34}$$

where the optimal policy π^* is used to form the state-action value function $Q_{\pi^*}(\boldsymbol{s}_k, a_k)$. The policy can be obtained via a greedy approach as follows:

$$V_{\pi^*}(\boldsymbol{s}) = \max_{a \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_k, a). \tag{35}$$

Upon convergence, actions can be selected based on the optimal policy and not the behavior policy as follows:

$$a_k = \arg\max_{a \in \mathcal{A}} Q_{\pi^*}(\boldsymbol{s}_k, a). \tag{36}$$

This completes our discussion on TD learning. In what follows, we discuss the MARL approaches as well as value function approximation using the proposed algorithms in the multi-agent environments.

4.1.3 Multi-Agent Setting

Within the context of MARL, we consider a scenario with N agents, each with its localized observations, actions, and states. In other words, Agent i, for $(1 \leq i \leq N)$, utilizes policy $\pi^{(i)}$, which is a function from the Cartesian product of its localized action set $\mathcal{A}^{(i)}$ and its localized observation set $\mathcal{Z}^{(i)}$ to a real number within zero and one. We use superset $\mathbb{S} = \{\mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(N)}\}$ to collectively represent all the localized states, $\mathcal{S}^{(i)}$, for $(1 \leq i \leq N)$. Likewise, supersets $\mathbb{A} = \{\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(N)}\}$ and $\mathbb{Z} = \{\mathcal{Z}^{(1)}, \ldots, \mathcal{Z}^{(N)}\}$ are used to jointly represent all the localized actions and local observations, respectively. Each agent makes localized decisions following the transition function $T : \mathbb{S} \times \mathcal{A}^{(1)} \times, \ldots, \times \mathcal{A}^{(N)} \rightarrow \mathbb{S}^2$. Consequently, an action is performed locally resulting in a new localized measurement and a localized reward $r^{(i)} : \mathbb{S} \times \mathcal{A}^{(i)} \rightarrow \mathbb{R}$. The main objective of each agent is to maximize its localized expected return $R^{(i)} = \sum_{t=0}^{T} \gamma^t (r^{(i)})^t$ over a termination window of T using a predefined discount factor of γ.

Traditional models like policy gradient or Q-Learning are not suitable for MARL scenarios [187], since the policy of an agent changes during the progress of the training, and the environment becomes non-stationary towards that specific agent's points of view. Consequently, most recently proposed platforms for multi-agent scenarios employ other strategies, where the agents' own observation (known as local information at the execution time) are exploited to learn optimal localized policies. Typically,

such methods do not consider specific communication patterns between agents or any differentiable model of the environment's dynamics [187]. Moreover, these models support different interactions between agents from cooperation to competition or their combination [187, 188]. In this context, an adaptation is made between the decentralized execution and centralized training to be able to feed the policy training steps with more available data to speed up the process of finding the optimal policy.

4.1.4 Multi-Agent Successor Representation (SR)

Within the context of SR, given an initial action $a^{(i)}$ and an initial state $\boldsymbol{s}^{(i)}$, the expected discounted future state occupancy of state $\boldsymbol{s}'^{(i)}$ is estimated based on the current policy $\pi^{(i)}$ as follows:

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, \boldsymbol{s}'^{(i)}, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right], \tag{37}$$

where $\mathbb{1}\{\cdot\} = 1$ if $\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}$; otherwise, it is zero. The SR can be represented with a $N_{\boldsymbol{s}^{(i)}} \times N_{\boldsymbol{s}^{(i)}}$ matrix when the state-space is discrete. The recursive approach used in Equation (33), can be leveraged to update SR as follows:

$$\begin{aligned}\boldsymbol{M}^{\text{new}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) = \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) + \\ \alpha\Big(\mathbb{1}[\boldsymbol{s}_k^{(i)} = \boldsymbol{s}'^{(i)}] + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, \boldsymbol{s}'^{(i)}, a_{k+1}^{(i)}) - \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)})\Big).\end{aligned} \tag{38}$$

After computation (approximation) of the SR, its inner product with the estimated value of the immediate reward can be used to form the state-action value function based on Equation (32), i.e.,

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \sum_{\boldsymbol{s}'^{(i)} \in \mathcal{S}^{(i)}} \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, \boldsymbol{s}'^{(i)}, a_k^{(i)}) R^{(i)}(\boldsymbol{s}'^{(i)}, a_k^{(i)}). \tag{39}$$

As a final note, it is worth mentioning an important characteristic of the SR-based approach, i.e., the state-action value function can be reconstructed based on the reward function. The developed MARL/MASR formulation presented here is used to develop the proposed MAK-TD/SR frameworks in the following sections.

4.2 The MAK-TD Framework

As stated previously, the MAK-TD framework, is a Kalman-based off-policy learning solution for multi-agent networks. More specifically, by exploiting the TD approach represented in Equation (34), the optimal value function associated with the i^{th} agent, for $(1 \leq i \leq N)$, can be approximated from its one-step estimation as follows:

$$Q_{\pi^{(i)*}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx r_k^{(i)} + \gamma \max_{a^{(i)} \in \mathcal{A}} Q_{\pi^{(i)*}}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)}). \tag{40}$$

By changing the variables' order, the reward at each time can be represented (modeled) as a noisy observation, i.e.,

$$r_k^{(i)} = Q_{\pi^{(i)*}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} Q_{\pi^{*}}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)}) + v_k^{(i)}, \tag{41}$$

where v_k is modeled as a zero-mean normal distribution with variance of $R^{(i)}$. By considering the local state-space of each agent, we use localized basis functions to approximate each agent's value function. Therefore, the following value function can be formed for Agent i, for $(1 \leq i \leq N)$,

$$Q_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}, \tag{42}$$

where term $\boldsymbol{\phi^{(i)}}(\boldsymbol{s}^{(i)}, a^{(i)})$ represents a vector of basis functions, $\pi^{(i)}$ is the policy associated with Agent i, and, finally, $\boldsymbol{\theta}_k^{(i)}$ denotes the vector of the weights. Substituting Equation (42) in Equation (41) results in

$$r_k^{(i)} = \Big[\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})^T\Big] \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}, \tag{43}$$

which can be simplified into the following linear observation model:

$$r_k^{(i)} = [\boldsymbol{h}_k^{(i)}]^T \boldsymbol{\theta}_k^{(i)} + v_k^{(i)}, \tag{44}$$

with

$$\boldsymbol{h}_k^{(i)} = \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \max_{a^{(i)} \in \mathcal{A}} \boldsymbol{\phi}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)}). \tag{45}$$

In other words, Equation (44) is the localized measurement (reward) of the i^{th}

agent, which is a linear model of the weight vector $\boldsymbol{\theta}_k^{(i)}$. For approximating localized weight $\boldsymbol{\theta}_k^{(i)}$, first we leverage the observed reward, which is obtained by transferring from state $s_k^{(i)}$ to $s_{k+1}^{(i)}$. Second, given that the noise variance of the measurement is not known a priori, we exploit MMAE adaptation by representing it with M different values ($R^{(j)^{(i)}}$, for $(1 \leq j \leq M)$. Consequently, a combination of M KFs is used to estimate $\hat{\boldsymbol{\theta}}_k^{(i)}$ based on each of its candidate values, i.e.,

$$\boldsymbol{K}_k^{(j)^{(i)}} = \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} \big(\boldsymbol{h}_k^{T^{(i)}} \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} + R^{(j)^{(i)}}\big)^{-1} \tag{46}$$

$$\hat{\boldsymbol{\theta}}_k^{(j)^{(i)}} = \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)} + \boldsymbol{K}_k^{(j)^{(i)}} \big(r_k^{(i)} - \boldsymbol{h}_k^{T^{(i)}} \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)}\big) \tag{47}$$

$$\boldsymbol{P}_{\boldsymbol{\theta},k}^{(j)^{(i)}} = \big(\boldsymbol{I} - \boldsymbol{K}_k^{(j)^{(i)}} \boldsymbol{h}_k^{T^{(i)}}\big) \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{T^{(i)}} \big(\boldsymbol{I} - \boldsymbol{K}_k^{(j)^{(i)}} \boldsymbol{h}_k^{T^{(i)}}\big) + \boldsymbol{K}_k^{(j)^{(i)}} R^{(j)^{(i)}} \boldsymbol{K}_k^{(j)^{T^{(i)}}} \tag{48}$$

where superscript j is used to refer to the j^{th} matched KF, for which a specific value ($R^{(j)^{(i)}}$) is assigned to model covariance of the observation model's noise process. The posterior distribution associated with each of the M matched KFs is calculated based on its likelihood function. All the matched a posteriori distributions are then added together based on their corresponding weights to form the overall posterior distribution given by

$$P^{(i)}(\boldsymbol{\theta}_k | \boldsymbol{Y}_k) = \sum_{j=1}^{M} \omega^{(j)^{(i)}} P^{(i)}(\boldsymbol{\theta}_k^{(i)} | \boldsymbol{Y}_k^{(i)}, R^{(j)^{(i)}}), \tag{49}$$

where $\omega^{(j)^{(i)}}$ is the j^{th} KF's normalized observation likelihood associated with the i^{th} agent and is given by

$$\omega^{(j)^{(i)}} = P^{(i)}(r_k^{(i)} | \boldsymbol{\theta}_{(k|k-1)}^{(i)}, R^{(j)}) = c^{(i)} . e^{\left[\frac{-1}{2}\left(r_k^{(i)} - \boldsymbol{h}_k^{T^{(i)}} \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)}\right)^T \left(\boldsymbol{h}_k^{T^{(i)}} \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} + R^{(j)^{(i)}}\right)^{-1} \big(r_k^{(i)} - \boldsymbol{h}_k^{T^{(i)}} \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)}\big)\right]}, \tag{50}$$

where $c^{(i)} = 1/(\sum_{j=1}^{M} w^{(j)^{(i)}})$. Exploiting Equation (49), the weight and its error

covariance are then updated as follows:

$$\hat{\boldsymbol{\theta}}_k^{(i)} = \sum_{j=1}^{M} \omega^{(j)(i)} \hat{\boldsymbol{\theta}}_k^{(j)(i)} \tag{51}$$

$$\boldsymbol{P}_{\boldsymbol{\theta},k}^{(i)} = \sum_{j=1}^{M} \omega^{(j)(i)} \left(\boldsymbol{P}_{\boldsymbol{\theta},k}^{(j)(i)} + (\hat{\boldsymbol{\theta}}^{(j)(i)} - \hat{\boldsymbol{\theta}}^{(i)})(\hat{\boldsymbol{\theta}}^{(j)(i)} - \hat{\boldsymbol{\theta}}^{(i)})^T \right). \tag{52}$$

To finalize computation of $\hat{\boldsymbol{\theta}}_k^{(i)}$ based on Equations (44)–(52), localized measurement mapping function $\boldsymbol{h}_k^{(i)}$ is required. As $\boldsymbol{h}_k^{(i)}$ is formed by the basis functions, its adaptation necessitates the adaptation of the basis functions. The vector of basis functions shown in Equation (42) is formed as follows:

$$\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}) = \left[\phi_1(\boldsymbol{s}_k^{(i)}), \phi_2(\boldsymbol{s}_k^{(i)}), \ldots, \phi_{N_b-1}(\boldsymbol{s}_k^{(i)}), \phi_{N_b}(\boldsymbol{s}_k^{(i)})\right]^T, \tag{53}$$

where N_b is the number of basis functions. Each basis function is represented by a RBF, which is defined by its mean and covariance parameters as follows:

$$\phi_n(\boldsymbol{s}_k^{(i)}) = \exp\{\frac{-1}{2}(\boldsymbol{s}_k^{(i)} - \boldsymbol{u}_n^{(i)})^T {\boldsymbol{\Sigma}_n^{(i)}}^{-1} (\boldsymbol{s}_k^{(i)} - \boldsymbol{u}_n^{(i)})\}, \tag{54}$$

where $\boldsymbol{u}_n^{(i)}$ and $\boldsymbol{\Sigma}_n^{(i)}$ are the mean and covariance of $\phi_n(\boldsymbol{s}_k^{(i)})$, for $(1 \leq n \leq N_b)$. Generally speaking, the state-action feature vector can be represented as follows:

$$\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = [\phi_{1,a_1}(\boldsymbol{s}_k^{(i)}), \ldots \phi_{N_b,a_1}(\boldsymbol{s}_k^{(i)}), \phi_{1,a_2}(\boldsymbol{s}_k^{(i)}), \ldots \phi_{N_b,a_{D^{(i)}}}(\boldsymbol{s}_k^{(i)})]^T, \tag{55}$$

where $\boldsymbol{\phi}(\cdot) : \mathcal{A}^{(i)} \times \mathcal{S} \rightarrow \mathbb{R}^{N_b \times D^{(i)}}$, and $D^{(i)}$ denotes the number of actions associated with the i^{th} agent. The state-action feature vector $\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)} = a_d^{(i)})$, for $(1 \leq d \leq D^{(i)})$ in Equation (55) is considered to be generated from $\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)})$ by placing this state feature vector in the corresponding spot for action $a_k^{(i)}$ while the feature values for the rest of the actions are set to zero, i.e.,

$$\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = [0, \ldots 0, \phi_1(\boldsymbol{s}_k^{(i)}), \ldots, \phi_N(\boldsymbol{s}_k^{(i)}), 0, \ldots 0,]^T. \tag{56}$$

Due to the large number of parameters associated with the measurement mapping function, the multiple model approach seems to be inapplicable. Alternatively, Restricted Gradient Descent (RGD) [118] is employed, where the goal is to minimize

the following loss function:

$$L_k^{(i)} = (\boldsymbol{\phi}^T(\boldsymbol{s}_k^{(i)}, a_k)\,\boldsymbol{\theta}_k^{(i)} - r_k^{(i)})^2. \tag{57}$$

The gradient of the objective function with respect to the parameters of each basis function is then calculated using the chain rule as follows:

$$\Delta \boldsymbol{u}^{(i)} = -\frac{\partial L_k^{(i)}}{\partial \boldsymbol{u}^{(i)}} = -\frac{\partial L_k^{(i)}}{\partial Q_{\pi^{*(i)}}} \frac{\partial Q_{\pi^{*(i)}}}{\partial \boldsymbol{\phi}^{(\boldsymbol{i})}} \frac{\partial \boldsymbol{\phi}^{(\boldsymbol{i})}}{\partial \boldsymbol{u}^{(i)}} \tag{58}$$

$$\text{and } \Delta \boldsymbol{\Sigma}^{(i)} = -\frac{\partial \boldsymbol{\Sigma}_k^{(i)}}{\partial \boldsymbol{u}^{(i)}} = -\frac{\partial L_k^{(i)}}{\partial Q_{\pi^{*(i)}}} \frac{\partial Q_{\pi^{*(i)}}}{\partial \boldsymbol{\phi}^{(\boldsymbol{i})}} \frac{\partial \boldsymbol{\phi}^{(\boldsymbol{i})}}{\partial \boldsymbol{\Sigma}^{(i)}}, \tag{59}$$

where calculation of the partial derivations is done leveraging Equations (42), (54), and (57). Therefore, the mean and covariance of the RBFs can be adapted using the calculated partial derivative as follows:

$$\boldsymbol{u}_n^{(i)} = \boldsymbol{u}_n^{(i)} - 2\lambda_{\boldsymbol{u}^{(i)}} \left(L_k^{(i)}\right)^{\frac{1}{2}} \boldsymbol{\theta}_k^{(i)^T} (\boldsymbol{\Sigma}_n^{(i)})^{-1} (\boldsymbol{s}_k^{(i)} - \boldsymbol{u}_n^{(i)}) \tag{60}$$

$$\boldsymbol{\Sigma}_n^{(i)} = \boldsymbol{\Sigma}_n^{(i)} - 2\lambda_{\boldsymbol{\Sigma}^{(i)}} \left(L_k^{(i)}\right)^{\frac{1}{2}} \boldsymbol{\theta}_k^{(\boldsymbol{i})^T} (\boldsymbol{\Sigma}_n^{(i)})^{-1} \times (\boldsymbol{s}_k^{(i)} - \boldsymbol{u}_n^{(i)})(\boldsymbol{s}_k^{(i)} - \boldsymbol{u}_n^{(i)})^T \boldsymbol{\Sigma}_n^{(i)^{-1}}, \tag{61}$$

where both $\lambda_{\boldsymbol{u}^{(i)}}$ and $\lambda_{\boldsymbol{\Sigma}^{(i)}}$ denote the adaptation rates. Based on [118], for the sake of stability, only one of the updates shown in Equations (60) and (61), will be applied. To be more precise, when the size of the covariance is decreasing (i.e., $L_k^{(i)^{\frac{1}{2}}} (\boldsymbol{\theta}_k^{(i)^T} \boldsymbol{\phi}(\cdot)) > 0$), the covariances of the RBFs are updated using Equation (61); otherwise, their means are updated using Equations (60). Using this approach, unlimited expansion of the RBF covariances is avoided.

Algorithm 1 The Proposed MAK-TD Framework

Learning Phase:
Set $\boldsymbol{\theta}_0, \boldsymbol{P}_{\boldsymbol{\theta},0}, \boldsymbol{F}, \boldsymbol{u}_{n,i_d}, \boldsymbol{\Sigma}_{n,i_d}$ for $n = 1, 2, \ldots, N$ and $i_d = 1, 2, \ldots, D$
Repeat (for each episode):
    Initialize $\boldsymbol{s}_k$
    Repeat (for each agent i):
      While $\boldsymbol{s}_k^{(i)} \neq \boldsymbol{s}_T$ **do**:
        $a_k^{(i)} = \arg\max_a \left(\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \boldsymbol{h}_k^{T(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \right)$
        Take action $a_k^{(i)}$, observe $\boldsymbol{s}_{k+1}^{(i)}, r_k^{(i)}$
        Calculate $\boldsymbol{\phi^{(i)}}(\boldsymbol{s}^{(i)}, a^{(i)})$ via Equations (53) and (54)
        $\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = \boldsymbol{\phi^{(i)}}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) - \gamma \arg\max_a \boldsymbol{\phi^{(i)}}(\boldsymbol{s}_{k+1}^{(i)}, a^{(i)})$
        $\hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)} = \boldsymbol{F}^{(i)} \hat{\boldsymbol{\theta}}_k^{(i)}$
        $\boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} = \boldsymbol{F}^{(i)} \boldsymbol{P}_{\boldsymbol{\theta},k-1}^{(i)} \boldsymbol{F}^{T(i)} + \boldsymbol{Q}^{(i)}$
        for $j = 1 : M$ **do**:
            $\boldsymbol{k}_k^{(j)(i)} = \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} (\boldsymbol{h}_k^{T(i)} \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} \boldsymbol{h}_k^{(i)} + R^{(j)(i)})^{-1}$
            $\hat{\boldsymbol{\theta}}_k^{(j)(i)} = \hat{\boldsymbol{\theta}}_{(\boldsymbol{\theta},k|k-1)}^{(i)} + \boldsymbol{k}_k^{(j)(i)} (r_k^{(j)} - \boldsymbol{h}_k^{T(i)} \hat{\boldsymbol{\theta}}_{(k|k-1)}^{(i)})$
            $\boldsymbol{P}_{\boldsymbol{\theta},k}^{(i)} = (\boldsymbol{I} - \boldsymbol{K}_k^{(j)(i)} \boldsymbol{h}_k^{T(i)}) \boldsymbol{P}_{(\boldsymbol{\theta},k|k-1)}^{(i)} (\boldsymbol{I} - \boldsymbol{K}_k^{(j)(i)} \boldsymbol{h}_k^{T(i)})^T + \boldsymbol{K}_k^{(j)(i)} R^{(j)} \boldsymbol{K}_k^{(j)^{T(i)}}$
        end for
        Compute the value of c and $w^{(j)(i)}$ by using $\sum_{j=1}^{M} w^{(j)(i)} = 1$ and Equation (50)
        $\hat{\boldsymbol{\theta}}_k^{(i)} = \sum_{j=1}^{M} w^{(j)(i)} \hat{\boldsymbol{\theta}}_k^{(j)(i)}$
        $\boldsymbol{P}_{\boldsymbol{\theta}_k}^{(i)} = \sum_{j=1}^{M} \omega^{(j)(i)} \left(\boldsymbol{P}_{\boldsymbol{\theta},k}^{(j)(i)} + (\hat{\boldsymbol{\theta}}^{(j)(i)} - \hat{\boldsymbol{\theta}}^{(i)})(\hat{\boldsymbol{\theta}}^{(j)(i)} - \hat{\boldsymbol{\theta}}^{(i)})^T \right)$
        RBFs Parameters Update:
        $L_k^{(i)} = (\boldsymbol{\phi}^T(\boldsymbol{s}_k^{(i)}, a_k)\, \boldsymbol{\theta}_k^{(i)} - r_k^{(i)})^2$
        if $L_k^{(i)\frac{1}{2}} (\boldsymbol{\theta}_k^{(i)^T} \boldsymbol{\phi}(\cdot)) > 0$ **then**:
          Update $\boldsymbol{\Sigma}_{n,a_d}$ via Equation (60)
        else:
          Update $\boldsymbol{u}_{n,a_d}$ via Equation (61)
        end if
      end while
Testing Phase:
Repeat (for each trial episode):
    While $\boldsymbol{s}_k \neq \boldsymbol{s}_T$ **do**:
      Repeat (for each agent):
        $a_k = \arg\max_a \boldsymbol{\phi}(\boldsymbol{s}_k, a)^T \boldsymbol{\theta}_k$
        Take action a_k, and observe $\boldsymbol{s}_{k+1}, r_k$
        Calculate Loss S_k for all agents
    End While

One superiority that the proposed learning framework shows over other optimization-based techniques (e.g., gradient descent-based methods) is the calculation of the uncertainty for the weights $\boldsymbol{P}_{\boldsymbol{\theta},k}^{(i)}$, which is directly related to the uncertainty of the value function. This information can then be used at each step to select the actions, leading to the most reduction in the weights' uncertainty. Using the information form of the KF (information filter [189]), the information of the weights denoted by $\boldsymbol{P}_{\boldsymbol{\theta},k}^{(i)}$ is updated as follows:

$$\boldsymbol{P}_{\boldsymbol{\theta},k}^{-1^{(i)}} = \boldsymbol{P}_{(\boldsymbol{\theta}_{,k|k-1})}^{-1}{}^{(i)} + \boldsymbol{h}_k^{(i)} R^{-1^{(i)}} \boldsymbol{h}_k^{T^{(i)}}. \tag{62}$$

In Equation (62), the second element, i.e., $\boldsymbol{h}_k^{(i)} R^{-1^{(i)}} \boldsymbol{h}_k^{T^{(i)}}$, represents the information received from the measurement. The action is obtained by maximizing the information of the weights, i.e.,

$$\begin{aligned} a_k^{(i)} &= \arg\max_a \left(\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) R^{-1^{(i)}} \boldsymbol{h}_k^{T^{(i)}}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \right) \\ &= \arg\max_a \left(\boldsymbol{h}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \boldsymbol{h}_k^{T^{(i)}}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \right). \end{aligned} \tag{63}$$

The second equality in Equation (63) is constructed as $R^{(i)}$ is a scalar. The projected behavior policy in Equation (63) is different from that in [124], where a random policy was proposed, which favored actions with less certainty of the value function. Although reducing the value function's uncertainty through action selection is an intelligent approach, it is less efficient in sample selection due to the random nature of such policies. Algorithm 3 briefly represents the MAK-TD framework proposed in this thesis.

4.3 The MAK-SR Framework

In the previous section, the MAK-TD framework is proposed, which is a MM Kalman-based off-policy learning solution for multi-agent networks. To learn the value function, a fixed model for the reward function is considered, which could restrict its application to more complex MARL problems. SR-based algorithms are appealing solutions to tackle this issue where the focus is instead on learning the immediate reward and the SR, which is the expected discounted future state occupancy. In the existing SR-based approaches that use standard temporal difference methods,

the uncertainty about the approximated SR is not captured. In order to address this issue, we extend the MAK-TD framework and design its SR-based variant in this section. In other words, MAK-TD is extended to MAK-SR by incorporation of the SR learning procedure into the filtering problem using KTD formulation to estimate uncertainty of the learned SR. Moreover, by applying KTD, we benefit from the decrease in memory and time spent for the SR learning and also sensitivity of the framework's performance to its parameters (i.e., more reliable) when compared to DNN-based algorithms.

Exact computation of the SR and the reward function is, typically, not possible within the multi-agent settings as we are dealing with a large number of continuous states. Therefore, we follow the approach developed in Section 4.2 and approximate the SR and the reward function via basis functions. For the state-action feature vector $\boldsymbol{\phi}(\boldsymbol{s}^{(i)}, a^{(i)})$, a feature-based SR, which encodes the expected occupancy of the features, is defined as follows:

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}, :, a^{(i)}) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) | \boldsymbol{s}_0^{(i)} = \boldsymbol{s}^{(i)}, a_0^{(i)} = a^{(i)}\right]. \tag{64}$$

We consider that the immediate reward function for pair $(\boldsymbol{s}^{(i)}, a^{(i)})$ can be linearly factorized as

$$r^{(i)}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) \approx \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})^T \boldsymbol{\theta}_k^{(i)}, \tag{65}$$

where $\boldsymbol{\theta}_k^{(i)}$ is the reward weight vector. The state-action value function (Equation (39)), therefore, can be computed as follows:

$$Q(\boldsymbol{s}_k^{(i)}, a_k^{(i)}) = {\boldsymbol{\theta}_k^{(i)}}^T \boldsymbol{M}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}). \tag{66}$$

The SR matrix $\boldsymbol{M}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)})$ can be approximated as a linear function of the same feature vector as follows:

$$\boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}_k^{(i)}, :, a_k^{(i)}) \approx \boldsymbol{M}_k \, \boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)}). \tag{67}$$

The TD learning of the SR then can be performed as follows:

$$\boldsymbol{M}^{\text{new}}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}_k, :, a^{(i)}_k) = \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}_k, :, a^{(i)}_k) + \alpha\big(\boldsymbol{\phi^{(i)}}(\boldsymbol{s}^{(i)}_k, a^{(i)}_k) + \gamma \boldsymbol{M}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}_{k+1}, :, a^{(i)}_{k+1}) - \boldsymbol{M}^{\text{old}}_{\pi^{(i)}}(\boldsymbol{s}^{(i)}_k, :, a^{(i)}_k)\big). \tag{68}$$

By defining the estimation structure of the SR and reward function, a suitable method must be selected to learn (approximate) the weight vector of the reward $\boldsymbol{\theta}^{(i)}$ and the weight matrix of the SR $\boldsymbol{M}$ for Agent i. The proposed multi-agent MAK-SR algorithm contains two main components: KTD-based weight SR learning and radial basis function update. For the latter, we apply the method developed in Section 4.2 to approximate the vector of basis functions via representing each of them as a RBF. The gradient of the loss function (57), with respect to the parameters of the RBFs, is calculated using the chain rule for the mean and covariance of RBFs using (60) and (61).

For KTD-based weight SR learning, the SR can be obtained from its one-step approximation using the TD method of Equation (68). In this regard, the state-action feature vector at time step k can be considered as a noisy measurement from the system as follows:

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}^{(i)}_k, a^{(i)}_k) = \boldsymbol{M}^{\text{new}}(\boldsymbol{s}^{(i)}_k, :, a^{(i)}_k) - \gamma \boldsymbol{M}(\boldsymbol{s}^{(i)}_{k+1}, :, a^{(i)}_{k+1}) + \boldsymbol{n}^{(i)}_k, \tag{69}$$

where $\boldsymbol{n}^{(i)}_k$ follows a zero-mean normal distribution with covariance of $\boldsymbol{R}^{(i)}_{\boldsymbol{M}}$. Considering Equations (67) and (69) together, the feature vector $\boldsymbol{\phi}(\boldsymbol{s}^{(i)}_k, a^{(i)}_k)$ can be approximated as

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}^{(i)}_k, a^{(i)}_k) = \boldsymbol{M}_k \underbrace{\Big[\boldsymbol{\phi}(\boldsymbol{s}^{(i)}_k, a^{(i)}_k) - \gamma \boldsymbol{\phi}(\boldsymbol{s}^{(i)}_{k+1}, a^{(i)}_{k+1})\Big]}_{\boldsymbol{g}^{(i)}_k} + \boldsymbol{n}^{(i)}_k. \tag{70}$$

s Matrix $\boldsymbol{M}_k$ is then mapped to a column vector $\boldsymbol{m}^{(i)}_k$ by concatenating its columns. Using the vec-trick characteristic of Kronecker product denoted by $\otimes$, then we can rewrite Equation (70) as follows:

$$\hat{\boldsymbol{\phi}}(\boldsymbol{s}^{(i)}_k, a^{(i)}_k) = ({\boldsymbol{g}^{(i)}}^T_k \otimes \boldsymbol{I})\boldsymbol{m}^{(i)}_k + \boldsymbol{n}^{(i)}_k, \tag{71}$$

where $\boldsymbol{I}$ represents an identity matrix of appropriate dimension. More specifically, Equation (71) is used to represent the localized measurements $(\boldsymbol{\phi}(\boldsymbol{s}^{(i)}_k, a^{(i)}_k))$ linearly

based on vector $\boldsymbol{m}_k^{(i)}$, which requires estimation. Therefore, we use the following linear state model:

$$\boldsymbol{m}_{k+1}^{(i)} = \boldsymbol{m}_k^{(i)} + \boldsymbol{u}_k^{(i)}, \tag{72}$$

to complete the required state-space representation for KF-based implementation. The noise associated with the state model (Equation (72)), i.e., $\boldsymbol{u}_k^{(i)}$, follows a zero-mean normal distribution with covariance of $\boldsymbol{Q}_{\boldsymbol{M}}$. Via implementing the KF's recursive equations, we use the new localized observations to estimate $\boldsymbol{m}_k^{(i)}$ and its corresponding covariance matrix $\boldsymbol{P}_{\boldsymbol{m}^{(i)},k}^{(i)}$. After this step, vector $\boldsymbol{m}_k^{(i)}$ is reshaped to form a $(L \times L)$ matrix in order to reconstruct Matrix $\boldsymbol{M}_k$. Equation (66) is finally used to form the state-action value function for associated with $(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$. Algorithm 2 summarizes the proposed MAK-SR framework.

Algorithm 2 The Proposed MAK-SR Framework

Learning Phase:
Initialize: $\boldsymbol{\theta}_0, \boldsymbol{P}_{\boldsymbol{\theta},0}, \boldsymbol{m}_0, \boldsymbol{P}_{\boldsymbol{M},0}, \boldsymbol{u}_n,$ and $\boldsymbol{\Sigma}_n$ for $n = 1, 2, \ldots, N$
Parameters: $\boldsymbol{Q}_{\boldsymbol{\theta}}, \boldsymbol{Q}_{\boldsymbol{M}}, \lambda_{\boldsymbol{u}}, \lambda_{\boldsymbol{\Sigma}},$ and $\{R_{\boldsymbol{\theta}}^{(j)}, R_{\boldsymbol{M}}^{(j)}\}$ for $j = 1, 2, \ldots, M$
Repeat (for each episode):
  Initialize $\boldsymbol{s}_k$
  Repeat (for each agent i):
    While $\boldsymbol{s}_k^{(i)} \neq \boldsymbol{s}_T$ **do**:
      Reshape $\boldsymbol{m}_k$ into $L \times L$ to construct 2-D matrix $\boldsymbol{M}_k$.
      $a_k^{(i)} = \arg\max_a \left(\boldsymbol{g}_k^{(i)}(\boldsymbol{s}_k^{(i)}, a) \boldsymbol{g}_k^{(i)^T}(\boldsymbol{s}_k^{(i)}, a^{(i)}) \right)$
      Take action $a_k^{(i)}$, observe $\boldsymbol{s}_{k+1}^{(i)}$ and $r_k^{(i)}$.
      Calculate $\boldsymbol{\phi}(\boldsymbol{s}_k^{(i)}, a_k^{(i)})$ via Equations (54) and (56).
      Update reward weights vector: Perform MMAE to update $\boldsymbol{\theta}_k^{(i)}$.
      Update SR weights vector: Perform KF on Equations (71) and (72) to update $\boldsymbol{m}_k^{(i)}$.
      Update RBFs parameters: Perform RGD on the loss function L_k to update $\boldsymbol{\Sigma}_n$ and $\boldsymbol{u}_n$.
    end while

It is worth mentioning that, unlike the DNN-based networks for multi-agent scenarios, the proposed multiple-model frameworks require far less memory due to their sequential data processing nature. In other words, storing the whole episodes' information for all the agents is not needed as the last measured data (assuming one-step

Markov decision process) can be leveraged given the sequential nature of the incorporated filters. Finally, note that the proposed MAK-SR and MAK-TD frameworks are designed for systems with a finite number of actions. One direction for future research is to consider extending the proposed MAK-SR framework to applications where the to action-space is infinite-dimensional. This might occur in continuous control problems [186, 190] where number of possible actions at each state is infinite.

4.4 RL-based Fusion Strategy

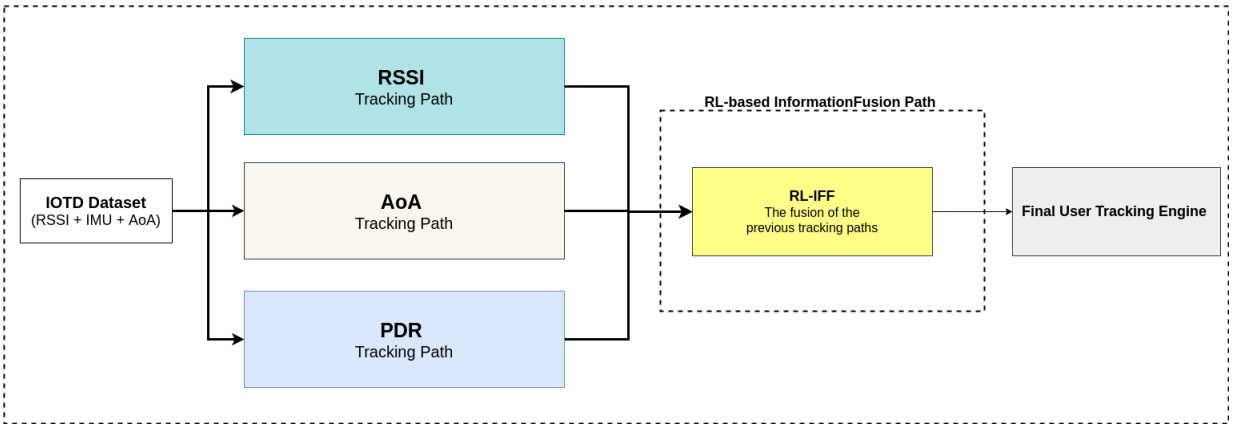


Figure 4.1: The proposed Localization Fusion framework.

In this section, we use the IoT-TD dataset introduced in chapter 3 and propose a track-fusion framework, that integrates estimated locations obtained from RSSI values with those computed based on the PDR and the AoA approaches. Fig. 4.1 represents the overl structure of the proposed fusion framework. As illustrated in Fig. 4.1, the fusion framework benefits from three different fusion scenarios between AoA, RSSI and PDR localization paths, which are then combined to form the final location estimates. The main objective is to find the optimal fusion weights, which is achieved by adjusting the RL learning procedure based on the following steps:

- Based on the information related to the target received from multi sensor and a generated signal via the RL solution, an action is taken and the corresponding reward will be calculated.
- The accumulative reward will be updated in each step taken in the environment.

By having the new target information, an action will be chosen leveraging an action selection scheme.

- Considering the time varying nature of the environment, an MDP is designed based on the fusion accuracy to address the challenges related to the different environment and movement scenarios.

Each specific indoor environment has its own structure, dimensions, and static/time-varying environmental variables, therefore, model-based fusion strategies become increasingly complex implementation-wise. In an alternative approach, we use Q-learning, insteat, to fuse the localization information gained from the RSSI, PDR, and AoA paths. The goal of such fusion task is to find the actions that maximize the reward gained from the environment to reach optimality in fused values. The data fusion process designed to be employed in this section is modeled as a MDP. In each time step, the RL-based fusion engine based on the user location is in a state s_k, by selecting an action from the Action set $\mathcal{A}$ and moving to the next state s_{k+1} will gain a scalar reward r_k based on the reward function $\mathcal{R}$. This learning scheme will continue by trial and error until reaching the terminal state.

The Q-function already is defined in (32), and the Q-learning is performed based on Bellman equation (34). The synchronized tracking information gained from three tracking paths (i.e., RSSI, PDR and AoA) will be fused (F_T) based on a specific weight as follows

$$\hat{\boldsymbol{x}}_k^{\text{Fused}} = \boldsymbol{w}^{\text{RSSI}}\boldsymbol{x}_k^{\text{RSSI}} + \boldsymbol{w}^{\text{PDR}}\boldsymbol{x}_k^{\text{PDR}} + \boldsymbol{w}^{\text{AoA}}\boldsymbol{x}_k^{\text{AoA}}, \tag{73}$$

where $\boldsymbol{x}_k^{\text{RSSI}}$, $\boldsymbol{x}_k^{\text{PDR}}$, $\boldsymbol{x}_k^{\text{AoA}}$ are the tracking results in the time stamp k and $\boldsymbol{w}^{\text{RSSI}}$, $\boldsymbol{w}^{\text{PDR}}$ and $\boldsymbol{w}^{\text{AoA}}$ are the associated fusion weight. In each time, this summation of the weights is static and is equal to one, i.e.,

$$\boldsymbol{w}^{\text{RSSI}} + \boldsymbol{w}^{\text{PDR}} + \boldsymbol{w}^{\text{AoA}} = 1. \tag{74}$$

The RL solution is used in fusion phase for updating the weights and finding the optimal weights to fuse the results of three tracking scenarios.

4.4.1 Data Fusion with Priori Knowledge

Considering the data gathered in different indoor locations and leveraging the IoT-TD dataset, we can have a priori knowledge of tracking scenarios in different indoor venues. There are many challenges related to the non-stationary RL solutions [191], specifically for the systems that are subjected to time-varying workloads. If the states in the tracking scenarios become defined based on the different physical locations or fingerprinting methods, the states would be continuous and hard to discretize. States' definition also can be very wide considering the type of the indoor environments and time varying user location scenarios. Considering the aforementioned challenges, the proposed RL-IFF is trying to deal with the time varying MDP, and transform it to a static MDP definition. In this regard, the localization accuracy error is considered as key metric to define the states and the proportional rewards gained by the agent. In what follows, state, action and reward function definition for the proposed RL solution are discussed thoroughly.

4.4.2 State, Action and Reward function Definition of the Proposed fusion Method

Considering an agent in position p_k ($p_k = (x_k, y_k)$), take a step in an environment and move to a new local position $p_{k+1}(p_{k+1} = (x_{k+1}, y_{k+1}))$, the tracking error can be the distance between the exact location of the user i.e., $p_{(r,k)} = (x_{(r,k)}, y_{(r,k)})$, and the newly fused estimation of the location, i.e., $p_{(e,k)} = (x_{(e,k)}, y_{(e,k)})$. This error ratio is used to define the states and the reward gained by the agent. We assume 100 states for this experiment, where the continuous error values are discretized to generate 100 states out of them. For the absolute error values between 0 to 1, we have 100 states as shown in Eq. (76), and for the absolute error values more than 1, the state would be 100.

$$\begin{aligned}\epsilon_k = \sqrt{(x_{(r,k)} - (\boldsymbol{w}^{\mathrm{RSSI}}\boldsymbol{x}_k^{\mathrm{RSSI}} + \boldsymbol{w}^{\mathrm{PDR}}\boldsymbol{x}_k^{\mathrm{PDR}} + \boldsymbol{w}^{\mathrm{AoA}}\boldsymbol{x}_k^{\mathrm{AoA}}))^2 + \\ (y_{(r,k)} - (\boldsymbol{w}^{\mathrm{RSSI}}\boldsymbol{y}_k^{\mathrm{RSSI}} + \boldsymbol{w}^{\mathrm{PDR}}\boldsymbol{y}_k^{\mathrm{PDR}} + \boldsymbol{w}^{\mathrm{AoA}}\boldsymbol{y}_k^{\mathrm{AoA}}))^2}\end{aligned} \tag{75}$$

$$s_k = \begin{cases} round(\epsilon_k * 100), & \text{if } 0 \leq \epsilon_k < 1 \\ 100, & \text{if } \epsilon_k >= 1 \end{cases} . \tag{76}$$

Considering the proposed RL solution's state definition, the reward function can be expressed as follows:

$$r_k = \begin{cases} 100, & \text{if } \epsilon_k = 0 \\ round(\frac{1}{round(\epsilon_k, 2)}), & \text{if } 0 < \epsilon_k < 1 \\ -100, & \text{if } \epsilon_k >= 1 \end{cases} . \tag{77}$$

Considering the reward function r_k equation where the reward is inversely proportional to the ϵ_k, the smaller distance between the actual location of the user and the estimated location, the larger the reward gained by the agent in the corresponding state would be. A significant negative reward will be issued for the ϵ_k values equal to or bigger than 1. The proposed RL-IFF aims to find the optimal weights to fuse the results of three tracking scenarios. Algorithm 3, briefly represents the proposed framework in this work. Different actions taken in this platform can be defined based on the weight adjustments. By modifying the weights, different fusion values can be calculated, and training can be performed based on the comparison between the new fused location estimation and the user's actual location in different states. Since there are three tracking scenarios and their underlying weights, based on Eq. (74), by assigning the new values to two of the weights, we can find the value of the third one. Consequently, the action function can be defined by $+$, $-$ and $<>$, denoting increased, decreased and unchanged operation on the weights respectively as follows:

$$\begin{cases} \boldsymbol{w}^{\mathrm{x}} = (1 + k\%)\boldsymbol{w}^{\mathrm{x}} & \text{if } \boldsymbol{w}^{\mathrm{x}}+ \\ \boldsymbol{w}^{\mathrm{x}} = (1 - k\%)\boldsymbol{w}^{\mathrm{x}} & \text{if } \boldsymbol{w}^{\mathrm{x}}- \\ \boldsymbol{w}^{\mathrm{x}} = \boldsymbol{w}^{\mathrm{x}} & \text{if } \boldsymbol{w}^{\mathrm{x}} <> \end{cases} , \tag{78}$$

where x is the tracking solution. Table 4.1, shows the possible actions regarding the modification of the weights.

Table 4.1: Actions in the RL-IFF for fusion strategy.

Action	Weight Modification
1	$\boldsymbol{w}^{\text{RSSI}}+$, $\boldsymbol{w}^{\text{AoA}}+$
2	$\boldsymbol{w}^{\text{RSSI}}+$, $\boldsymbol{w}^{\text{AoA}}-$
3	$\boldsymbol{w}^{\text{RSSI}}+$, $\boldsymbol{w}^{\text{AoA}}$ <>
4	$\boldsymbol{w}^{\text{RSSI}}-$, $\boldsymbol{w}^{\text{AoA}}-$
5	$\boldsymbol{w}^{\text{RSSI}}-$, $\boldsymbol{w}^{\text{AoA}}+$
6	$\boldsymbol{w}^{\text{RSSI}}-$, $\boldsymbol{w}^{\text{AoA}}$ <>
7	$\boldsymbol{w}^{\text{RSSI}}$ <>, $\boldsymbol{w}^{\text{AoA}}+$
8	$\boldsymbol{w}^{\text{RSSI}}$ <>, $\boldsymbol{w}^{\text{AoA}}-$
9	$\boldsymbol{w}^{\text{RSSI}}$ <>, $\boldsymbol{w}^{\text{AoA}}$ <>

Algorithm 3 The Proposed RL-based Information Fusion Framework: RL-IFF

Learning Phase:
Input: The tracking results of the applied tracking paths, i.e., the RSSI path; the PDR path, and; the AoA path;
Output: The optimal set of weights, fused data;
Initialize $Q = 0$ with random weights $\boldsymbol{w}^{\text{RSSI}}$, $\boldsymbol{w}^{\text{AoA}}$ and $\boldsymbol{w}^{\text{PDR}}$;
Set parameters γ and α for Q-learning;
Initialize state s_0, leveraging the initial weights in Eq. (76);
Repeat for each episode:
  $a_k \leftarrow$ epsilon greedy action selection in s_k
  Take action a_k, observe the next state s_{k+1};
  Calculate the gained reward using Eq. (77);
  $Q(\boldsymbol{s}_k, a_k) \leftarrow (1-\alpha) Q(\boldsymbol{s}_k, a_k) + \alpha \Big(r_k + \gamma \arg\max_{a \in \mathcal{A}} Q(\boldsymbol{s}_{k+1}, a_k) \Big)$;
  $\boldsymbol{w}^{\text{RSSI}}$, $\boldsymbol{w}^{\text{AoA}}$ and $\boldsymbol{w}^{\text{PDR}} \leftarrow$ optimal weights that maximize $Q(\boldsymbol{s}_{k+1}, a_k)$;
  $s_k \leftarrow$ optimal new state s^*;
end for
return $\boldsymbol{w}^{\text{RSSI}}$, $\boldsymbol{w}^{\text{AoA}}$ and $\boldsymbol{w}^{\text{PDR}}$; fused data p_k .

4.5 Experiments and results

The performances of the proposed MAK-SR and MAK-TD frameworks are evaluated in this section, where a multi-agent extension of the OpenAI gym benchmark is utilized. Figure 4.2 illustrates snapshots of the environment utilized for evaluation of the proposed approaches. More specifically, a two-dimensional world is implemented to simulate competitive, cooperative, and/or mix interaction scenarios [192]. The utilized benchmark is currently one of the most standard environments to test

different multi-agent algorithms, where time, discrete action space, and continuous observations are the basics of the environment. Such a multi-agent environment is a natural curriculum in that the environment difficulty is determined based on the skills of the agents cooperating or competing. The environment does not have a stable equilibrium, therefore, allowing the participating agents to become smarter irrespective of their intelligence level. In each step, the implemented environment provides observations and rewards once the agents performed their actions. The proposed platforms are implemented on a computer with a 3.79 GHz AMD Ryzen 9, 12-core processor. The frameworks are evaluated via several experiments, which are implemented through the OpenAI Gym multi-agent RL benchmarks. The parameters related to the proposed MAK-SR and MAK-TD are set randomly. In the designed deep models, the learning rate is set as 0.001, and the models are trained with the mini-batches of size 128 using Adam Optimizer. $MADDPG$ and $DDPG$ are based on the Actor-Critic approach. DQN and $DDPG$ receive an observation as input consisting of the current state, next state, gained reward, and the action taken by the agents at each step in the environment. For $MADDPG$, based on the received state data (current and next state) and the actions taken by all the agents, the future return is approximated considering all the agent's policies.

In what follows, we discuss different multi-agent environments exploited in this work as well as the experimental assumptions considered during testing of the proposed methods. Finally, the results of the experiments will be represented and explained.

4.5.1 Environments

In the represented multi-agent environments, we do not impose any assumption or requirement on having identical observations or action spaces for the agents. Furthermore, agents are not restricted to follow the same policy π while playing the game. In the environments, a different number of agents and possible landmarks can be placed to establish different interactions such as cooperative, competitive, or mixed strategies. The strategy in each environment is to keep the agents in the game as long as possible. Each test can be fully cooperative when agents communicate to maximize a shared return, or can be fully competitive when the agents compete to achieve different goals. The mixed scenario for the predator–prey environments

(a variant of the classical predator–prey) is defined in a way that a group of slower agents must cooperate against another group of faster agents to maximize their returned reward. Each agent takes a step by choosing one of five available actions, i.e., no movement, left, right, up, and down, transiting to a new state, and receiving a reward from the environment. Moreover, each agent will receive a list of observations in each state, which contains the agent's position and velocity, relative positions of landmarks (if available), and its relative position to other agents in the environment. That is how an agent knows the position and general status of the agents (friends and adversaries), enabling the decision-making process of that agent. As shown in Figure 4.2, each environment has its own margins. An agent that leaves the area will be punished by -50 points, the game will be reset, and a random configuration will be initiated to start the next state, which begins immediately. The red agents play the predator role and receive $+100$ points intercepting (hunting) a prey (small green agents). The green agents that are faster than red agents (predators) will receive -100 points by each interception with the red ones. As their job is to follow the prey, the predators will be punished proportionally to their distance to the prey (green agents). In contrast, the opposite will happen to the green agents as they keep the maximum distance from the predators. The proposed MAK-TD/SR frameworks are evaluated against DQN [112], DDPG [113], and MADDPG [187]. We evaluate the algorithms in terms of loss, returned discounted reward, and the number of collisions between agents.

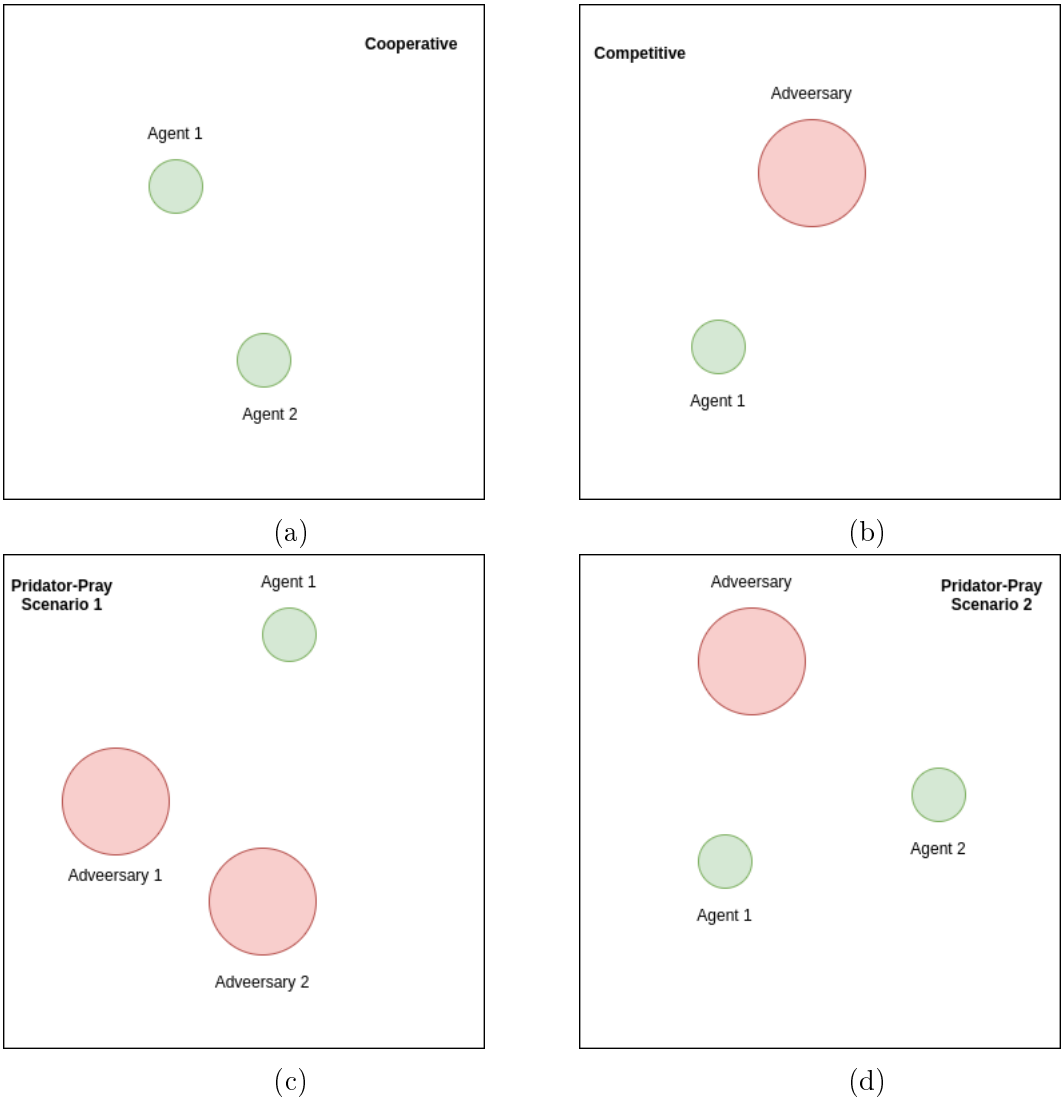


Figure 4.2: Different multi-agent scenarios implemented within the OpenAI gym. (**a**) Cooperation Scenario (**b**) Competition Scenario (**c**) Predator-Prey 2v1, and (**d**) Predator-Prey 1v2

4.5.2 Experimental Assumptions

In the proposed frameworks, we exploit related RBFs based on the different agents' sizes of observations and a bias parameter. The size of the observation vector at each local agent (localized observation vector), which represents the number of global and local measurements available locally, varies across different scenarios based on the type and the number of agents present/active in the environment. Irrespective of size of the localized observation vectors, the size of the localized feature vectors, which

represents the available five actions, is considered to be 50. Mean and covariance of the RBFs are initialized randomly for all the agents in all the environments. For example, consider a Predator–Prey scenario with 2 preys optimizing their actions against one predator. In this toy-example (discussed for clarification purposes), considering 9 RBFs together with localized observation vectors of size 12 for the predator and 10 for the preys, the mean vector associated with the predator and the preys are of dimensions 9×12 and 9×10, respectively. Consequently, for this Predator–Prey scenario, $\boldsymbol{\mu}$, which is initialized randomly contains three agents with random values with the mean size $((9,12),(9,10),(9,10))$ and the covariance, $\boldsymbol{\Sigma} = (\boldsymbol{I}_{12}, \boldsymbol{I}_{10}, \boldsymbol{I}_{10})$ where $\boldsymbol{I}_{12}$ and $\boldsymbol{I}_{10}$ are the identity matrices of size (12×12) and (10×10), respectively. Based on Equation (56), the vector of basis function is represented as follows:

$$\boldsymbol{\phi}(\boldsymbol{s}_k, a_k = -2) = [0,\ldots,0,0,\ldots,0,1,\phi_{1,a_d},\ldots\phi_{9,a_d},0,\ldots,0,0,\ldots,0]^T, \quad (79)$$

$$\boldsymbol{\phi}(\boldsymbol{s}_k, a_k = -1) = [0,\ldots,0,0,\ldots,0,0,\ldots,0,1,\phi_{1,a_d},\ldots\phi_{9,a_d},0,\ldots,0]^T, \quad (80)$$

$$\boldsymbol{\phi}(\boldsymbol{s}_k, a_k = 0) = [0,\ldots,0,0,\ldots,0,0,\ldots,0,0,\ldots,0,1,\phi_{1,a_d},\ldots\phi_{9,a_d}]^T, \quad (81)$$

$$\boldsymbol{\phi}(\boldsymbol{s}_k, a_k = +1) = [0,\ldots,0,1,\phi_{1,a_d},\ldots,\phi_{9,a_d},0,\ldots0,0,\ldots,0,0,\ldots,0]^T, \quad (82)$$

and $$\boldsymbol{\phi}(\boldsymbol{s}_k, a_k = +2) = [1,\phi_{1,a_d},\ldots,\phi_{9,a_d},0,\ldots,0,0,\ldots,0,0,\ldots0,0,\ldots0]^T, \quad (83)$$

where ϕ_{l,a_d} is calculated based on Equation (55) for ($l \in \{1,2,\ldots,9\}$. , γ, In all the scenarios, the time step chosen to be 10 milliseconds and the discount factor is 0.95. The transition matrix is initiated to $\boldsymbol{F} = \boldsymbol{I}_{50}$, and for the process noise covariance, a small value of $\boldsymbol{Q}_k = 10^{-7}\boldsymbol{I}_{50}$ is considered. The covariance matrix associated with the noise of the measurement model is selected from the following set:

$$R^{(i)} \in \{0.01, 0.1, 0.5, 1, 5, 10, 50, 100\}. \quad (84)$$

For initializing the weights, we sample from a zero mean Gaussian initialization distribution $\mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{P}_{\boldsymbol{\theta},0})$, where $\boldsymbol{\theta}_0 = \boldsymbol{0}_{50}$ and $\boldsymbol{P}_{\boldsymbol{\theta},0} = 10\boldsymbol{I}_{50}$. By considering the aforementioned initial parameters, each experiment is initiated randomly and consists of 1000 learning episodes together with 1000 test episodes. Given small number of available learning episodes, the proposed MAK-TD/SR frameworks outperformed their counterparts across different metrics including sample efficiency, cumulative reward, cumulative steps, and speed of the value function convergence.

4.5.3 MAK-TD/SR Results

Initially, the agents are trained over different number of episodes, after which 10 iteration each of 1000 episodes is implemented for testing to compute different results evaluating performance and efficiency of the proposed MAK-TD/SR frameworks. First, to evaluate stability of the incorporated RBFs, a Monte Carlo (MC) study is conducted where 10 RBFs are used across all the environments. The results are averaged over multiple realizations leveraging MC sampling as shown in Tables 5.1–5.4.

Table 4.2: Total loss averaged across all the episodes and for all the four implemented scenarios.

Environment	MAK-SR	MAK-TD	MADDPG	DDPG	DQN
Cooperation	8.93	2.4088	9649.84	10561.16	10.93
Competition	0.43	4.9301	10158.18	10710.37	107.39
Predator–Prey 1v2	0.005	1.9374	6816.34	6884.33	8.21
Predator–Prey 2v1	8.87	1.2421	7390.18	6882.2	10.24

Figure 4.4b shows the rewards gained by all the agents in a Predator–Prey environment. It is worth mentioning that the average number of the steps taken by all the agents in the defined environments is also represented in Table 5.4, showing MAK-SR remarkable results in contrast with the other algorithms. Results related to cumulative distance walked by the agents (computed by multiplying the number of the steps by 0.74 meter for each step) are also shown in Figure 4.10 for different environments admitting superiority of the MAK-SR framework in contrast with other solutions. The loss function associated with each of the five implemented methods is shown in Figure 4.5.

Table 4.3: Total received reward by the agents averaged for all the four implemented scenarios.

Environment	MAK-SR	MAK-TD	MADDPG	DDPG	DQN
Cooperation	-16.0113	-23.0113	-69.28	-66.29	-39.96
Competition	-0.778	-13.358	-63.30	-61.34	-14.49
Predator–Prey 1v2	-0.0916	-13.432	-46.17	-20.53	-23.451
Predator–Prey 2v1	-0.081	-17.0058	-55.69	-49.41	-44.32

Table 4.4: Average steps taken by agents per episode for all the environments based on the implemented platforms.

Environment	MAK-SR	MAK-TD	MADDPG	DDPG	DQN
Cooperation	14.03	12.064	7.377	7.369	15.142
Competition	17.59	17.48	7.36	7.18	11.98
Predator–Prey 1v2	14.78	12.36	6.21	7.69	10.02
Predator–Prey 2v1	9.94	9.773	6.25	7.12	8.46

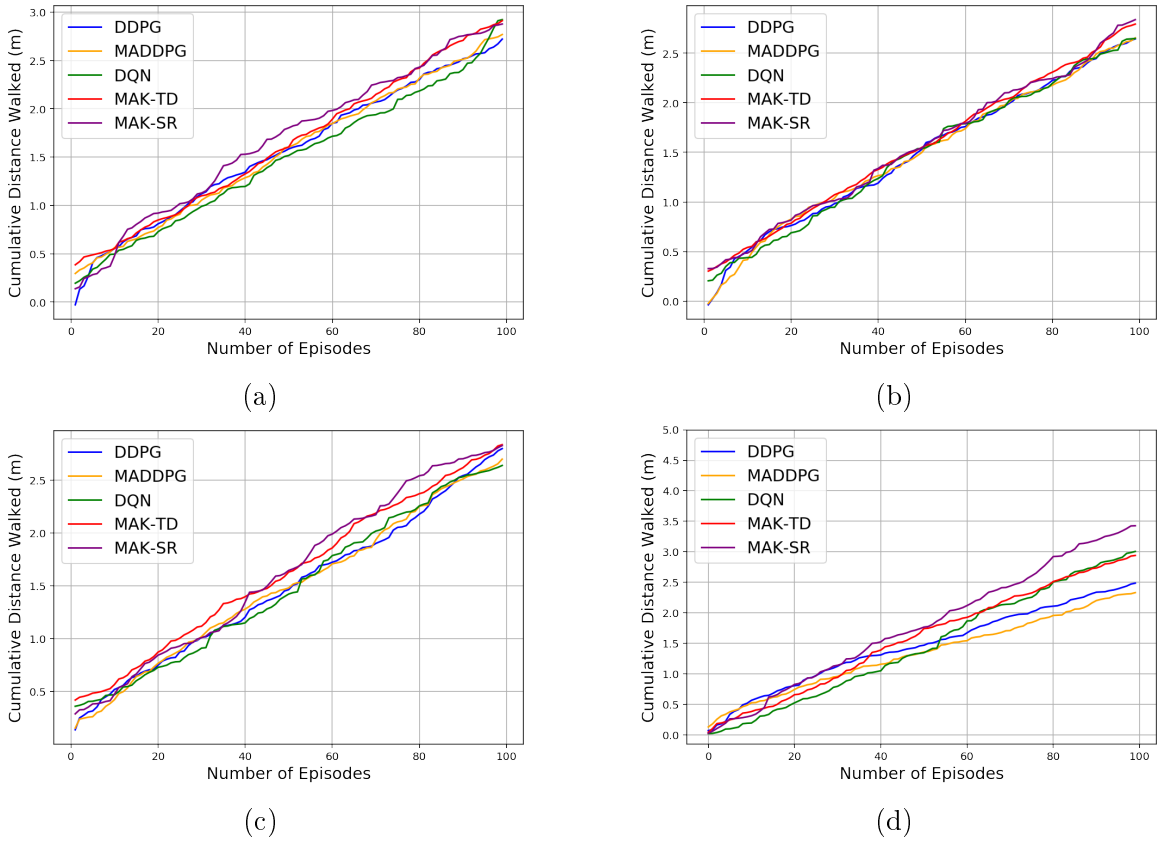
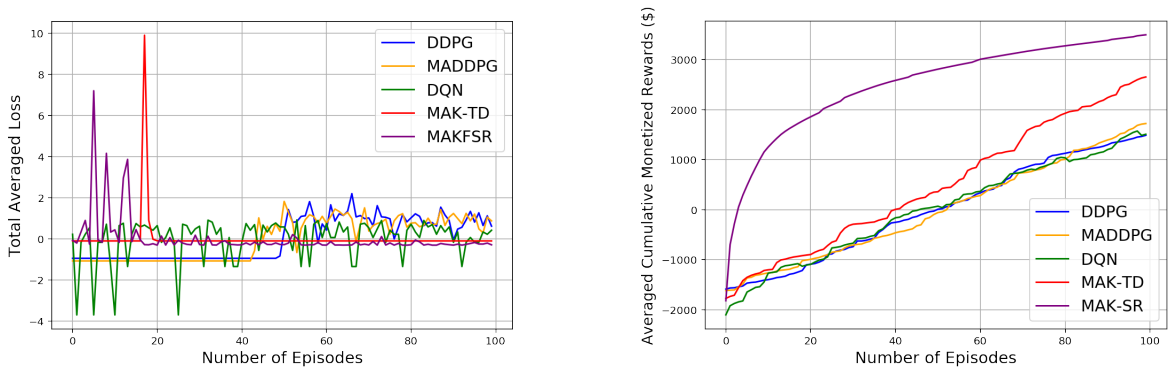
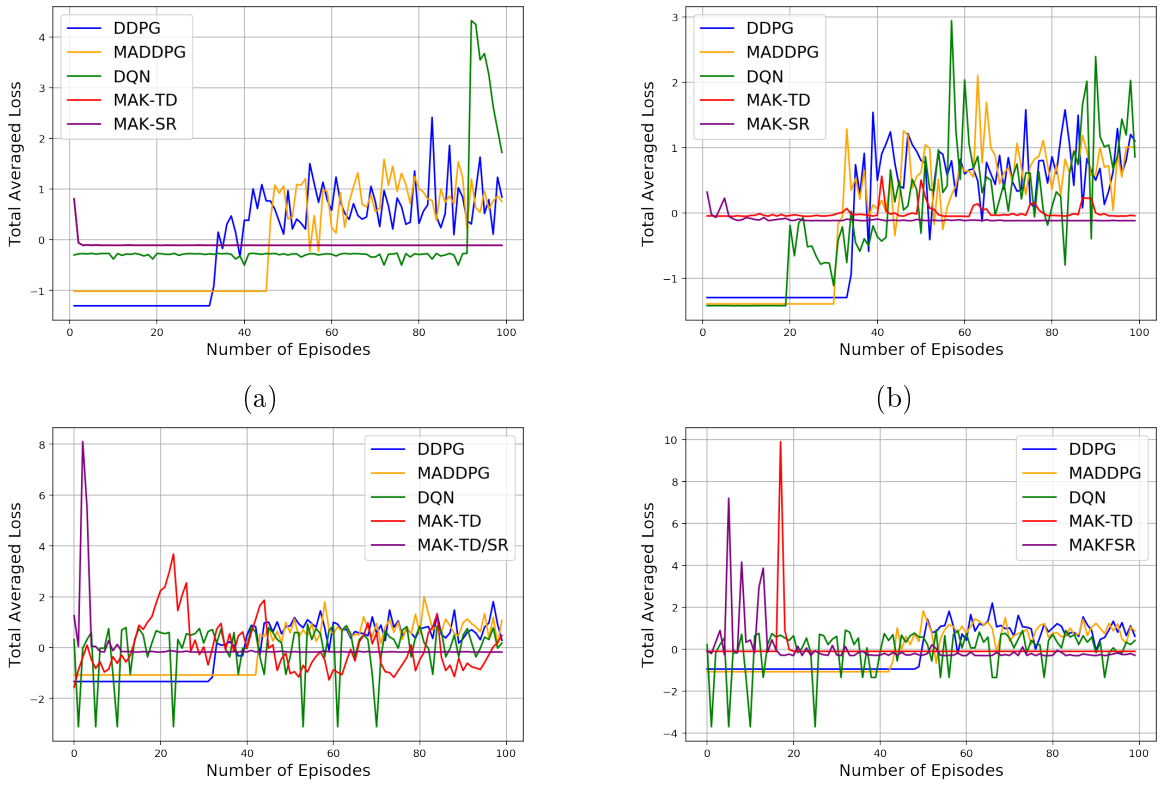


Figure 4.3: Cumulative distance walked by the agents in four different environments based on the five implemented algorithms (**a**) Cooperation. (**b**) Competition. (**c**) Predator–Prey 2v1. (**d**) Predator–Prey 1v2.



(a) (b)

Figure 4.4: The Predator–Prey environment: (**a**) Loss. (**b**) Received rewards.



(a) (b)

(c) (d)

Figure 4.5: Four different normalized loss functions results for all the agents in the for the four algorithms in four different environments: (**a**) Cooperation. (**b**) Competition. (**c**) Predator–Prey 2v1. (**d**) Predator–Prey 1v2.

4.5.4 SR-based RL Discussions

The results shown in Section 4.5.5 illustrate the inherent stability of the utilized RBFs and the proposed MAK-TD and MAK-SR frameworks. Capitalizing on the results of Tables 5.1–5.4, the MAK-SR can be considered as the most sample-efficient approach. It is worth noting that although MAK-SR outperforms the MAK-TD approach, we included both, as the learned representation is not transferable between optimal policies in the SR learning. For such scenarios, MAK-TD is an alternative solution providing, more or less, similar performance to that of the MAK-SR. To be more precise, when solving a previously unseen MDP, a learned SR representation can only be used for initialization. In other words, the agents still have to adjust the SR representation to the policy, which is only optimal within the existing MDP. This limitation urges us to represent the MAK-TD as another trusted solution.

As it can be seen from Table 5.1, the average loss associated with the proposed MAK-SR is better than that of the MAK-TD. Both frameworks, however, outperform their counterparts, which can be attributed to their improved sample selection efficiency. This excellence can also be seen for the Predator–Prey $1v2$ environment in Figure 4.4a. The calculated losses mostly have small values after the beginning of the experiments, indicating stability of the implemented frameworks. As can be seen, other approaches cannot provide that level of performance that is achieved by MAK-SR and MAK-TD with such low number of training episodes in this experiment. The other three DNN-based approaches can reach such an efficiency with a much greater amount of experience (more than $10,000$ experiments) and use much more memory space to save the batches of the information.

As can be seen in Table 5.5 and Figure 4.4b, the rewards gained in the MAK-SR are also better than those of the MAK-TD and are much higher than the other approaches. This can be considered exceptional considering the limited utilized experience. For all other environments, this better performance in the gained reward can be seen in Figure 4.6 where four different reward functions for five discussed algorithms in four experiment environments are shown. As expected, the performance of each model improves over time as being trained through different training episodes. The proposed MAK-SR and MAK-TD provide exceptional performances given the small number of training episodes utilized in these experiments. $MADDPG$, $DDPG$, and DQN, however, fail to achieve the same performance level.

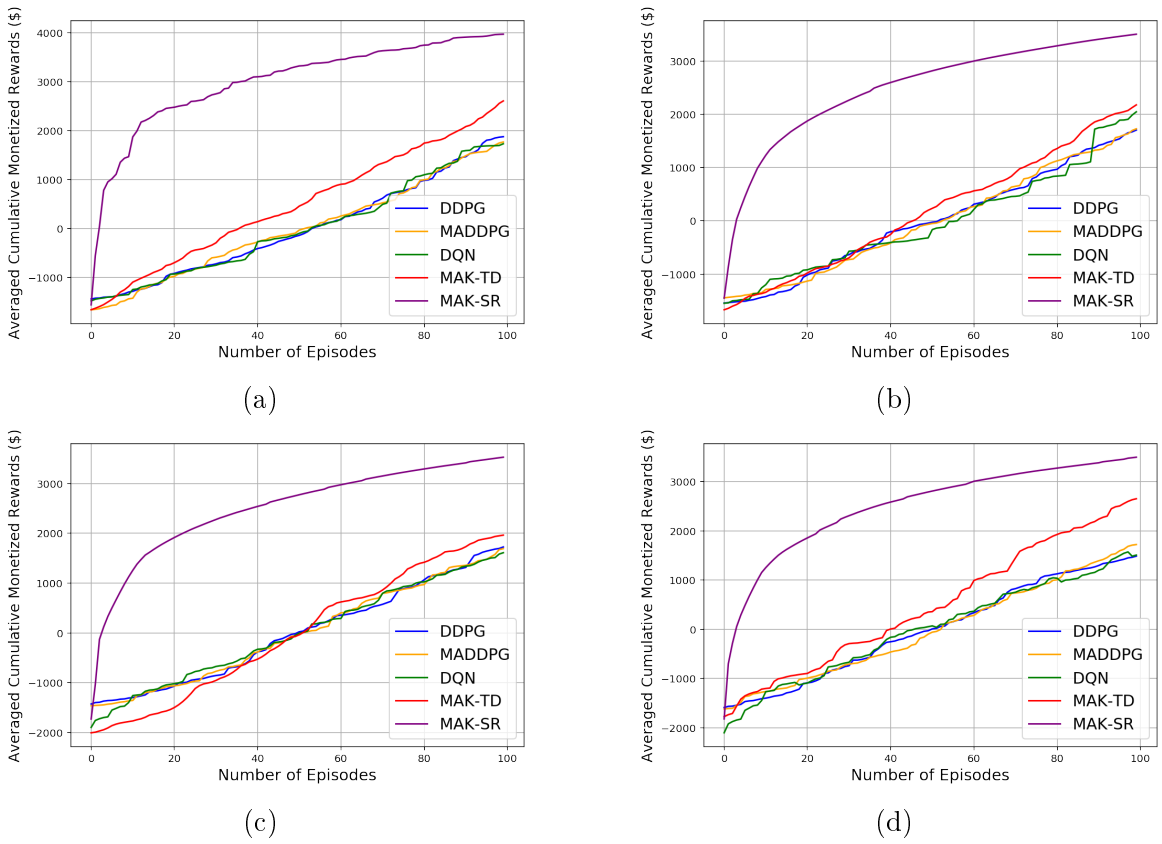


Figure 4.6: Four different reward functions results for all the agents for the five algorithms in four different environments: (**a**) Cooperation. (**b**) Competition. (**c**) Predator–Prey 2v1. (**d**) Predator–Prey 1v2.

Evaluating reliability of the proposed learning frameworks is of significance to verify their applicability in real-world scenarios. A reliable learning procedure should be able to provide consistency in its performance and generate reproducible results over multiple runs of the model [135]. Generally speaking, performance of RL-based solutions, particularly DNN-based approaches, are highly variable because of their dependence on a large number of tunable parameters. Hyperparameters, implementation details, and environmental factors are among these parameters [193]. This can result in unreliability of DNN-based RL algorithms in real-world scenarios compared to the proposed frameworks that are less dependent on parameter selection and fine-tuning. To better illustrate reliability of the proposed frameworks, another experiment is conduced where the initial parameters in each run are generated randomly.

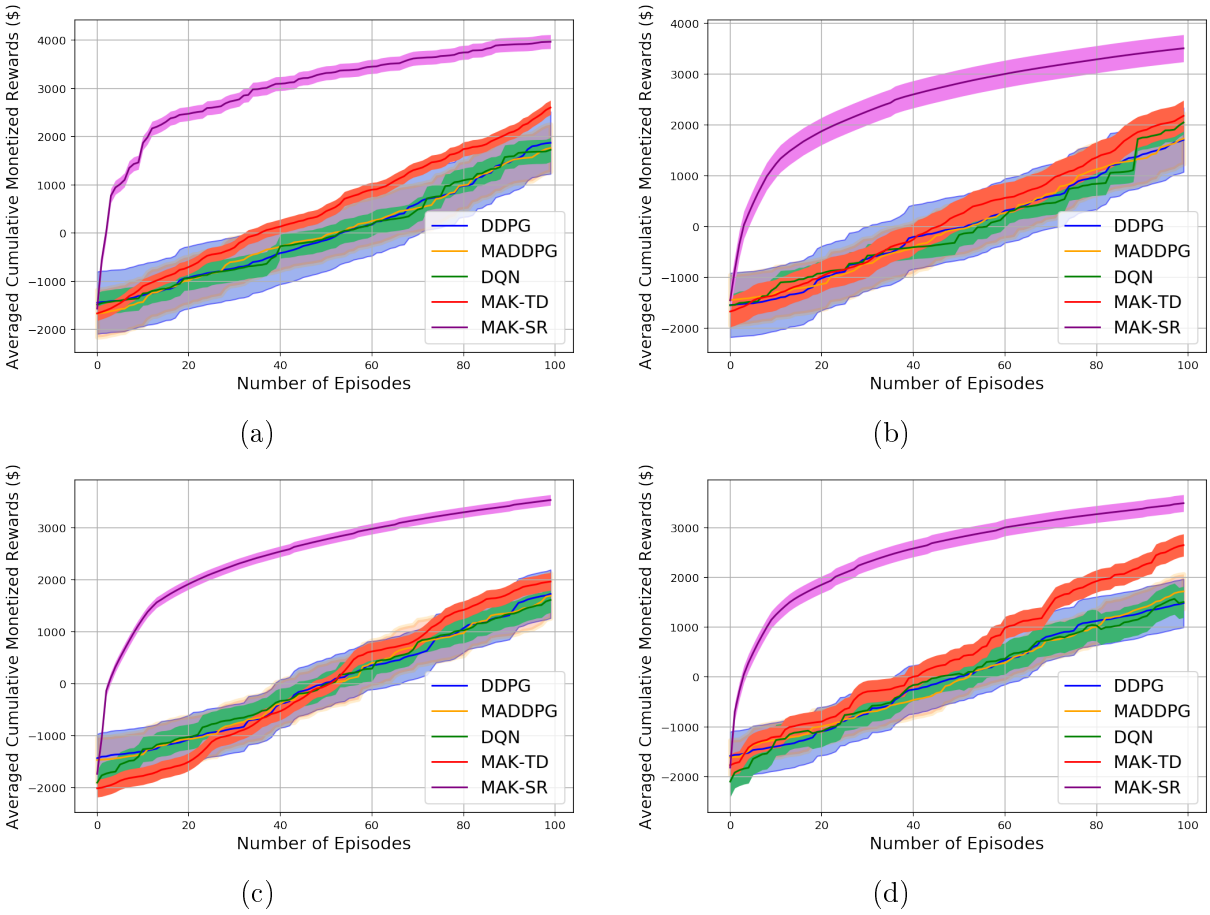


Figure 4.7: The mean (solid lines) and standard deviation (shaded regions) of cumulative episode's reward for the four algorithms in four different environments: (**a**) Cooperation. (**b**) Competition (**c**) Predator–Prey 2v1. (**d**) Predator–Prey 1v2.

More specifically, we have repeated each test 10 times consisting of 1000 learning episodes together with 1000 test episodes. A reliable RL algorithm should be consistent in regenerating performance across different training sessions, i.e., reproducibility feature. As can be seen from Figure 4.7, for all four test scenarios (i.e., cooperative, competitive, and mixed strategies) DNN-based methods (*MADDPG*, *DDPG*, and *DQN*) have higher variance illustrating their sensitivity to the underlying parameters that can be attributed to reduced reliability. As can be seen from Figure 4.7, MAK-SR outperforms other approaches in terms of the received awards. In both MAK-SR and MAK-TD algorithms, positive effect of uncertainty usage in the action selection procedure is noticeable. The ability to produce stable performance across different episodes is another aspect for investigating reliability of RL models.

Stability of different models can also be compared through Figure 4.7. It can be seen that the proposed MAK-SR algorithm is more stable than its counterparts as fewer sudden changes occur during different episodes.

With regards to potential future works, on the one hand, the proposed frameworks can be implemented and applied to higher-dimensional MARL environments, e.g., large-scale IoT applications such as indoor localization scenarios in unconstrained environments. One interesting scenario here is to consider a heterogenous network of multiple agents using different tracking/localization algorithms with application to Contact Tracing (CT). Another direction for future research is to focus on optimization of the current SR-based solution. In its current form, the SR weight matrix is approximated by mapping into a one-dimensional vector and applying KF leveraging the KTD framework. For application to higher dimensions, this vectorized approach can result in potential information loss as such more complex approximation techniques should be developed while being mindful of potential computation overhead.

4.5.5 Experimental Results Information Fusion Indoor Localization

In this section, we evaluate various localization techniques using the ground truth IoT-TD outlined in Chapter 3. These localization methods are based on AoA, RSSI, Pedestrian Dead Reckoning (PDR), and the proposed RL-based Fusion approach (IoT-TD).

To the best of our knowledge, current indoor localization and tracking applications are designed for predefined user movements. However, tracking random movements or movements with changes in direction or walking patterns remains a significant challenge. These predefined scenarios do not provide a fair comparison with the ground truth data, which reveals the user's exact location during the experiment. As a result, these scenarios inherently add a default error as the user's movement pattern is already assumed.

In this section, we evaluate the proposed tracking methods, including the RSSI path, PDR path, AoA path, and the RL-based information fusion framework, IoT-TD. The evaluations are performed on a computer with a 3.79 GHz AMD Ryzen 9, 12-core processor. Each experiment is conducted in all environments and repeated 50 times to ensure accuracy. The RL-based information fusion scenario undergoes 1000000

learning episodes and 1000 test episodes to improve accuracy and assess the reliability of the proposed fusion approach.

The RL-IFF is utilized to determine the optimal weights for information fusion, resulting in improved tracking performance. The effectiveness of the fusion approach is shown in Figs.4.8,4.9 and 4.10. The results of the localization estimation in three

Table 4.5: The MSE of the location estimation based on the dataset gathered in the FIRST environment by comparing the ground truth (collected with the Vicon system) and the proposed RL-IFF.

Movement Scenario	AoA	RSSI	PDR	RL-IFF
Rectangular	0.01979	0.12994	0.21997	0.003973
Random	0.02508	0.14502	0.17303	0.00495
Diagonal-A	0.0439	0.1961	0.2925	0.00685
Diagonal-B	0.01992	0.16002	0.10029	0.00504

Table 4.6: The MSE of the location estimation based on the dataset gathered in the SECOND environment by comparing the ground truth (collected with the Vicon system) and the proposed RL-IFF.

Movement Scenario	AoA	RSSI	PDR	RL-IFF
Rectangular	0.02193	0.05307	0.09464	0.006393
Random	0.10133	0.09432	0.16103	0.00799
Diagonal-A	0.11368	0.12331	0.19331	0.00994
Diagonal-B	0.06854	0.11268	0.1105	0.00863

Table 4.7: The MSE of the location estimation based on the dataset gathered in the THIRD environment by comparing the ground truth (collected with the Vicon system) and the proposed RL-IFF.

Movement Scenario	AoA	RSSI	PDR	RL-IFF
Rectangular	0.12379	1.63892	1.5869	0.0077
Random	0.02923	0.1123	0.14699	0.00745
Diagonal-A	0.01468	0.22397	0.29195	0.00799
Diagonal-B	0.01698	0.06556	0.0961	0.00648

different environments are presented in Tables 4.5, 4.6, and 4.7. These tables compare the Mean Squared Error (MSE) in meters of the primary localization approaches, including AoA, RSSI with KF-PF and PDR, and the proposed RL-IFF framework. The results show that the RL-IFF consistently outperforms the other algorithms, demonstrating the high potential of the proposed RL-based solution.

Table 4.8: Sample weight adjustment offered by the proposed RL-IFF, for information fusion of tracking scenarios.

Environment	Movement Scenario	$\boldsymbol{w}^{\mathbf{AoA}}$	$\boldsymbol{w}^{\mathbf{RSSI}}$	$\boldsymbol{w}^{\mathbf{PDR}}$
First Environment	**Rectangular**	0.1302776	0.7583678	0.1113545
	Random	0.513865	0.825227	-0.339092
	Diagonal-A	0.2742709	0.4277489	0.2979801
	Diagonal-B	0.1022387	0.5218031	0.3759582
Second Environment	**Rectangular**	0.072684	1.0238779	-0.0965619
	Random	0.1701491	0.6915486	0.1383023
	Diagonal-A	-0.3430738	0.734489	0.6085848
	Diagonal-B	0.3344642	0.465767	0.1997688
Third Environment	**Rectangular**	-0.693659	1.0109855	0.6826736
	Random	0.0188074	0.77122	0.2099726
	Diagonal-A	0.0858805	0.5444015	0.3697179
	Diagonal-B	0.2267579	0.7415241	0.0317179

In Table 4.8, a sample of the weight adjustment task performed by the RL-IFF framework is provided. This table displays the fusion weights, $\boldsymbol{w}^{\mathrm{AoA}}$, $\boldsymbol{w}^{\mathrm{RSSI}}$, and $\boldsymbol{w}^{\mathrm{PDR}}$, for each environment and moving scenario. The accuracy of the models can be improved by increasing the number of particles (N_p), but this comes at the cost of additional computation overhead. In offline modes, this trade-off can be considered, but in online modes, the limitations make the number of particles a sensitive factor that cannot be increased for higher accuracy.

The data collection sessions for these experiments took place in three different environments to account for potential interference and environmental impacts. The results are discussed based on these three environments and the accuracy of the models is evaluated.

First Environment

The first environment (3.5 meters to 3.5 meters) is a large room with minimal obstacles and positional items during the experiments. The walls surrounding the environment are mainly made of wooden material covered with wall paintings. The results of different fusion strategies based on various movement scenarios are shown in Fig.4.8. The final results of these fusion strategies are plotted in black to compare with other basic approaches, particularly with the ground truth results shown in blue. The results in the environment with the fewest obstacles are expected to outperform

the others. As seen in Table4.5, the final MSE of the proposed RL-IFF is less than that of the other two environments for all trajectories. The other environments are more susceptible to noise and potential conflicts due to obstacles and the effects of digital and electronic systems installed.

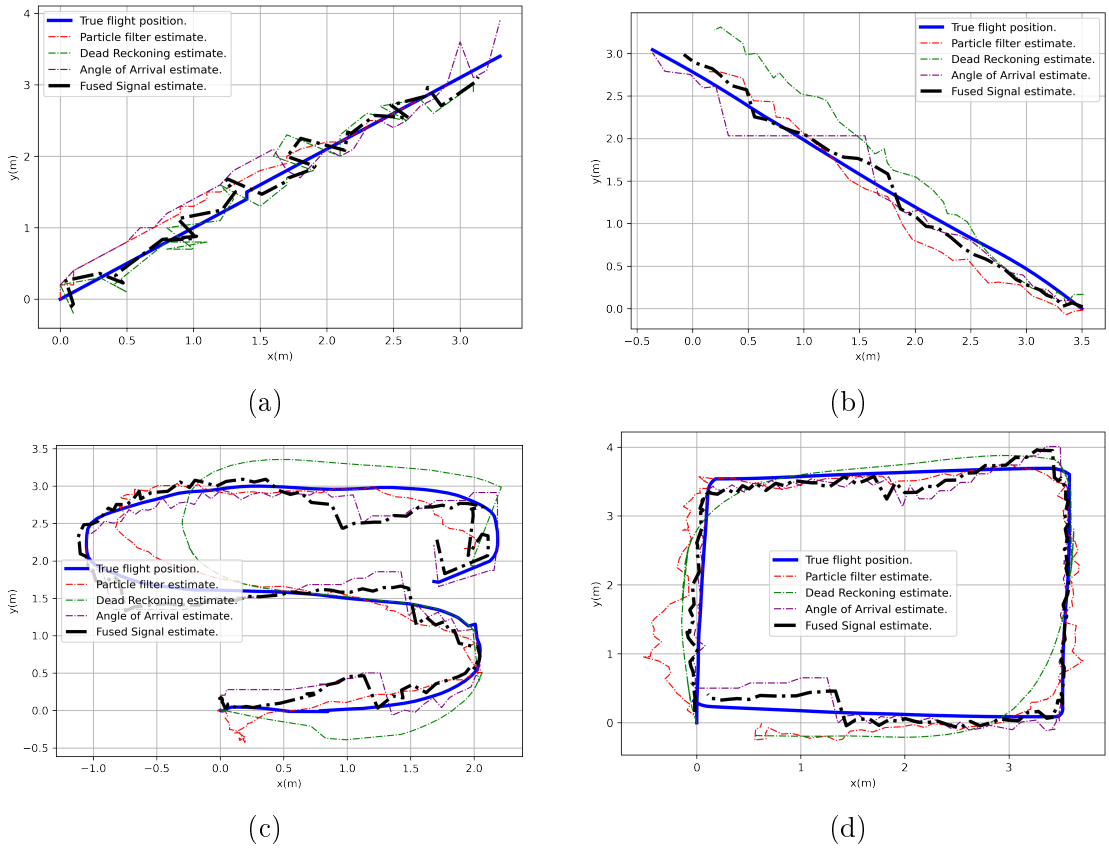


Figure 4.8: RL-IFF Results for Four different movement scenarios in the first environment: (a) Diagonal A; (b) Diagonal B; (c) Random, and (d) Rectangular trajectories.

Second Environment

The Second environment (2.5 meters to 3.5 meters) is a smaller room with various obstacles and positional objects present during the experiment. The walls in this environment are composed of a combination of concrete and wooden materials. Additionally, there are various computer systems connected to the Internet and electronic devices that are active during the experiment. The results of the proposed RL-IFF are shown in Table 4.6 and Fig. 4.9. In these figures, the RL-IFF results are plotted in black and compared to the ground truth of the target's movements, which are

shown in blue. This comparison provides a clear illustration of the accuracy of the RL-IFF approach in this particular environment, which is characterized by obstacles, electronic devices, and potential sources of interference.

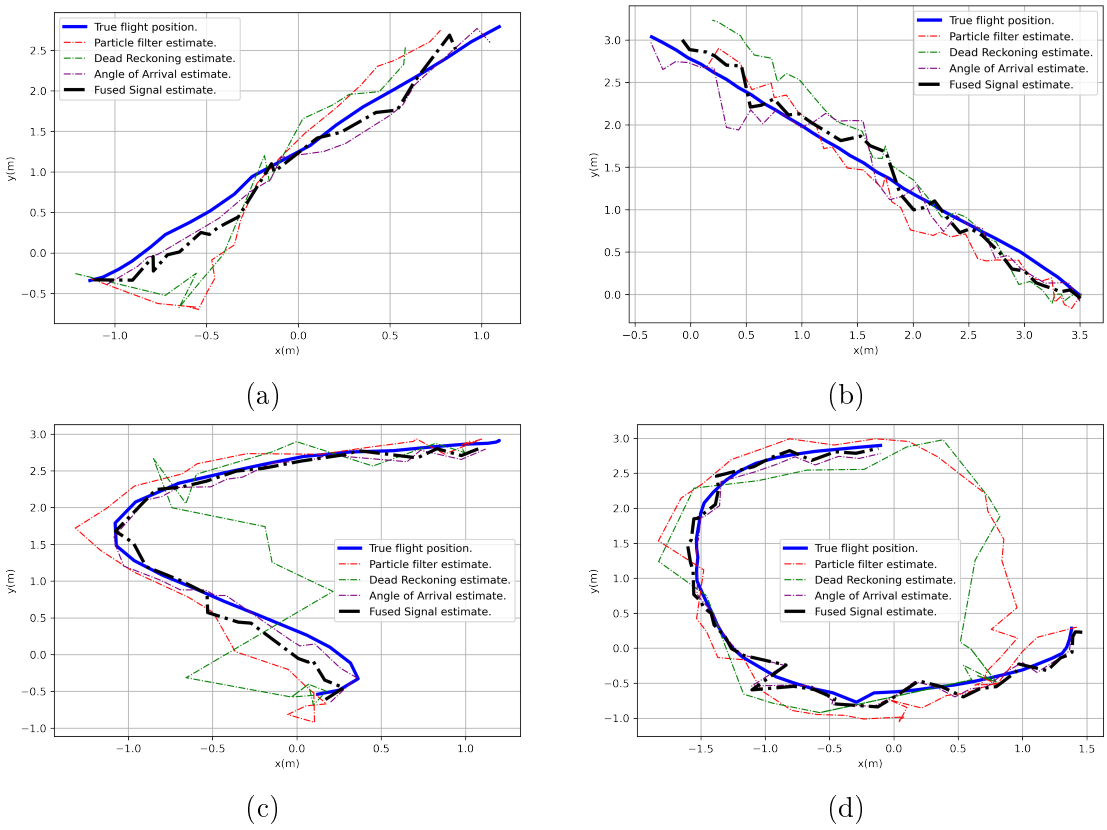


Figure 4.9: RL-IFF Results for Four different movement scenarios in the second environment: (a) Diagonal A; (b) Diagonal B; (c) Random, and (d) Rectangular trajectories.

Third Environment

The third environment, with dimensions of 2.5 meters to 3.5 meters, presents a different challenge for user tracking. This environment, unlike the first, is a small room surrounded by metal plate walls, creating a noisy atmosphere. Additionally, numerous electrical devices were installed during the data gathering session, further complicating the tracking process. The results of the proposed RL-IFF framework in this environment can be seen in Table 4.7 and Fig. 4.10. The RL-IFF result is plotted in black and compared to the ground truth in blue, making it easier to evaluate the performance of the framework. The results show that the RL-IFF framework outperforms other tracking solutions, providing the best tracking result in this challenging

environment.

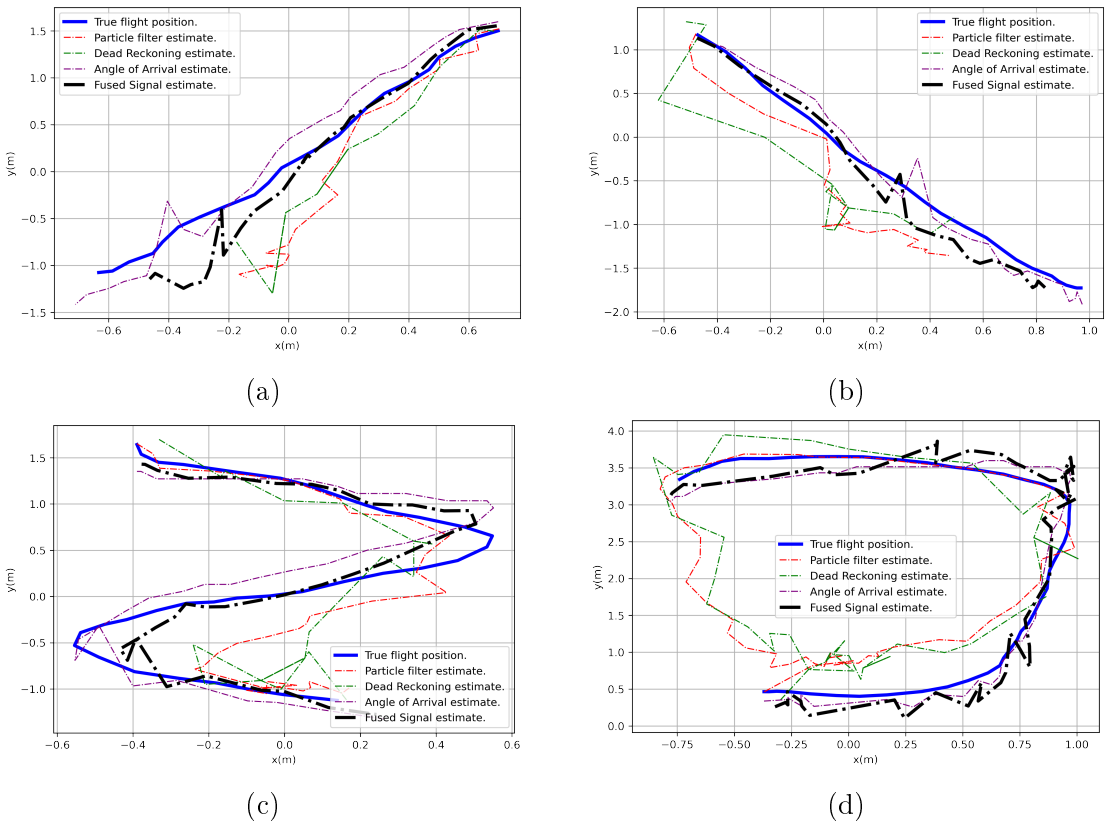


Figure 4.10: RL-IFF Results for Four different movement scenarios in the third environment: (a) Diagonal A; (b) Diagonal B; (c) Random, and (d) Rectangular trajectories.

Reliability of the Proposed RL Solution

To evaluate the effectiveness of the proposed RL-IFF in real-world scenarios, it is crucial to assess its reliability. A trustworthy learning procedure should consistently generate reproducible results across multiple runs of the model [2,194]. Unfortunately, the performance of RL-based solutions, especially those based on DNN, can be highly variable due to the dependence on a large number of tunable parameters such as hyperparameters, implementation details, and environmental factors [2].

To address this challenge, the proposed RL-based fusion network, based on Q-learning, is compared with other weight adjustment solutions. In one approach, weights are randomly selected in each iteration of the test, while in the other approach, all weights are set equal and the fused tracking result is considered the average value of all three tracking paths (i.e. RSSI, PDR, and AoA). The RL fusion

scenario consists of 1 million learning episodes and 1000 test episodes, and each test was repeated 50 times for all experiments.

As shown in Figure 4.11, for all four movement scenarios (i.e., Rectangular, Random, and two Diagonal), random weight adjustment and fixed averaged weights have higher variance, indicating their unreliability. In contrast, the proposed RL-IFF outperforms other approaches in terms of tracking the user and is a reliable approach to fuse the information gathered from multiple tracking scenarios. Additionally, the ability to produce stable performance across different episodes is another aspect of investigating the reliability of RL models, as shown in Figure 4.11. The proposed RL-IFF algorithm is more stable than its counterparts, with fewer sudden changes occurring during different episodes.

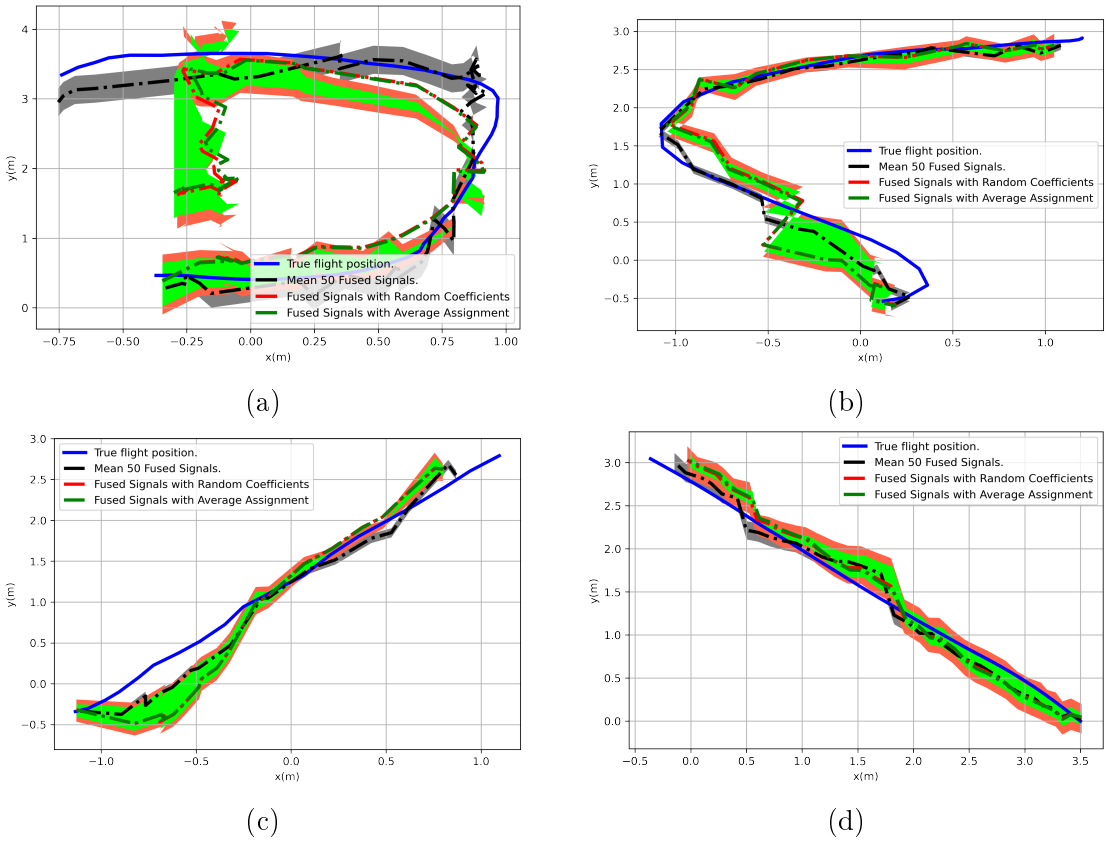


Figure 4.11: Evaluating the relaibility of the proposed RL-IFF comapred to random weight adjustement and averaged weight selections schems for four different sample movement scenarios: (a) Diagonal A; (b) Diagonal B; (c) Random, and (d) Rectangular trajectories.

4.6 Conclusions

In this chapter, we aim to tackle the challenges associated with MARL by introducing new solutions and utilizing the RL-based information fusion strategy to enhance the accuracy of indoor localization. This is achieved by combining the benefits of the different indoor tracking/localization approaches discussed in Chapter 3. To begin, we formulate the problem in Section 4.1. Next, we propose the MAK-TD framework in Section 4.2 and its SR-based variant, the MAK-SR framework in Section 4.3 as efficient alternatives to DNN-based MARL solutions.

The proposed frameworks address the limitations of DNN-based MARL techniques, such as sample inefficiency, memory problems, and the lack of prior information, by integrating Kalman temporal difference, multiple-model adaptive estimation, and successor representation for MARL problems. This integration accommodates changes in the reward model and overcomes the issues related to overfitting and high sensitivity to parameter selection. Furthermore, an active learning mechanism is implemented to balance exploration and exploitation, by utilizing the obtained uncertainty of the value function.

In Section 4.4, we introduce the RL-based information fusion framework, RL-IFF, to provide an efficient information fusion strategy that combines the AoA, PDR, and RSSI approaches to improve the accuracy of the tracking scenarios. This is achieved by leveraging the IoT-TD dataset, which provides reliable data for indoor tracking tasks. The novelty of the proposed RL-IFF framework lies in its integration of RL-based optimization with indoor tracking tasks, allowing us to take advantage of the benefits of different localization approaches to boost the accuracy of the final tracking scenarios.

Chapter 5

Blockchain-Enabled System for Indoor Contact Tracing in Epidemic Control

In Chapter 4, we delved into SR and SR-based MARL networks and proposed a stable framework for MARL. In alignment with the final objective of the thesis, Chapter 5 integrates our findings from the previous chapters into developing a trustworthy SCT framework. We introduce a novel framework, the TB-ICT, designed with the central aim of ensuring trustworthiness and security in the domain of Smart Contact Tracing. This robust framework, crafted with privacy as a priority, not only integrates the indoor localization solutions discussed in earlier chapters but also incorporates blockchain technology, reinforcing the security aspect.

The TB-ICT framework's design aims at preserving privacy and ensuring the integrity of the underlying CT data from unauthorized access. In section 5.1, we provide a comprehensive overview of the TB-ICT framework's architecture. Section 5.2 focuses on the data-driven localization model of the TB-ICT framework, which is based on BLE sensor measurements.

Finally, in Section 5.3, we detail the TB-ICT blockchain platform. This fully decentralized and innovative platform employs a dynamic Proof of Work (dPoW) credit-based consensus algorithm, coupled with a Randomized Hash Window (W-Hash) and dynamic Proof of Credit (dPoC) mechanisms. Lastly, in Section 5.4, we analyze the

security, complexity, and scalability of the TB-ICT framework, highlighting its effectiveness in managing and controlling the spread of contagious diseases in indoor environments. In summary, this chapter emphasizes the need for a secure, trustworthy, and privacy-oriented SCT framework, thereby answering the central question posed in our thesis.

5.1 The TB-ICT Framework

In this section, first, we present the 6 sub-systems designed to converge BLE-based CT with a blockchain platform. Afterwards, to facilitate practical implementation of the TB-ICT, assumptions used to develop the framework together with the underlying requirements for deployment are discussed. Finally, technical specifications utilized for the development of the proposed TB-ICT are presented.

5.1.1 TB-ICT Model and Convergence Analysis

The structure of the TB-ICT framework, as shown in Fig. 5.2, consists of the following two main components: (i) Indoor localization platform, and; (ii) The Blockchain network. To integrate blockchain within the proposed TB-ICT framework, we need an address generation subsystem that can represent the temporal identity of the underlying nodes. These addresses are used in the designed blockchain platform to send transactions as the inputs of the network that can be mined and added to the blocks. The blockchain network benefits from stable nodes to mine the blocks and this is how the user's contact data can be added to the blocks in the case of an infection. For the miners to be able to approve an infection claim, a data pool is used to query the claim made in a transaction sent to the blockchain. More specifically, to converge the blockchain platform, the following subsystems are designed and added to the TB-ICT framework:

• ***Network's Members***: Nodes that have the application on their device can join the platform and be a part of the network. There is a stable sub-network consisting of long-running nodes in the environment, which act as the core network, and are referred to as the "seed nodes". A new unstable node is not forced to connect to one of these stable nodes, however, it can utilize them for fast discovery of the other nodes in the designed network. This can speed up joining procedure for the light nodes in

the network allowing them to send transactions in the platform.

• ***Credit-based Platform***: To quantify the behavior of the nodes in the network, a credit-based platform is defined to calculate credit of the nodes. This approach enables the blockchain system to prioritize the high credit nodes to have a lower difficulty level to mine the blocks in the platform. The dPoW solution is used to integrate this credit-based mining eligibility approach in the platform. This is how real-world behavior of the nodes, e.g., practicing healthy distancing solutions can directly be integrated into the blockchain network.

• ***Unique Temporary Identity Generation Scheme***: A signature generation platform is designed to create temporary private/public keys and the related addresses to send a transaction in the blockchain network between entities. These signatures are generated by leveraging the ambient environmental data.

• ***Trace Transaction Scheme***: Each user's contact information (its proximate distance with other users within past 14 days) is kept on that user's handheld device. In case of infection, transaction containing the infection notification will be sent to its contacts' addresses stored in the user's device. This is how the stored contact tracing data will be sent as a transaction.

• ***Blockchain Transaction Submission Phase***: When a user known by her/his address is approved as an infected individual, the related private key of the user will be used to sign the transaction containing the contact's data (generating digital signature). Trace transactions will be issued by the infected user to all her/his close contacts' addresses and each of these transactions will be signed by related private keys. The public key of the user verified as an infected person by the authorized testing center is used by the miners to verify the digital signature and consequently the transaction.

• ***Mining, and Verification Phase***: Stable nodes and high-credit nodes can mine the blocks after verifying the claims in the transactions considering the data submitted in the Infected Users Pool (IUP) database. Each transaction, where the claim of the infection is approved will be verified as a true claim and will be added to the blockchain. Fig. 5.1 shows the flowchart of the TB-ICT. As can be seen, to provide the CT service, the BLE enabled devices running the TB-ICT application perform different phases, which are detailed later on.

Figure 5.1: Flaw chart of the proposed TB-ICT solution.

5.1.2 Assumptions

To satisfy functionality and privacy features necessary for implementation of a practical CT solution, the TB-ICT is implemented based on the following assumptions/requirements:

- *Transmission of Data*: A user device running the application can send non-connectable advertisement data to the nearby devices.
- *Data Reception*: A user device running the TB-ICT application frequently scans for nearby devices. In other words, the device running the TB-ICT application

is enabled to receive non-connectable advertisement data from nearby devices that run the TB-ICT application.

- *Active BLE Nodes*: While a user is running the application, process of scanning nearby BLE devices is active.

- *BLE Randomization*: Data transmission via the TB-ICT has no interference with the BLE randomization scheme.

- *Information Content*: The only information being exchanged between nearby devices is the unique signature, i.e., time-varying temporary address generated via the public/private key generation component of the blockchain module. In other words, no location data is exchanged through BLE advertisement packets as such there is no possibility for another device or an attacker to exploit the transferred data.

In addition to aforementioned assumptions/requirements, we considered following conditions as well:

- The proposed TB-ICT platform is designed to be used in limited indoor locations and it is assumed that the user in the indoor venue will not be surrounded by a large number of users in a very crowded situation. This will guarantee the ability of the application to trace the contacts in a real-time fashion.

- It is assumed that the expiration time based on the disease spreading time window is 14 days. In other words, each user's contact information, i.e., its proximate distance with other users within 14 days is kept locally. We also assume that the *immediate* distance for the proximity between users is less than 1-meter, and *near* is for distances between 1 to 2 meters. These proximity distances are assumed to be close contact of a user. Only the information about the users in the close contact will be saved in the users' smart phone.

- It is assumed that an enough percentage of the population under control in the indoor environment enroll in the automated contact tracing for outbreak control.

- To be realistic and to have robust localization, we assume a complex indoor environment without presence of LoS links. The AoA generated signals are affected by AWGN with different SNRs.

- Although there is no study in this work to model the heterogeneous level of smartphone/application usage, e.g., age or level of income, it is assumed that this application will not be used by users under or over a specific age range (under 10 or above 80 years old).

5.1.3 Design Objectives

Finally, technical specifications utilized for development of the proposed CT solution are as follows:

- Users' TB-ICT application uses the BLE non-connectable advertisement mode to disclosing sensitive data.

- In the non-connectable advertising mode, three advertisement channels will be used by a BLE node in a periodic fashion defined by the system's advertisement interval (T_a) to transmit the advertisement packets. Term T_a is the broadcasting frequency of the advertisement's packet, which has a memory space of 47 bytes, but only 31 bytes of this space can be used as a payload to host data and transfer it to another device.

- The BLE-based application advertisement task is configured to happen in Advertise Mode Low Latency.

- The only data that will be generated and transmitted in the TB-ICT platform is the unique address (unique temporary node's address) of close contacts. Consequently, this signature can be defined using 31 bytes or less of the available memory space.

- In the localization platform, we consider an indoor environment without presence of Line of Sight (LoS) links, modelled by Rayleigh fading channel, affected by Additive White Gaussian Noise (AWGN) with different Signal to Noise Ratios (SNRs).

- The smartphone will only stay active during the defined scanning window T_w. Generation interval T_g, advertising interval T_a, and scanning interval T_s are triggered periodically to exchange the data packets. When $T_a < T_s$, the possibility to see the advertising packet sent by neighboring smartphones is high.
- While the advertisement (T_a), scanning intervals (T_s), and scanning window (T_w) can be configured, it is recommended to keep T_a between 100 to 200 milliseconds to advertise about 10 to 5 packets per second.
- The period of address rotation for the advertiser is designed to be a random value between 5 to 10 minutes.

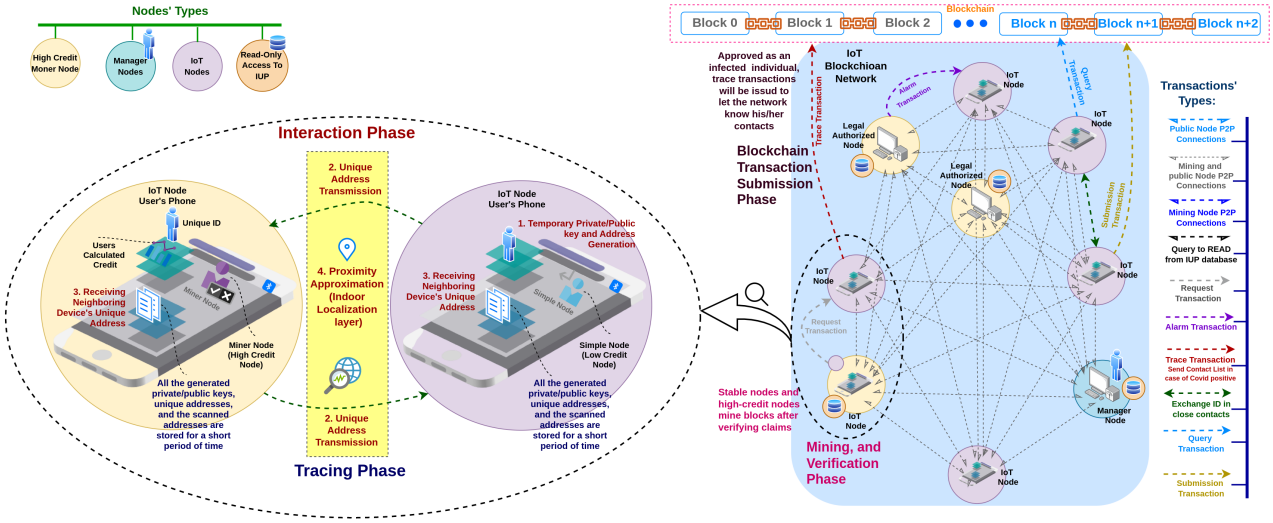


Figure 5.2: TB-ICT Blockchain network and Indoor Localization Platform.

5.2 The TB-ICT Localization Model

In this section, we focus on the localization phase, where as can be seen from Fig. 5.2, the IoT nodes (the BLE-enabled mobile devices) exchange their unique ID with other nodes within their local neighborhood. To have an efficient indoor CT model, it is essential to implement a robust localization framework with high accuracy. Next, we briefly outline targeted challenges in indoor localization:

• One of the most important challenges we face in indoor environments is the effect of obstacles, such as walls and humans, on the transmitted BLE signals. More precisely,

the transmitted signal can be reflected, refracted, and diffracted, and a number of phase delayed and power attenuated versions of the same signal is received by the mobile user. It is, therefore, of significant importance to consider a wireless signal model to be adopted with a dense indoor environment full of obstacles. Despite the existing research works [57–60], which can only cope with noise, we consider a highly dense indoor environment without presence of LoS links, modelled by Rayleigh fading, affected by AWGN with different SNRs.

• Another important issue associated with real indoor environments is the interference between mobile devices. Although the existing research works [57] assumed a single-user scenario, our wireless signal model is more realistic, which can be used to accommodate multi-user indoor environments.

• Considering a more complex experimental testbed with noise, interference, shadowing, and small scale fading, potentially leads to large localization error, which cannot be completely mitigated by existing CSI-based models [57, 60]. Furthermore, using complex signal processing approaches for modelling the multi-path and path-loss effects in indoor environments is not time efficient, especially for real-time applications. Therefore, by using DNN models, and generating a dataset adopted with a dense indoor environment full of obstacles, the effects of multi-path and path-loss can be learned based on the train dataset. Consequently, there is no need for complex and precise analytical models, therefore, we introduced a CNN-based approach, where the dataset is generated in the presence of noise and Rayleigh fading channel.

• Finally, in real indoor environments, mobile devices and BLE beacons are not always located along the same line, which in turn leads to issues related to elevation angle. Although the elevation angle of the incident signal is not utilized for the location estimation, its destructive effect on the location accuracy should be considered. By considering a 3-D indoor environment as the experimental testbed, therefore, the effects of elevation angle is considered on the train dataset.

5.2.1 CNN-based AoA Localization Framework

In this subsection, first we formulate BLE wireless signal model. Then, we present our proposed CNN-based AoA localization framework relying on BLE technology. Fig. 5.3 illustrates different components of the proposed localization platform, i.e., (1) BLE transmitter; (2) BLE Wireless channel; (3) BLE receiver; (4) AoA estimation; (5)

Data Preparation, and; (6) CNN-based localization, which are introduced below:

BLE Transmitter

The baseband version of the transmitted signal $s_b(t)$ is calculated as [47]

$$\begin{aligned} s_b(t) &= s_i^b(t) + j s_q^b(t) \\ &= \sqrt{\frac{2E}{T}} \Big\{ \cos(\phi(t) + \phi_0) + j \sin(\phi(t) + \phi_0) \Big\}. \end{aligned} \tag{85}$$

where terms E and T represent the energy and the time interval of the transmitted symbol, respectively. Term ϕ_0 is the initial phase of the transmitted signal. Finally, term $\phi(t)$, denoting the phase deviation, is expressed as [47]

$$\phi(t) = \frac{\pi h}{T} \int_{-\infty}^{t} \sum_{n=-\infty}^{+\infty} s[n] p(\tau - nT) d\tau, \tag{86}$$

where h, known as the modulation index, is between 0.45 to 0.55 in BLE standard. Terms $s[n] = \pm 1$ and $p(t)$ denote the baseband pulse sequence and Gaussian Filter (GF), respectively. Finally, the transmitted BLE signal $s(t) = Re\{s_b(t) e^{j2\pi f_c t}\}$, is modulated by Gaussian Frequency-Shift Keying (GFSK) [47], i.e.,

$$s(t) = \sqrt{\frac{2E}{T}} \cos\left(2\pi f_c t + \phi(t) + \phi_0\right), \tag{87}$$

where $2.4 \leq f_c \leq 2.48$ GHz denotes the carrier frequency.

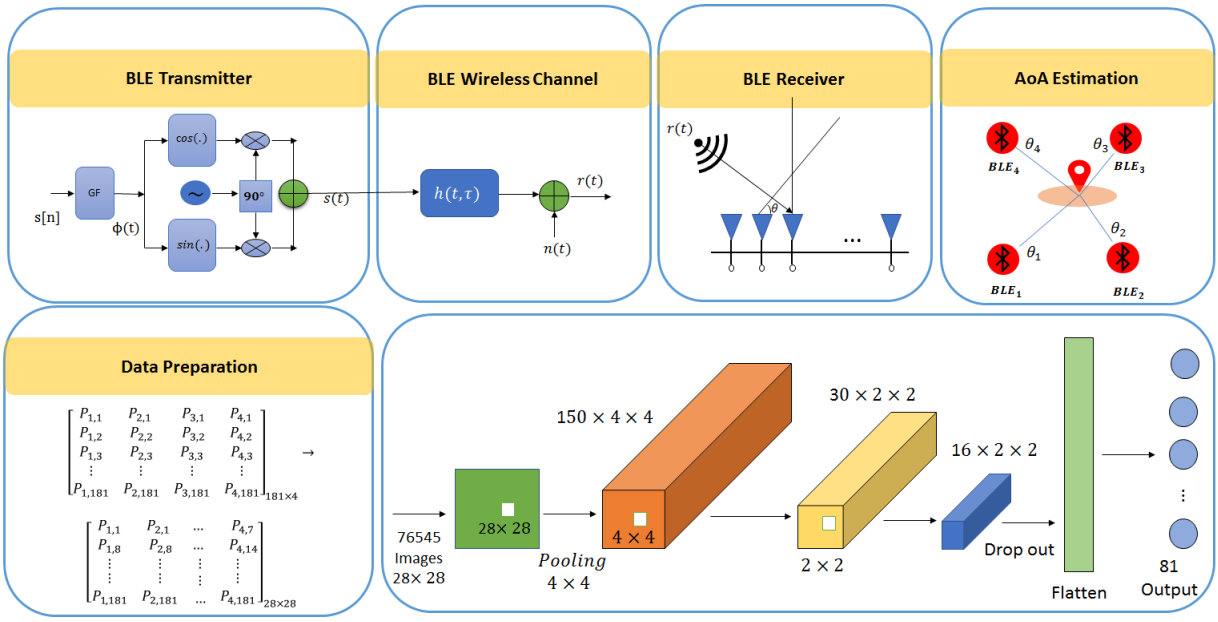


Figure 5.3: Block diagram of the BLE transceiver, wireless channel model, and the proposed CNN-based AoA localization framework.

BLE Wireless Channel

Because of the obstacles in indoor environments, a number of phase delayed and power attenuated versions of the transmitted BLE signal, which are affected by AWGN, are received by the BLE beacons. Therefore, the received BLE signal can be expressed as [47]

$$r(t) = \sum_{k=1}^{N(t)} \rho_k(t,\tau) s(t - \tau_k(t)) + n(t), \tag{88}$$

where $N(t)$ represents the number of detachable paths. Terms $\rho_k(t,\tau)$ and $\tau_k(t)$ denote the attenuation and the delay of the k^{th} path, respectively. In addition, term $n(t)$ represents the AWGN channel, which is modelled by $n(t) \sim \mathcal{N}(0,\sigma^2)$. In a BLE positioning system, the location of users can be obtained from the estimated CIR, [47], i.e.,

$$h(t,\tau) = \sum_{k=1}^{N(t)} \rho_k(t,\tau)\delta(t - \tau_k(t)). \tag{89}$$

BLE Receiver

To extract the angle of the incident signal, BLE beacons are required to be equipped with an antenna array, where commonly the Linear Antenna Array (LAA) is used. In a typical LAA, there are N_e number of elements with the same distance d_0, where the transmitted BLE signal is received by distinct elements with different phase differences. The discrete received signal by element e, sampled at time slot m, denoted by $r_e[m]$, is obtained as [195]

$$r_e[m] = s^{'}[m]\Theta(\theta, \phi)[m] + n[m], \tag{90}$$

where $\mathbf{\Theta}(\theta, \phi)$ denotes the array vector, expressed as [195]

$$\begin{aligned}\mathbf{\Theta}(\theta, \phi) = \exp([-j\frac{2\pi d}{\lambda}\cos\theta\cos\phi, \\ -j\frac{2\pi d}{\lambda}\sin\theta\cos\phi, -j\frac{2\pi d}{\lambda}\sin\phi]^T),\end{aligned} \tag{91}$$

where θ and ϕ represent the azimuth and elevation angles, respectively. Term d indicates the space between two consecutive elements of the LAA, which is equal to $\frac{\lambda}{2}$, with $\lambda = c/f_c$. Finally, term $c = 3 \times 10^8\ m/s$ is the speed of light. By assuming M samples in each received signal, we have [195]

$$\mathbf{r} = [r_1[m] \dots r_{N_e}[m]]^T, \tag{92}$$

$$\text{and} \quad \mathbf{s}^{'} = [s_1^{'}[m] \dots s_{N_e}^{'}[m]]^T. \tag{93}$$

Therefore, the received signal can be expressed as follows

$$\mathbf{r} = \mathbf{\Theta}(\theta, \phi)\mathbf{s}^{'} + \mathbf{n}. \tag{94}$$

AoA Estimation

Given the angle of the incident signal through a subspace-based angle estimation algorithm, we propose to use a CNN model for localization. In this regard, the spatial spectrum of the received signal, denoted by $\mathbf{p}(\theta, \phi, t)$, is calculated in each

time slot as follows

$$\mathbf{p}(\theta, \phi, t) = \frac{1}{\mathbf{\Theta}^H(\theta, \phi, t)\mathbf{E}_N\mathbf{E}_N^H\mathbf{\Theta}(\theta, \phi, t)}, \tag{95}$$

where $\mathbf{E}_N$ is the noise eigenvectors of the covariance matrix $\mathbf{R} = E[\mathbf{r}, \mathbf{r}^H]$, and $\mathbf{r}$ denotes the received signal.

Data Preparation

In each time slot, a reshaped angle image is the input of the CNN, which is generated by $\mathbf{P}(\theta, \phi, t) = [\mathbf{p}_1(\theta, \phi, t), \ldots, \mathbf{p}_{N_b}(\theta, \phi, t)]$, where N_b indicates the number of available BLEs in the user's vicinity. Since $0 \leq \theta \leq 180$, there are 181 samples in $\mathbf{p}_i(\theta, \phi, t)$, where the minimum value of $\mathbf{p}_i(\theta, \phi, t)$ indicates the incident angle θ. Fig. 5.3(b) illustrates a typical angle image, containing 724 sets of AoA measurements generated by the subspace-based algorithm. As can be seen in Fig. 5.3(b), the original angle image is of size 4×181, reshaped to be an square angle image of size 28×28 by zero padding. This completes presentation of the localization platform of the TB-ICT. Next, we focus on the blockchain module.

5.3 The TB-ICT Blockchain Platform

The TB-ICT framework is a blockchain-based infrastructure where the ledger holding all the transactions, is a public open ledger for those having access to the application itself. Once a user joins the network, in the sign up phase, she/he creates a blockchain account and receives public and private keys (P_k, S_k) within the initialization of the account as the node's unique identifier. In this framework, private key is changed over time based on the signature generated by the application, which is constantly searching for other close BLE devices to monitor ambient environmental features. The time-variant key generation scheme is comprehensively discussed in Sub-section V-A. The generated key is used to sign the transactions and create unique public identifiers to be placed in the BLE advertisement packets and to be shared with neighboring nodes. The initial block download process, which is the process of downloading blocks that are new to a user, is much easier in the TB-ICT platform. This is due to the nature of the network and its specific use case as only the last 14 days'

information/transactions are necessary to be stored. In the TB-ICT framework, there is a stable core consisting of long-running nodes in the environment. Based on the indoor venue, these full nodes (contain an entire copy of the ledger) are those in charge of that specific indoor location and also technical or non-technical service providers of that location. Health and public authorities, health centers, and hospitals who have access to the COVID-19 test results are the default stable nodes, which can also participate as miners in the network. For example, one of the indoor locations that the proposed TB-ICT platform is designed to be used is universities. The technical IT solution providers of the university, the authorities in charge of the management of the network, the health center nodes, and the technical support center of the university are the stable long-running nodes in the network. These stable nodes which are part of the network since the beginning of the platform and the creation of the first blocks are called "seed nodes". A new unstable node (light node that does not hold full copies of the ledger) is not forced to connect to one of these stable nodes, however, it can utilize them for fast discovery of other nodes in the designed network. No matter how many nodes join the network or leave the platform, how many of them are active, or acting in an offline manner, the whole network will be always up and running based on the seed nodes. Generally speaking, in the TB-ICT framework, nodes are divided into manager static nodes, authorized static nodes, and dynamic nodes. In what follows, each of these entities, their role, and their level of credibility is discussed in detail:

• ***Manager Static Nodes:*** These nodes are the stable nodes that started the network and have the responsibility to manage other static nodes and maintain stability of the whole network. Manager nodes are specific full nodes that can approve other authorized nodes to be added or deleted from the network. For example, in a university setting, the genesis block can be generated by the school's managers as the main initial nodes. Other authorized manager sections of the university and the healthcare-related nodes can be added based on the approval of these nodes and the credential which is given to them. These nodes have high credits and always play the role of full nodes in the network. The private key related to these nodes and their derived public keys are hard-coded into the software installed by these nodes and eligible authorized static nodes. Each manager node can keep track of all authorized nodes in the network by signing a transaction TX of available authorized static nodes' public

keys $\left(PK_{AN_1}, PK_{AN_2}, \ldots, PK_{AN_n}\right)$ with its secret key. Since the manager uses his secret key, which cannot be forged to sign the transaction, other authorized nodes can simply fetch the list of the authorized nodes in the blockchain network published by the manager.

• ***Authorized Static Nodes:*** These full nodes are legally authorized units that have read access to data in the Infected Users Pool (IUP). These nodes have an initial credit higher than the network threshold and can act as the miners in the network, keeping the whole platform stable. In the case of any fraudulent activity, they will receive a high negative score and will lose their credibility to mine in the network. The infected persons' data, which is saved in the IUP, is distributed among these nodes and cannot be changed. When an authorized node receives data about the infection of a node, this data will be broadcasted to the other authorized nodes. These nodes will validate any claim of infection, and in case of any false claim, the claiming node will be punished by receiving negative credit points. All the transactions in the network will be validated and picked to be mined and added to a block via stable nodes, i.e., manager static nodes, authorized static nodes, and high credit dynamic nodes, which are discussed next.

• ***Dynamic Nodes:*** These nodes, as power-constrained BLE-enabled smartphones, are public users of the application. Dynamic nodes are the basic entities that are used to apply the AoA-based proximity approximation scheme designed in the TB-ICT localization platform. These nodes can send and receive different transactions in the network, including but not limited to Submission Transaction (ST), when starting as node in the network; Trace Transaction (TT), when claiming the infection; Query Transaction (QT), when checking their status of infection to see whether or not they were in close contact with an approved infected node, and; Alarm Transaction (AT), when they receive an alarm mentioning their possible infection based on the data approved in the network. The level of the credibility of dynamic nodes varies based on their behavior in the system. Since these devices are power constraint nodes, they mainly act as light nodes in the blockchain. If their credit score is higher than the network-defined credit threshold, similar to the Authorized Static Nodes, they can play the role of high credit miner nodes. In different phases of the TB-ICT, dynamic nodes can perform the following actions:

• *Interaction Phase*:

- *Temporary Private/Public key and address Generation*: To generate private/public keys, a signature is created in different time intervals exploiting the environmental features received through the advertisement packets of other BLE-enabled devices in the environment. BLE-enabled mobile devices scan for these advertisement packets in different time intervals and also propagate their internally created unique addresses via their non-connectable advertisement packets. These data packets change over time based on the user's location or neighboring BLE devices.

- *Unique Address Transmission*: A mobile device in the network periodically broadcasts the advertising packet containing its own unique address via non-connectable advertising channels. These broadcasting events happen in the advertising intervals (T_a).

- *Receiving Neighboring Device's Unique Address*: A periodic scanning process also happens to leverage the three advertising channels to scan for the other neighboring devices' advertisement packets. These scanning events happen between broadcasting intervals.

- *Proximity Approximation*: While a BLE packet of another device is received, the proximity to that user will be estimated via the proposed AoA-based localization scheme. If the users are in close contact, the received unique address will be stored on the local database.

• *Tracing Phase*: All the generated private/public keys and unique addresses, and the scanned addresses are stored for only a short period of time in the local device's database. As the stored data does not contain any personal information about the owner, it is impossible to find or trace anything about users and their personal information or traces and locations in the system. These data will also be deleted from the devices after the short expiration time, which can remove any historical data and keep the local storage required light and applicable.

• *Blockchain Transaction Submission Phase*: When a user known by her/his address is approved as an infected individual, the related private key of the user will be used to sign the transaction containing the contacts' data (generating a digital signature). The infected user will issue trace transactions to all the close contacts' addresses, and each of these transactions will be signed by related private keys. The public

key of the user verified as an infected person is used by the miners, i.e., high credit and authorized static nodes, to verify the digital signature and consequently the transaction.

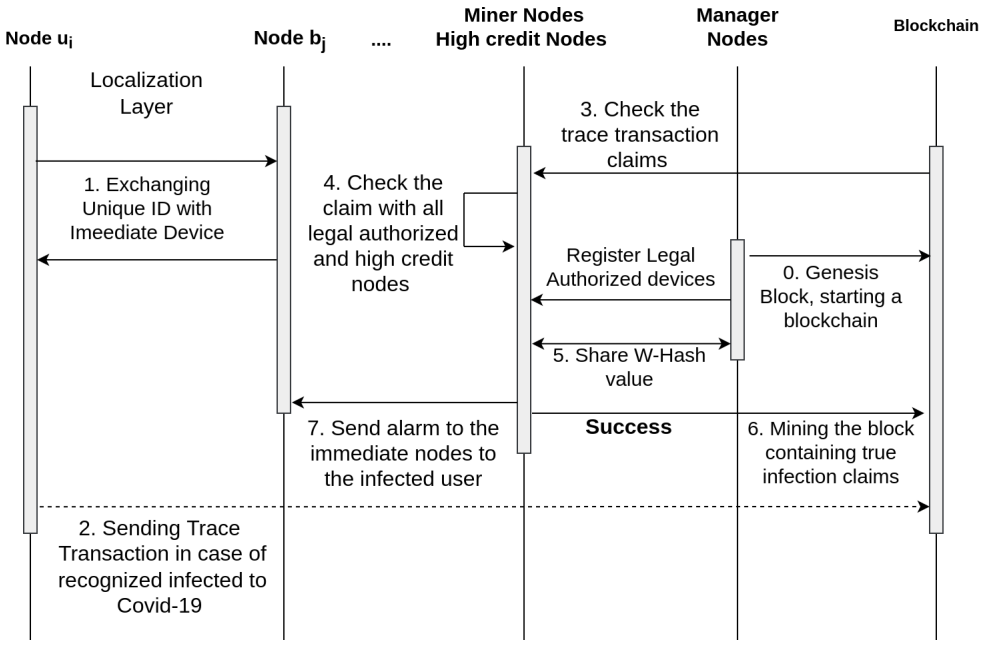


Figure 5.4: Interconnections within the TB-ICT network.

Other nodes can submit different transactions about the status of the network (e.g., QT and ST) and/or claim an infectious status as a transaction (i.e., the TT). Interconnections between different nodes are shown in Fig. 5.4. The proposed TB-ICT as an alternative blockchain platform exploits a time variant signature generation Scheme, a Dynamic PoW (dPoW) Credit-based consensus algorithm along with a Randomized Hash Window (W-Hash), and Dynamic Proof of Credit (dPoC).

5.3.1 Time Variant Signature Generation Scheme

Generally speaking, the ICT platform works based on BLE-based data observed by a smartphones in the ambient environment. Typically, encryption approaches use the raw data provided by the users, which can lead to data leakage if the encryption approach is compromised. In the TB-ICT platform, a privacy-preserving signature generating scheme is designed to address this issue and improve privacy. This signature generation scheme will then be used to generate temporary private/public keys

and the related addresses to send a transaction in the blockchain network between entities. While a user's device is filtering the neighboring BLE devices, batches of data are obtained from ever-changing environmental proximity sensing information, which is exploited to create temporary signatures (discussed in details in the following subsection). These generated signatures in different time intervals will then be used to create unique addresses for secure exchange of BLE packets of the close contacts without revealing sensitive personal data. In other words, it will be used in the procedure of creating different temporary private keys, public keys and addresses to send the transaction between nodes. The public keys will finally be used to sign the blockchain transactions in case of infection with COVID-19 while keeping the data secure, immutable, anonymous, and distributed. Next, the type of communication in the TB-ICT and its privacy preserving scheme are described:

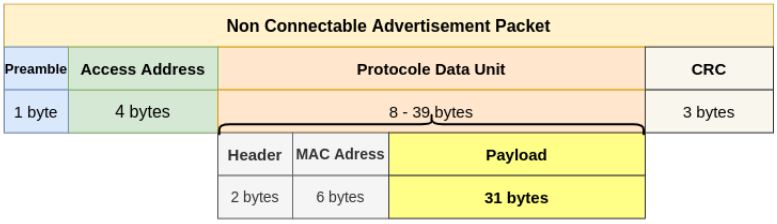


Figure 5.5: Size of the advertising packet (47 bytes) however and available bytes 31 bytes for the actual payload.

Type of Communication in TB-ICT

In the TB-ICT framework, Bluetooth Low Energy (BLE) is used as the communication technology. In brief, BLE is developed with reduced power consumption for short-range wireless communication over 2.4GHz industrial, scientific, and medical (ISM) frequency band [13]. Almost all smartphones are now equipped with this technology, making BLE an ideal communication medium for providing different IoT-based services such as indoor localization and CT. BLE spectrum has 40 frequency channels with 2 MHz channel spacing, 37 $(0-36)$ of which are for data transmission, and 3 $(37-39)$ are considered the advertising channels where frequency hopping is employed to diminish the interference effects. Communication-wise, BLE introduces two communication modes: (i) Non-connectable advertising and; (ii) Connectable advertising [196]. In the non-connectable mode, no connection request is accepted by

other BLE devices, which makes packet transfer safer and more secure. In this work, therefore, the non-connectable mode is used as the communication system. To adopt the non-connectable communication mode in the proposed TB-ICT framework, the BLE-enabled modules are configured to periodically broadcast advertisement packets via BLE 3 advertising channels. Other BLE devices, which are scanning periodically, can receive the transmitted packets while they are active within the broadcast range. While in non-connectable advertisement mode, it is not required for the nodes to fully connect to each other by accepting a BLE connection request, there are still security concerns. One major issue is limited payload capacity to encapsulate data for transmission to other devices, i.e., a data packet is limited to 47 bytes, as shown in Fig. 5.5, with only 31 bytes being available in the payload to add and transfer user's temporary address. In the proposed platform, unique temporary addresses with the size of 20 bytes will be encapsulated in the payload of the non-connectable adverting signals and transferred through the non-connectable advertising channels. While continuous scanning can increase the rate of receiving packets, this approach has a negative impact on energy consumption. Therefore, we consider the scenario where an activated smartphone stays active only during a predefined scanning window T_s and can only receive advertisement packets which have overlap with its scanning interval. Generation interval T_g, advertising interval T_a, and scanning interval T_s are triggered periodically to exchange the data packets. When $T_a < T_s$, the possibility of seeing the advertising packet sent by neighboring smartphones is considerably high [13]. While the scanning window is long enough, with a higher broadcasting frequency than the scanning frequency, it is more likely that packets from a device reach the scanning window of another one.

TB-ICT Privacy Preserving Scheme

As discussed earlier, in the TB-ICT application (implemented on users' devices) the BLE non-connectable advertisement mode is used to prevent any full connection to be in danger of revealing sensitive data. The only data that will be transmitted in the TB-ICT platform is the unique address (unique temporary node's address), which is defined to consume less than 31 bytes of the memory space. It is worth nothing that, one node's address will be changing over time and in the blockchain adding to the privacy preserving characteristics of the proposed framework. The

aforementioned unique address is computed based on localized public keys, which are computed via private keys. The latter is formed based on localized signature. In what follows, we describe computation of signatures, private, and public keys. *(i) Signature Generation Step*: To create a signature for private key generation, time-averaged RSSI values received through the advertisement packets of other BLE devices (static/dynamic) in the environment is used. These data packets change over time based on the user's location or the BLE devices' location. When a node joins the network and the application starts discovering its close contacts, a 32 bytes signature will be generated. We interchangeably refer to a node's signature as the ambient environmental vector. Every few seconds, the signature will be changed based on the user's new location and relative distance to the surrounding sensible BLE devices. Let $B = \{b_1, b_2, \ldots, b_m\}$, be a set of BLE-enabled devices observed by a user's smartphone (excluding the user itself) at a specific time. The observed vector of the contacts (neighboring BLE devices) for the user u is,

$$\boldsymbol{o}_u(t) = (P_e(d_{b_1}), P_e(d_{b_2}), \ldots, P_e(d_{b_m}))^T \tag{96}$$

where $\boldsymbol{o}_u(t) \in \mathbb{R}^m$ and length of vector m is dependent on the size of $B = \{b_1, b_2, \ldots, b_m\}$. Term $P_r(d)$ is assumed to be a function returning the time-averaged RSSI values from a BLE device located at a certain distance from the smartphone. Dictionary $\Psi \in \mathbb{R}^{32\times m}$ as a known secret transformation key is considered to transform the m-dimensional vector to a 32-dimensional vector, i.e.,

$$\Psi = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \ldots & \psi_{1,m} \\ \psi_{2,1} & \psi_{2,2} & \ldots & \psi_{2,m} \\ \vdots & \vdots & \ldots & \vdots \\ \psi_{32,1} & \psi_{32,2} & \ldots & \psi_{32,m} \end{pmatrix} \tag{97}$$

Unique signature vector $s(t) \in \mathbb{R}^{32}$, is calculated by multiplying this dictionary by the j^{th} observer vector as follows

$$s(t) = \sum_{b_j \in B} \boldsymbol{\Psi}_j P_e(d_{b_j}), \tag{98}$$

where $\mathbf{\Psi}_j = (\psi_{1,j}, \psi_{2,j}, \ldots, \psi_{31,j})^T \in \mathbb{R}^{32}$ is the column vector from the dictionary. The generated signature will be checked and used in order to create a private key. One major step for creating the private key is finalized.

(ii) Private Key Generation Step: A random combination of the 32 bytes generated unique signature will be chosen to create the private key. After this step, there are two other main steps to generate the related public key and address, i.e., (i) build a 64 bytes public key from the generated private key, and; (ii) Drive the 20 bytes unique address from the public key.

(iii) Public Key Generation Step: Elliptic Curve Digital Signature Algorithm (ECDSA) is used to generate the public key from the private key.

(iv) Address Generation Step: As the address of a node in a blockchain network (Ethereum Key generation scheme [197]) is 20 bytes, this address can be exchanged via non-connectable BLE packets. Consequently, the Keccak-256 hash of the public key is calculated to generate a 32 bytes string. The last 20 bytes of this string will be taken and considered as the address of this entity as follows

$$A(t) = \beta_{96\ldots255}(KEC(ECDSAPUBKEY(rnd(s(t))))), \tag{99}$$

where rnd is the function that generates a 32 bytes random combination of the generated $s(t)$, and $ECDSAPUBKEY$ is the function deriving a public key from the private key (i.e., $rnd(s(t))$) and KEC is Keccak-256 hash of the public key which will return a 32 bytes string. $\beta_{96\ldots255}$ will extract the last 160 bits of the generated string as the address of the node. Considering the positive infection situation, a user wants to send a trace transaction to the network. When a user known by her/his address is approved as an infected individual, the related private key of the user will be used to sign the transaction containing the contacts' data (generating a digital signature). The addresses are stored for only a short period, and the close contacts related to each of the temporary generated addresses are saved in the local device database. The infected user will issue trace transactions to all the close contacts' addresses, and each of these transactions will be signed by related private keys. The public key of the user verified as an infected person by the authorized testing center is used by the miners to verify the digital signature and, consequently, the transaction. Fig. 5.6 shows the procedure of creating different temporary public keys and validating the digital signatures. This approach will enable the user to keep his/her identity secure and submit

various transactions using a non-constant private key. This non-stationary signature generation scheme allows the platform to be secure against different threats in the networks, including key attacks, replay attacks, impersonation attacks, modification, and man in the middle attacks [198].

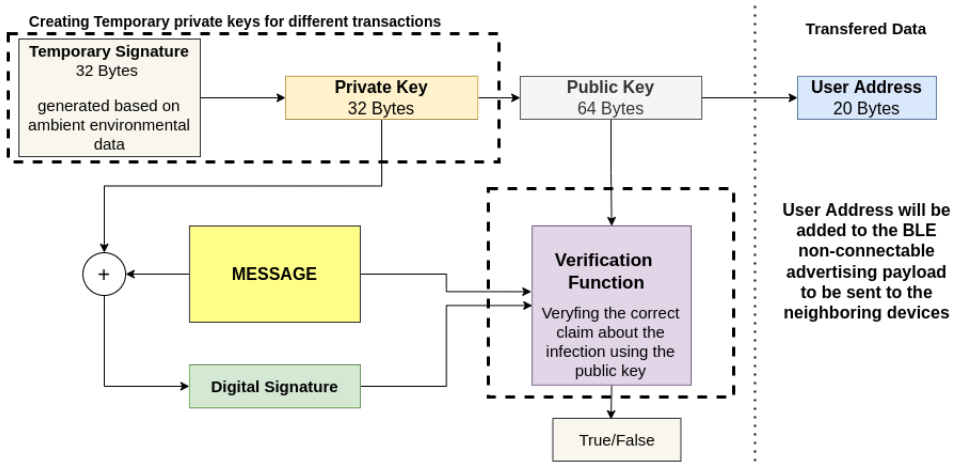


Figure 5.6: Signature and private/public key and address generation and transaction verification scheme.

5.3.2 Randomized Hash Window

The introduced Randomized Hash Window, referred to as the W-Hash, is a mechanism to mine a block, which is similar in nature to the randomized Sliding Window Algorithm (SWA) and the check-points concept [76]. The SWA is inherently an approach to simplify processing of the incoming sequence of data. More specifically, when dealing with large time-stamped data streams, previously received sequences can be outdated after a certain time and loose their value to be added to the underlying analysis procedures. In different use-cases, the SWA considers a fixed (predefined) window (n) of the received data to apply the analysis and discard the older data sequences. By reducing the data window, focus is given to more recent data as such considerably lowers the data space [199]. In its randomized version [199], a random value n of the sliding window will be generated at each epoch. While SWA, is used to reduce the previously used number of time-stamped sequences for simplification purposes, W-Hash intentionally increases the sequence of previously received data to make the platform more complex and boost the difficulty level to make the mining

procedure harder. More specifically, a miner continuously mines the current block with a number, i.e., the nonce value, to meet the predefined Difficulty Level (DL) and produces the target hash string. W-Hash makes this hashing process harder, i.e., the W-Hash mechanism slides through the blockchain ledger similar to the sliding window algorithm. In other words, as can be seen in Fig. 5.7, the miner should hash the current block along with the sliding window of the previous blocks. For instance, assume W-Hash is 14, and the current block number is 100, the miner should consider the concatenation of block 100 with previous 13 blocks (from 87 to 99) to create the window size of 14. Within the initializing phase of the TB-ICT network, when a W-Hash is larger than the current number of blocks in the blockchain, the W-Hash will be set to zero. Otherwise, the current block will be hashed based on the previous succeeding blocks determined and concatenated considering the W-Hash. In the long term, when more blocks become added to the blockchain, the W-Hash, which is a value between 0 to 100, can be applied to every block. The drawback of such an approach is that the overhead complexity is of $O(n)$, where n denotes the W-Hash value, when compared to the classical blockchain platforms with a constant PoW difficulty level. To address this issue, we propose dPoW difficulty managed by a credit-based evaluation mechanism as discussed next.

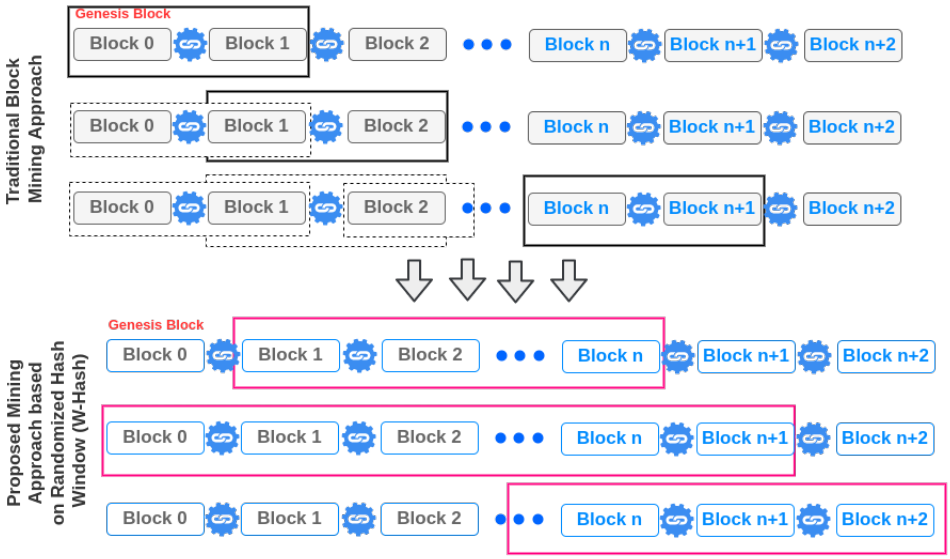


Figure 5.7: TB-ICT randomized hash window (W-Hash) solution to enhance the security of the proposed Blockchain network.

5.3.3 Dynamic Proof of Work (dPoW)

As stated previously, there is a trade-off between security and scalability in blockchain networks. The proposed W-Hash solution is an innovative approach to maximize the security of the network. However, W-Hash directly affects the underlying computational complexity, mining costs, and mining time. As stated in [76], a dynamic approach can be employed to define the different levels to control the rate of the incoming communication traffic and the mining difficulty. Therefore, Dynamic PoW consensus approach and credit-based difficulty level (DL) determination for the nodes in the network is a solution to the added complexity of the W-Hash. Intuitively speaking, the users in the network with higher credit can exploit a much easier difficulty level in the system helping them to easily mine the blocks, and the confidential W-Hash value n will be shared with them securely. On the other hand, a malicious miner now requires both $n-1$ blocks and the W-Hash value n to mine a new block. A fraud process and/or faulty behaviors are, therefore, costly as they result in high negative credits resulting in the higher difficulty.

In the TB-ICT framework, we consider two difficulty levels to be applied for the nodes in different circumstances. As can be seen in Table 5.1, Difficulty level 1 (DL_e) has the lowest difficulty achieved by considering the first significant digit of the block hash string, which is calculated by incrementing the nonce continuously and check the results until meeting the difficulty constraints. The first level of difficulty DL_e is assigned to the legally authorized nodes and the nodes in the network with a credit larger than a predefined threshold (α_d). Details of the credit evaluation process will be covered later in Section V-D. Due to the low level of difficulty, even devices with limited computational power (if they meet the constraint set by the credit threshold) can participate in the mining process. The second difficulty level (DL_h) is four times harder than DL_e, which is achieved by considering the first four significant digits of the target hash string. Level 4 is considered for the low incoming communication traffic and for dishonest nodes with low credits.

Table 5.1: Difficulty Level (DL) of the Proposed Dynamic PoW.

DL	**Size (bits)**	**Target Hash**	**Difficulty**
DL_e	4	SHA256[0:1]	Low
DL_h	16	SHA256[0:4]	High

Parameters of the dPoW: Similar in nature to Bitcoin, data encryption is implemented in the TB-ICT exploiting Elliptic Curve Cryptography (ECC) with Elliptic Curve Digital Signature Algorithm (ECDSA). The block size in the blockchain of Bitcoin is restricted to 1 megabyte. Unlike the conventional blockchain model [198], which has a fixed block size, TB-ICT is designed to have a variable block size considering incorporation of the dPoW approach in its consensus algorithm. Given the variable block size of TB-ICT, we set an upper bound of 1 megabyte for the block size. The variable block size is formulated as follows

$$\begin{aligned} S_{B_c} = &Overhead_{B_c} + \\ &[(Data + EncryptionOverhead) \times N_{data}], \end{aligned} \tag{100}$$

where S_{B_c} is the current block size, $Overhead_{B_c}$ is the block metadata (number of bytes representing a block), $EncryptionOverhead$ is the number of bytes for current block data encryption, and N_{data} is the number of data samples added to the current block. The difficulty level is

$$D_{B_c} = \frac{DL}{TargetHash}, \tag{101}$$

where D_{B_c} represents the difficulty level of the current block, DL is the level of the difficulty in the dPoW, i.e., DL_e or DL_h, and the $TargetHash$ is the required string to mine a block. The mining interval of the blocks, representing the average time required to mine a block, can be formulated as

$$Intervel_B = \frac{D_{B_c} \times 2^b}{HashRate}, \tag{102}$$

where $HashRate$ is a miner's computation power for hash iteration generation per second, and 2^b is responsible for difficulty level (DL) bits control, i.e., 16 bits for the DL_h and 4 bits for DL_e. Table 5.1, represents the difficulty level in the proposed dynamic structure.

5.3.4 Dynamic Proof of Credit (dPoC)

Consider a CT network with N_{nodes} number of nodes. Node i, for $(1 \leq i \leq N_{nodes})$, is assigned a credit $CR_i(t)$ at time instant $t > 0$, which is a parameter altering in

real-time based on the node's interactions and behavior in the environment. Normal Behaviors, i.e., following the physical distancing regulations and sending transactions after being recognized as an infected person, will gradually increase the user's credit value over time. On the other hand, abnormal behaviors will diminish the node's credit over time. Abnormal behaviors can be classified into two main categories:

- *Disease-Related Violations:* Breaching the physical distancing regulations and other related rules, which are mostly bond to the proximity-based localization results.

- *Anomalous Behavior in Blockchain Network:* Different anomaly behaviors in the networks including attacks or wrong claims, e.g., false claims of infection to the COVID-19 or failing to share contact information in the case of being diagnosed with COVID-19.

Node i, based on its behavior, can receive negative or positive scores resulting in the following credit

$$Cr_i(t) = Cr_i^{Prox}(t) + Cr_i^{-}(t), \tag{103}$$

where t is the current time index, $Cr_i^{Prox}(t)$ is the proximity-based credit (which can be positive or negative), and $Cr_i^{-}(t) \leq 0$ is the received negative credit score, which is calculated based on malicious behaviors of the node in the network. Following Section 5.2, let $\mathcal{O}_i(t) = \{o_1(t), o_2(t), \ldots, o_{N_i(t)}(t)\}$, denote the set of $N_i(t)$ BLE-enabled devices observed by the i^{th} user (excluding the user itself) at time t. Every few seconds, this set of devices can be changed as the user moves within the environment. Based on the distance between user i and a neighboring BLE-enabled device, Node $o^{(j)}$, $(1 \leq j \leq N_i(t))$, we consider two categories, i.e., immediate nodes (less than 2 meters) and non-immediate node (near or far, when the distance is more than 2 meters). We define $P_i(o^{(j)})$ as the credit assigned to Node i based on its proximity to the neighbouring Node $o^{(j)} \in \mathcal{O}_i(t)$ as follows

$$P_i(o^{(j)}(t)) = \begin{cases} \frac{-1*\lambda^-}{L(i,o^{(j)}(t))}, & \text{Immediate, } L(i,o^{(j)}(t)) < 2 \\ \frac{L(i,o^{(j)}(t))}{\lambda^+}, & \text{Near/Far, } L(i,o^{(j)}(t)) \geq 2 \end{cases}, \tag{104}$$

where $L(i, o^{(j)}(t))$ denotes the distance between Agent i and its contact $o^{(j)}(t)$ at time t. Terms λ^- and λ^+ are the credit calculation coefficients that are defined

based on the policy utilized to punish and reward the users. In a difficult situation when the disease is highly contagious and preventing policy should be considered more seriously, both λ^- and λ^+ can have a higher value. As can be seen, keeping a healthy distance will be rewarded by a positive credit. On the other hand, in case of immediate proximity, the node will be charged with a higher negative credit score proportional to its closeness to the other contact. These coefficients allow the model to manage credit weight distribution. Considering Eq. (104), the proximity credit assigned to Node i is calculated as follows

$$Cr_i^{Prox}(t) = \sum_{o_j(t) \in \mathcal{O}_i(t)} P_i(o^{(j)}(t)), \tag{105}$$

where $Cr_i^{Prox}(t)$ is adjusted based on the level of healthy preventive behavior practiced by Node i. To measure the total credit value of a node based on the localization result, this value will be added to the previously calculated proximity-based credit to calculate the final credit. Based on the scanning and advertising intervals of the BLE devices, every few seconds, the new credit value of the user will be updated and added to the previously calculated proximity-based credit.

Disease-related violations are incorporated into the credit score based on the proximity results obtained from the localization model. Term $Cr_i^-(t)$ is negatively proportional to the number of anomalous behaviors of Node i, to account for anomalous behaviors, i.e.,

$$Cr_i^-(t) = -\sum_{l=1}^{m_i} \omega . \frac{\Delta T}{t - t_l} \tag{106}$$

where m_i is the total number of abnormal behaviors performed by user i, t_l indicates time point of the l^{th} abnormal behavior, and ΔT represents a unit of time. Term $\omega(x)$ is the penalty coefficient and is computed as follows

$$\omega = \begin{cases} \omega_{FC}, & \text{If a false claim is made} \\ \omega_{SC}, & \text{If a contact violation happened} \\ \omega_{NA}, & \text{If a network attack happened} \end{cases}, \tag{107}$$

where all coefficients (ω_{FC}, ω_{SC} and ω_{NA}) can be adapted based on the network's punishment policy. A node's credit can be measured via Eq. (103) by calculating

$Cr_i^-(t)$ and $Cr_i^{Prox}(t)$. In the case of abnormal claims or malicious behavior of a node, the absolute value of $Cr_i^-(t)$ becomes a large value, which will make continuous attacking impossible for a faulty node. The difficulty for a specific node is

$$D_u(t) = \begin{cases} DL1, & \text{if } Cr_i(t) \geq \alpha_d \\ DL2, & \text{if } Cr_i(t) < \alpha_d \end{cases}. \tag{108}$$

A new block is designed to be concatenated with n_{wh} number of former blocks through the dPoW consensus replication approach. Let B_j represent the j^{th} block of the blockchain, B_c the current block that should be mined, and N_c as its nonce value. Hash of the current block (B_c) is formulated as

$$H_c = h\left(\sum_{j=1}^{n_{wh}-1} B_{c-j} + [B_c + N_c]\right), \tag{109}$$

where H_c is the hash of the current block and $\sum_{j=1}^{n-1} B_{c-j}$ denotes the $n-1$ W-Hash blocks. If H_c meets the difficulty level requirement, i.e., prefix zero length, the block can be mined, and the nonce value is successfully found. Adjusting the difficulty level based on the prefix zero is already discussed in Section V-C. A larger minimum required prefix zero makes it more difficult to find the nonce value and to mine a block. Algorithm 4 summarizes the block mining procedure of the TB-ICT framework.

Algorithm 4 TB-ICT Block Mining Algorithm

Result: Target Hash String

initialization
Check Eligibility of the Node:
if $Cr\boldsymbol{u} \geq \alpha_d$ **then**
    High Credit Node
    set target, $T =$ "0"
else
    Low Credit Node
    set target, $T =$ "0000"
end
set nonce: $N_c = 0$
while mining **do**
    for block.index $\in [0, 100]$ **do**
        $H_c = h\left(\sum_{j=1}^{n_{wh}-1} B_{c-j} + [B_c + N_c]\right)$
        if $H_c == SHA256[T]$ **then**
            break
        else
            N_c++
        end
    end
end

5.4 Security, Complexity, and Scalability Analysis of the TB-ICT Framework

5.4.1 Security Analysis

In this section, we analyze vulnerabilities and thread models of blockchain, illustrating how the proposed TB-ICT overcomes these threats based on its underlying design. While blockchain networks can be utilized to design secure IoT-based platforms, there are some vulnerabilities that blockchain-based platforms suffer from [73, 85]. Generally speaking, attacks on the blockchain can be categorized into application-based or application-free attacking models [198]. In the latter category, the vulnerabilities within the application are targeted by adversaries. In the application-free attacking model, on the other hand, the blockchain itself is the attacking target. To be more precise, the attacking scenarios can be classified into the following different threat

models [85], i.e., Identity-based Attacks, Manipulation-based Attacks, Reputation-based Attacks, Service-based Attacks, Linkage Attack, Enumeration Attack, Location Confirmation Attack, and Device Tracking Attack. The major threat analysis model considered here is the STRIDE (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege) [200]. In what follows, STRIDE threats in the context of the proposed TB-ICT framework is introduced.

1. Spoofing: In spoofing, another entity (e.g., a person or a system) is impersonated in the network, and the authenticity is violated. Identity-based attacks can be considered in this category. In this context, the identity of an authorized user is forged to enable an adversary to gain access to the system and manipulate the network. The following attacks that can cause spoofing in the network:

- *Key Attack*: When the private key of a user in the network is leaked, the attacker can use this key to impersonate that user's identity to perform fraudulent activities in the network and consequently in the designed TB-ICT platform, causing a significant reduction in the credit of that user. Our proposed mechanism is designed to create temporary private keys to prevent such an attack. To be more precise, a time-varying signature (32 bytes) is first generated based on the ambient features, and a secret key (32 bytes) is then randomly generated/extracted from this signature. Such a signature generation scheme is repeated in short time intervals resulting in the generation of temporary (time-varying) private keys, which can inherently prevent some major attacking scenarios in the network. Consider a situation where a private key is leaked; in a short while, this private key will be changed, and it would not be authorized to sign anything or create any address in the network. The data generated in the network is time-stamped, and there is no way to gain any trust in the network based on the stolen key. In summary, to prevent the key attack, the proposed framework relies on its time-varying private key generation scheme. Furthermore, elliptic curve encryption is employed to calculate the hash function, which further improves the resiliency of the proposed model against key attacks.

- *Replay Attack*: In this attack, by spoofing the identity of two connected nodes, an adversary can create an interception in the transferred data packet and relay the data to the destination without any alternation [198]. In other words, the advertised message via BLE modules can be captured by a fraudulent node and

can be used at another location at a different time. This can mainly be used in CT platforms to create false-positive results during the tracing transaction in a positive infected case. When the private keys are generated without sufficient randomness, the key leakage can cause such a problem. The proposed TB-ICT platform is resilient against such replay attacks as it uses a temporary time-stamped private/public key generation scheme. Consequently, conducting such an attack is unlikely as a node cannot claim via previously revoked addresses. To be more precise, the address and the description of the message stem from a tracing event are generated with a detailed time description. The adversary fails in this attack unless she/he can temper the message content and also forge a valid signature, which is impossible due to the utilized temporary private key generating scheme and time-stamped messaging events.

- *Impersonation Attack*: This attack can happen in the case of private key leakage. In this situation, an adversary node can masquerade as an authorized entity and try to conduct unauthorized behavior or apply false claims. The temporary private/public key generating scheme of the TB-ICT couple with the credit-based dynamic PoW solution makes this attack unlikely. Furthermore, the non-monetized credit-based definition of the platform reduces the incentive of the nodes to perform such faulty operations. Literature-wise, the following three approaches [198] are recommended to prevent impersonation attack: (i) Defining temporary private keys for defined sessions [201]; (ii) Distributed cooperation solution based on an incentified scheme [202], and; (iii) Replacing the ECDSA signature with another signature solutions based on the attributes of the nodes [203]. First, the signature generation scheme of the TB-ICT (aligned with approach (i)) is an attribute-based solution, which provides a defense against such an attack. Second, the credit-based Dynamic PoW solution of the TB-ICT introduces an incentive-based mechanism (aligned with approach (ii)) that controls the tendency of fraudulent activities/claims specifically in the mining process, making the attack very costly for an unauthorized and low-credit node.

- *Sybil Attack*: In this attack, an adversary targets taking over the network by creating several fake identities. To prevent such an attack in the TB-ICT, we considered the following mechanisms. For the manager and authorized entities

in the network, which are high-credit nodes by definition and are mostly involved in the mining and approving procedures, a referral-only system based on the credit-based approach is designed to check the eligibility of these nodes. This referral solution has a probationary period and reputation checking stage where authorized nodes are approved to be in the network. Each manager node also is enabled to keep track of all authorized nodes in the network by signing a transaction TX with its secret key, of the available Authorized Nodes' public keys $\left(PK_{AN_1}, PK_{AN_2}, \ldots, PK_{AN_n}\right)$. This will keep the network reliable. Although there are some other solutions, such as Lagrange Interpolation in signature generating method [204], our proposed solution is using a platform to approve the validity and integrity of the transactions. Any claim regarding the infection and any transaction submitted based on the CT solution will be approved by the IUP database via the authorized and high-credit miner nodes. To be more precise, when a node is recognized as an infected one by the authorities, all its addresses in the past 14 days will be uploaded to the IUP, at this stage two verification steps are considered to be employed to finally submit a transaction in the blockchain: (i) Verification of the infected node to be sure its addresses can be found in the IUP, and; (ii) Verification of the claim with the node receiving the claim, which is issued automatically through the application after approving the presence of such an address in the contact's database. This method would highly prevent Sybil attack from happening. Another solution is proposed in [205], which can be a future research direction. In [205], the authors address the vulnerability against Sybil attack by proposing an immutable temporary chain of interaction between agents and defining a mechanism to calculate the trustworthiness of the nodes. In our proposed solution, the trustworthiness of the nodes is computed based on a credit-based solution, and a temporary localized database keeps the contacts' history of each agent that is used for the correctness and integrity of the claims and the whole network. This is impossible for an adversary to create many fake identities to submit false claims and tamper with the data in the network.

- *Man-in-the-Middle Attack*: The man-in-the-middle attack can happen when the private key is forged, and an adversary, by intervening in a mutual connection by spoofing the identity of both entities, can control the transactions exchanged

between them and perform fraudulent operations. This attack can be categorized as Manipulation-based attacks as well. In our represented framework, the messages are exchanged between the nodes in a non-connectable advertising mode, and the transactions are submitted to the network using a temporary address definition. There is no static private key for the entities and forging one key provides limited time access for a fraudulent node as in a short time-stamp an alternation will be applied to all the keys and addresses. These features are protecting the proposed framework from this attacks.

2. ***Tampering***: In tampering, the data is modified via a malicious node, which is violating the integrity of the data in the network. Manipulation-based attacks are considered in this class where data is achieved and tampered through unauthorized access to the network. Man-in-the-middle, overlay, modification, and false data injection attacks can be included in this category and can tamper with data in the network [85]. The latter, i.e., false data injection attack, is mostly seen in the IoT sensory networks and smart grid solutions [204] where the sensory measurements are, typically, tampered with to endanger data integrity and data aggregation steps in IoT networks. Our proposed solution, however, is not working based on data aggregation scenarios and evaluations based on the sensory data. The AoA-based localization solution leveraged the smartphone's sensory data to find the proximity of the users, and no raw sensory data is submitted as a transaction in the network. Moreover, the TB-ICT's verification solution and its credit-based dynamic PoW approach can prevent any false injection claims from threatening the integrity of the network. In the case of overlay attack, which is mainly happening in the monetized blockchain platforms, the resistance against this attack is guaranteed as the transactions are added with a time-stamp to ensure their uniqueness. The modification attack is also unlikely in our designed platform since, similar to [201], the proposed solution is exploiting a temporary private key generating scheme based on the ambient environmental information. The man-in-the-middle attack can also result in tampering with the data and be categorized as a manipulation-based attack. In the proposed TB-ICT framework, messages are transmitted between nodes by leveraging a non-connectable advertising mode, and temporary addresses are used to submit transactions. There is no constant private key for the entities in the network, and even forging one identity would not be a long-term success for the adversary as all the keys and addresses will be changed

in a short period of time. Consequently, TB-ICT can be considered completely safe against manipulation-based attacks.

3. Repudiation: In Repudiation, a user can deny taking an action. Protecting the system from this threat requires a robust authentication and accurate design to attribute users to their related actions. Reputation-based attacks can be considered in this category, where an adversary tries to affect the network by changing its reputation positively. This attack can happen in our proposed network when the low-credit nodes try to falsify their credit score to send false claims and fraudulent transactions. The reputation-based attacks can be classified into two main classes, i.e., (i) Hiding Blocks Attack, and; (ii) Whitewashing. Under the hiding blocks attack, those transactions that positively affect the node's reputation will be exposed via the node. This is an unlikely attack in our proposed solution, as all the interactions of the nodes in the network are performed in an online fashion via the AoA-based proximity approximation algorithm, based on which each node's credit is calculated. On the other hand, when a node has a positive infection result, a trace transaction will be sent to the network, and this can be done automatically when a node approves the infection in the application. If a node refuses to send the transaction to the network, its credit will be affected negatively, limiting its access to the network. In the whitewashing attacking scenario, an agent with a negative reputation (very low credit) can remove its virtual identity and create a new clean identity. Although there is no way to prevent such a faulty behavior from happening, in the proposed framework, based on the formulated credit-based scheme, those newly added nodes to the network will start with a random, very low credit score, and it takes time for a node to gain credit based on its behavior in the network.

4. Information Disclosure: Unauthorized access to the sensitive data is addressed via information disclosure. As discussed earlier, in the proposed TB-ICT platform, private keys of the nodes, the user's real identity, and the users' real location can be considered sensitive data. However, no personal information of the users or their contacts is sent as a transaction to the ledger or saved in personal devices. Moreover, transferring of the messages between nodes happens in a non-connectable advertising mode, and a temporary address is used to submit transactions to the network. Private keys are changed constantly, and data is encrypted in the framework. Linkage attack, enumeration attack, location confirmation attack and device tracking attack which

all want to disclose sensitive data, can be categorized in this class and are thoroughly discussed here.

- *Linkage Attack*: Here, the attacker, by leveraging the information collected in the network, is trying to reveal the anonymous identity of a node. In our context, the identity of the infected user or her/his close contact can be targeted, but this attack is impossible in our proposed framework as there is no information stored directly about the identity of a node or its contact list. All the data that is stored and then used in the blockchain is the temporary addresses related to those nodes. There is not even a constant private key to sign the transactions, and all the keys are time-varying.

- *Enumeration Attack*, can happen when an adversary tries to approximate the number of infections based on the nodes who successfully submitted their tracing transaction after being recognized to be infected. Because there are multiple temporary identities without any chance to find and match them for a specific user, finding such information is impossible in the network. In other words, there can be many addresses related to an infected person and its close contacts, and tracking the unique identities is not a logical operation.

- *Location Confirmation Attack*: There is a tendency for the adversary nodes to find or approve the presence of a specific node in a specific location. This attack can happen by leveraging the unique identifier of a user, such as a smartphone used in the CT procedure. This is an impossible attack as well since the TB-ICT platform never uses unique or constant identifications, and the platform is designed based on a temporary key generation scheme.

- *Device Tracking Attack* As most smart CT solutions employ the BLE sensory data to find the proximity and track a device, this information can be misused to find a movement pattern or track a device. The BLE-enabled application itself uses a randomized MAC address in short time intervals, and the platform generates temporary behaviors in short time-stamps in non-connectable advertising mode to prevent such an attack.

5. Denial of Service (DoS): In DoS attacks, the ability of the system to provide its expected performance and respond to the requests is degraded and destroyed. In

CT specific use case, DoS attack can happen by creating repeated false advertising where an adversary can try to consume the storage capacity and increase the power consumption of another BLE device. Similar to other smart CT solutions, the proposed TB-ICT framework is vulnerable to this attack, and fake interactions can be sent via an agent in the network through this kind of attack. On the other hand, in the proposed blockchain network, the size of the blocks is considered to be a limited value, and also checking scheme is applied for the transaction input to find the attribute signature. These make the proposed blockchain design more resilient against such attacks. DoS is categorized as a service-based attack, which is another common class of attacks that can happen in blockchain-based IoT platforms. Refusal to Sign Attack, Double-Spending, and collusion are categorized in this class of attacks. In the proposed solution, as there is no monetized structure, double-spending and collusion attacks are not applicable.

6. Elevation of Privilege: This can happen when a fraudulent node can perform unauthorized actions by escalating the privilege via gaining elevated access to the resources and protected functionalities of a platform. In a blockchain network, there are different levels of access (e.g., account on a distributed ledger or smart contacts) that can be compromised and gained by an unauthorized user. In the proposed TB-ICT framework, however, the credit level of the users and their accesses are designed carefully. The temporary private/public key generating scheme of the TB-ICT couple with the credit-based dynamic PoW solution makes accessing the account of the users and forging the identity unlikely. This design can make the TB-ICT framework more resilient to this vulnerability.

5.4.2 Computational Complexity and Consistency Analysis

Considering Eq. (109) and Algorithm 4, hash of the current block (B_c) (if the block is not already mined) is the hash of the concatenation of the current block and $n_{wh}-1$ of the previous blocks with the nonce value. n_{wh} is the current $W-Hash$ value, which is defined and shared between the high-credit and legally authorized nodes based on their credit. To mine the block, miners compete to solve a mathematical puzzle and find the target hash string by adding the nonce in each step via a trial and error procedure. The target hash string in the TB-ICT platform has 64 digits in hexadecimal base. A compressed representation (called “Bits") of the target hash

string hex value, e.g., $0x1b0404cb$ is stored in the blocks. The hexadecimal value is calculated based on this packed form. For example 0000 0000 $2FF4$ $04CB$ 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000, hexadecimal string of length 64 and the hash value should be less than this value to mine a block. The leading zeros (eight leading zeros 0000 0000) in the target hash string represent the difficulty level. As can be seen in Table 5.1, for the nodes with higher credits and the legally authorized nodes, the first significant digit (1 digit) of the block hash string is considered the leading zeros. This value will be four significant digits of the block target hash string for the nodes with bad credit scores. To be more precise, lower target values will lead to the greater difficulty of block mining. Assuming the length of target hash string is b bits, for a $W - Hash$ value of size n_{wh}, computational complexity is of $O(n_{wh} \times t_{TH})$. Term t_{TH} is the required time to generate the target hash string in the search space of 2^b. Therefore, computational complexity is of $O(n_{wh} \times 2^b)$. The malicious node is now forced to find the hashes for all the possible values of the $W - Hash$ sizes, which directly increases its computational complexity and, consequently, the attack's cost. The complexity in this scenario is increased as

$$Complexity = O(n_{wh} \times (n_{wh} \times t_{TH})) = O(n_{wh}^2 \times 2^b) \tag{110}$$

This attribute also makes other possible attacks such as double-spending and spamming infeasible. Without the W-Hash, the complexity of the network to conduct an attack would be as low as $O(t_{TH})$, while this complexity, after applying the hash window, will be $O(n^2 \times t_{TH})$, which requires $n \times n$ more test and operation to find the block's target hash string. Consequently, the security of the network is significantly improved. On the other hand, based on the credit-based platform, any malicious attempts to even break this difficulty level will have an irreversible and high credit punishment, causing even higher computational costs to the faulty nodes.

Network Consistency: In brief, in the proposed framework, we cannot assume that all the miner nodes are in the same state at each time. In other words, due to the inherent message delays [206], miners may see different status for the network's latest updates, e.g., some miners may have the most recent mined block while others have access to an older version. Requiring the vote of all miners to mine a block as a solid condition for reaching consensus on the longest chain would be a complex condition to be reached and gain consistency in the network. Achieving agreement

on the current chain by ignoring a small number (i.e., T number) of unconfirmed blocks at the end of the chain is a solution to have another version of the consistency requirement. This notion of consistency is named T-consistency [207]. Honest nodes, which are qualified based on the proposed credit-based approach, participate in this decision-making process. In this context, we prove that by choosing a sufficiently large T value, there is no chance for a confirmed block to get lost from the blockchain. This would prove the consistency of the network.

In our W-Hash-based dynamic PoW solution, consider p as a mining hardness, i.e., m number of miners successfully mine a block with a single random oracle query with probability of p. It is assumed that all the miners participating in this process have the same computational power in mining the blocks. If the protocol is proceeded in different rounds (time-stamped steps), while each honest node as a miner gets one random oracle query in each round, a ρ fraction of dishonest nodes (adversaries) get ρm number of queries [208]. Honest miners perform their query operations in an asynchronous (parallel) fashion while the adversary miners are allowed to perform sequential query operations [208]. Let us assume the probability of honest miners mining a block in a round to be as follows

$$q = 1 - (1 - p)^{(1-\rho)m}, \tag{111}$$

and the expected number of blocks that can be mined by dishonest nodes be $\gamma = \rho m p$. Considering the case when $p \ll \frac{1}{m}$ then we have $q \approx p(1 - \rho)m$ and consequently $\frac{p}{q} \approx \frac{1-\rho}{\rho}$. In this case, considering that $\exists \delta > 0$ such that

$$q(1 - (2\theta + 2)q) \geq (1 + \delta)\gamma, \tag{112}$$

where θ is the network latency, then T-consistency can be approved in a random oracle query model except in exponentially small probability in T. Consequently, until the adversaries' computational power is less than half of the total computational power there would be some p for any θ to satisfy the consistency of the proposed platform. If $p > \frac{1}{\rho m \theta}$ this consistency will fail.

5.4.3 Scalability Analysis

For blockchain platforms, scalability (Transactions Per Second (TPS) rate) is highly dependent on the block size. There is, however, a trade-off between block size and speed, consequently the scalability of the whole network, i.e., increasing size of the blocks will result in decrease of the speed and throughput of the nodes. To improve scalability of the proposed blockchain-based solution, applying scaling of blocks could be beneficial. More specifically, to scale a blockchain network, there are two main approaches, each of which has its own pros and cons [209], as follows:

• ***On-Chain Scaling***: In an on-chain approach, the blockchain itself is being changed or redesigned to address the scalability issues. This can be achieved via the following techniques:

- Shrinkage of each transaction's data [210] to increase the number of transactions that can be encapsulated in each block. Another similar shrinkage method, applied in the Bitcoin platform, is the Segregated Witness update known as SegWit [211], which alters the network's capacity.

- Increase the block generation rate [210], which in turn can effectively increase the TPS rate. This approach, however, comes with the drawback that the newly generated blocks should be propagated in the network to enable the nodes to know the full status of the network.

- Creating a seamless connection between discrete blockchains and sharding transactions to enable different nodes to verify specific shards in a parallel fashion. This method suffer from unreliability issues, i.e., it can take years to implement such an approach on platforms like Ethereum. Moreover, there are many security concerns regarding these solutions, e.g., sharding itself can increase the possibility of double-spending, and 51% attack.

• ***Off-Chain Scaling***: In off-chain solutions, no change is made directly on the blockchain itself, instead they are implemented on top of the blockchain layer to address main chain issues. This can be achieved via the following techniques:

- Lightning Network [212] creates channels between a network of nodes to transact the data and only interact with the main chain in some time intervals. This can make the interaction between the nodes faster and cheaper.

- Branching off the main blockchain and creating side-chains for certain tasks to free up the overall network's bandwidth.

Off-chain scaling solutions are not considered in this work, as the temporary identities of the nodes in the designed IoT platform can make the solution hard to implement and more complexity. Moreover, for the lightning networks to be successful [212], there should be considerable liquidity to be locked up in the network, yet the proposed solution is not a monetized framework. Consequently, the on-chain scalability solutions are considered to be used in the TB-ICT. Among the aforementioned on-chain solutions, creating a seamless connection between discrete blockchains and sharding transactions is complex in terms of implementation and the overall network. Additionally, this approach suffers from security concerns as such is considered in this work. Shrinkage of each transaction's data is the main approach implemented in the TB-ICT. Transactions are the packets of the possibly infected contacts' addresses, and no more data is pushed to the blocks.

Next, we discuss differences and similarities between the transaction's data structure of the proposed TB-ICT and that of conventional blockchain platforms, i.e., Bitcoin and Ethereum.

• ***Data Shrinkage Solution***: In Ethereum [197], all transaction types have the following fields as their logical structure:

- *type* (T_x), which is $EIP-2718$ transaction type.
- *nonce* (T_n), is a scalar value that consumes up to 32 bytes and is equal to the number of transactions sent via sender.
- *gasPrice* (T_p), which is another scalar value that consumes up to 32 bytes and is the number of the Wei to be paid for the computation costs to execute the related transaction.
- *gasLimit* (T_g), is another scaler value that needs more than 32 bytes defining the amount of the gas required to be paid up-front to execute a transaction before applying any computation.
- *to* (T_t), the 20 bytes recipient address.

- *value* (T_v), which has up to 32 bytes size and is the number of Wei to be transferred to message call's recipient or newly created account in case of creating a contract.
- *r, s*, denoted by T_r and T_s, each of 32 bytes, are values related to transaction's signature and the sender's determination.
- *Data*, which ranges from 0 to unlimited bytes, but due to the gas price and transaction cost, is highly limited.

Such a logical data structure is, however, longer in practical scenarios as it is Recursive Length Prefix (RLP)-encoded for serialization across Ethereum's execution layer. Considering the above definitions, the transaction size is still considerable in Ethereum and is much higher than 100 to 200 bytes [213]. This situation is more complicated for the Bitcoin. While there is no block size limitation in the Ethereum design, we have this limitation in Bitcoin. A Bitcoin transaction deals with Unspent Transaction Outputs (UTXO) as such it needs references to the previous UTXO hashes. To reach an specific value to be transferred via the Bitcoin network, there are one or more UTXOs to sum up and create that specific amount. Moreover, when a transaction is issued in Bitcoin, there could be more than one output. This case gets even worse when we assume there is no SegWit solution [211]. All these results in the Bitcoin's average transaction size to be about 600 bytes [214], which is larger than that of the Ethereum's transaction size. The TB-ICT framework is designed to be as light as possible. More specifically, no data is exchanged via transactions. Furthermore, when a transaction is approved and mined, essentially, it means that the receiver was in close contact with a person that is recognized to be infected with the disease. Therefore, there is no monetized structure, there is no gasPrice, gasLimit, value and data exchanged via transactions. The credit of a node itself is designed to be used in the network only when a node is trying to be a miner, otherwise, the local application keeps tracking proximities. Considering all these design architectures, the logical structure of a transaction in the TB-ICT can be described as follows:

- **type** (T_x), which is the transaction type ranging from trace transaction to query transaction.
- **nonce** (T_n), is a scalar value that consumes up to 32 bytes and is equal to the number of transactions sent via sender.

- **to** (T_t), which is the 20 bytes recipient address.
- **r, s**, denoted by T_r and T_s, each of 32 bytes, are values related to transaction's signature and sender's determination.

This structure keeps the transaction light. In extreme peak transaction scenarios when the TPS could be around 580, approximately 3 blocks needed to be mined in each second, which is an achievable state. Furthermore, for the easier difficulty level (DL_e), the average time required to mine a block for all different W-Hash values is very small, therefore, mining higher number of blocks can be handled easily.

To summarize, the following key features of the proposed TB-ICT framework contributes to its higher TPS compared to other well-known blockchain networks such as Ethereum: *(i) Improved Consensus Protocol*: In the TB-ICT, we proposed an innovative consensus protocol, i.e., dPoC to improve the network's throughput. The dPoC solution is the core engine of the TB-ICT framework and does not require miners to solve hard cryptographic algorithms using massive computational power. On the contrary, it ensures consensus through the selection of validators according to credits in the network. Adopting the dPoC consensus could substantially boost the capacity of the proposed solution alongside improving security and decentralization. The proposed dPoC consensus approach is converged with the IoT network and the application-specific design to provide contact tracing services for epidemic control. In short, the high achieved TPS, is not only because of adjusting hash difficulty but mainly due to the proposed dPoC consensus approach that is coupled with the underlying IoT network and its application-specific design. *(ii) Reduced Transaction Size and Elimination of Nonessential Concepts Affecting Scalability*: The transaction size in the proposed solution is 2 to 4 times less than that of the average transaction size in the Ethereum network (without considering the data field). This, however, is not the main reason behind the higher scalability of the proposed framework as described above in the context of improving the consensus protocol. With regards to the transaction size, in the Ethereum network, the concept of "gas" and its associated fee ("gasPrice") for each transaction is a solution to prevent spamming of the network and is essential to handle the limitation of space and time for mining a block. The gas limit, however, can drastically prevent many transactions from being mined and added to the blockchain as such, the TPS reduces significantly. Unlike Ethereum, there is no concept such as gas in the proposed TB-ICT. In other words, the TB-ICT framework

inherently is not a monetized framework. It is a blockchain-based IoT framework for exchanging a small amount of data. To elaborate more on the relation between the gas price and the scalability issue in the Ethereum networks, we can review the Ethereum Layer 2 solutions [215] proposed to address the scalability challenge. In these solutions, transactions are rolled up into a single transaction submitted to the Ethereum Mainnet in order to reduce gas fees for users and make Ethereum more accessible. In the proposed TB-ICT framework, there is no need to leverage such a solution as the framework is designed based on dPoC consensus protocol and all the barriers for the higher TPS such as gas are removed. In short, the following items are key factors contributing to the high TPS ratio of the TB-ICT compared to Ethereum:

- Implementation of the dPoC consensus protocol instead of the PoW consensus and non-dynamic DL.
- Elimination of nonessential concepts affecting scalability, i.e., no gas limitation, no gas price, and no transaction cost, which highly boosts TPS of the TB-ICT.
- Simpler transaction data structure instead of resorting to complex data structures.
- Elimination of the data section, which is included in other platforms, making the TB-ICT almost ten times lighter in transaction size.
- Efficient block validation structure via high credit nodes with low DL instead of difficult and time-consuming block validation of the Ethereum.

5.5 Experiments

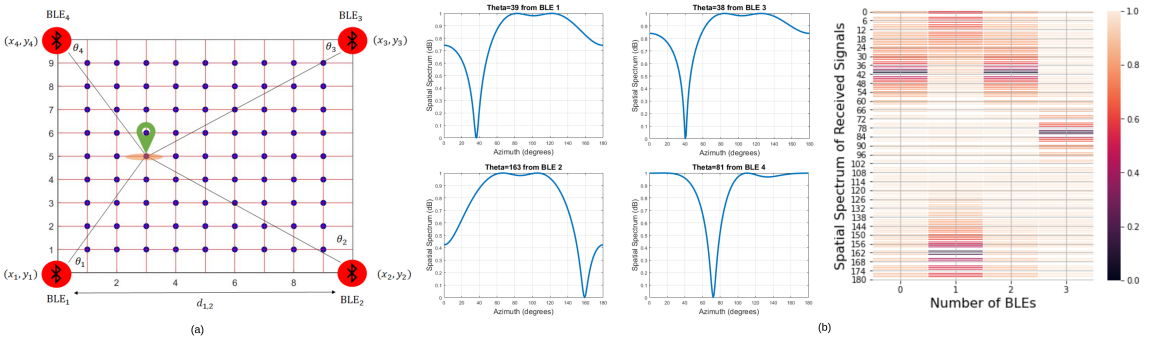
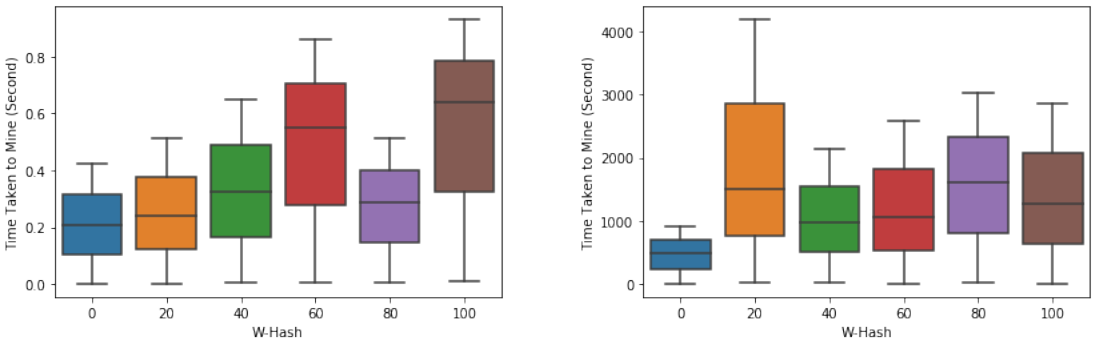


Figure 5.8: (a) Experimental data collection of the CNN-based AoA localization framework. (b) An angle image, used as the input of the CNN-based framework.

To evaluate the proposed TB-ICT framework, we considered a real experimental testbed in a rectangular indoor area $(10 \times 10)\ m^2$, divided into 400 square zones each with dimension of $(0.5 \times 0.5)\ m^2$ (see Fig. 5.8(a)). There are 16 BLE beacons, where the distance between each BLE beacon is equal to 4 m. There exists $1,000$ number of users, where their movements are modeled by a random walk.



(a) (b)

Figure 5.9: Exploiting dPoW for different W-Hash values to mine 120 sample blocks: (a) Difficulty Level DL_e employed for high-credit and legally authorized nodes. (b) Difficulty Level DL_h applied for low credit nodes (e.g., Malicious Nodes).

Table 5.2: LIST OF PARAMETERS.

Notation	**Unit**	**Value**	**Notation**	**Unit**	**Value**
DL_e	symbols	4	TPS	quantity/seconds	Positive Real Number
DL_h	symbols	16	$TargetHash$	-	SHA-256 hash algorithm
$Min.Time$	seconds	Positive Real Number	$No.Interactions$	-	Positive Integer
$Max.Time$	seconds	Positive Real Number	$W-Hash$	symbols	$\{0, 20, 40, 60, 80, and 100\}$
$Ave.Time$	seconds	Positive Real Number	$Credit$	symbols	Integer

Table 5.1 represents the list of parameters used in the experiments. Six different W-Hash values (0, 20, 40, 60, 80, and 100) are considered to evaluate the proposed system for its two different difficulty levels, i.e., DL_e and DL_h as are described in Table 5.1. Fig. 5.9 shows the different times spent to mine a block over different W-Hash values. The W-Hash 0, represents the classical PoW process while the other W-Hash values, concatenate the n_{WH} number of blocks with the current block in order to enter the mining process. These plots are formed based on localization data of $1,000$ indoor users where all the nodes are active (have movements in the environment) and have several interactions. All the interactions are considered to be collected and used to calculate the users' credit over time exploiting Algorithm 4. All

the proximities affect the credit value of the user where the close interactions will push negative scores to the credit, while near or far proximities (distances between 2 to 10 meters) will bring positive points to the user's credit. However, when a user is recognized to be infected with COVID-19, only the contacts in immediate distance to that user will be reported in a transaction to the blockchain. This will keep the block's incoming data light and improves the system's throughput. In our current evaluation setup, we mine 120 blocks and import a mean of 200 number of transactions to each of them. Fig. 5.9, shows the time spent to mine the block for different difficulty levels. Table 5.4, represents the maximum, minimum, and average time required to mine a block based on different difficulty levels. The time spent to mine these blocks based on the W-Hash values completely defines the superiority of the dynamic approach to make the mining process easier for high-credit nodes and legally authorized entities. As the dPoW is based on randomly selecting the W-Hash value for the current block to make the anomaly behavior harder to conduct, the average time spent to mine the block based on two difficulty levels can be a reasonable metric to contrast the applicability of the system. DL_e is far much smaller than DL_h, providing a much easier procedure for the high credit nodes to mine a block. While in Bitcoin [208], every 10 minute, a new block can be mined, this process in the TB-ICT framework is much easier and sooner for the high credit nodes. A high credit miner in this network can mine more than one block in each second and add the necessary transactions to the network in a timely fashion. In contrast, this process for the malicious nodes is much longer. The times needed to mine a block for the high credit and low credit nodes in the evaluation setup is shown in Table 5.4 and are discussed later in Section VII-C.

Table 5.3: Difficulty Level (DL), in the Proposed Dynamic PoW.

DL	**Size (bits)**	**Target Hash**	**Min. Time**	**Max. Time**	**Ave. Time**	**TPS**
DL_e	4	SHA256[0:1]	0.0018	0.9287	0.34501	579.69
DL_h	16	SHA256[0:4]	3.823	4986.3987	1286.6703	0.155

5.5.1 Localization Performance

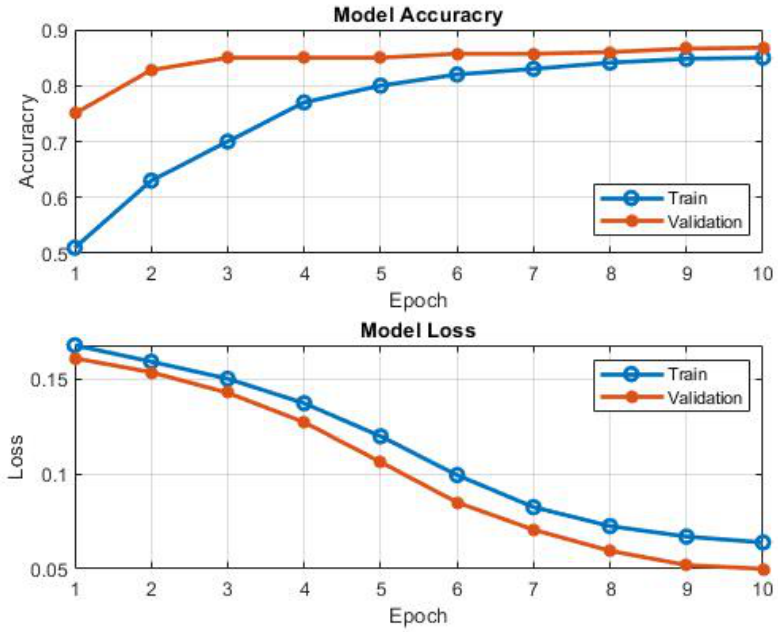


Figure 5.10: Accuracy and loss of the proposed CNN-based AoA scheme.

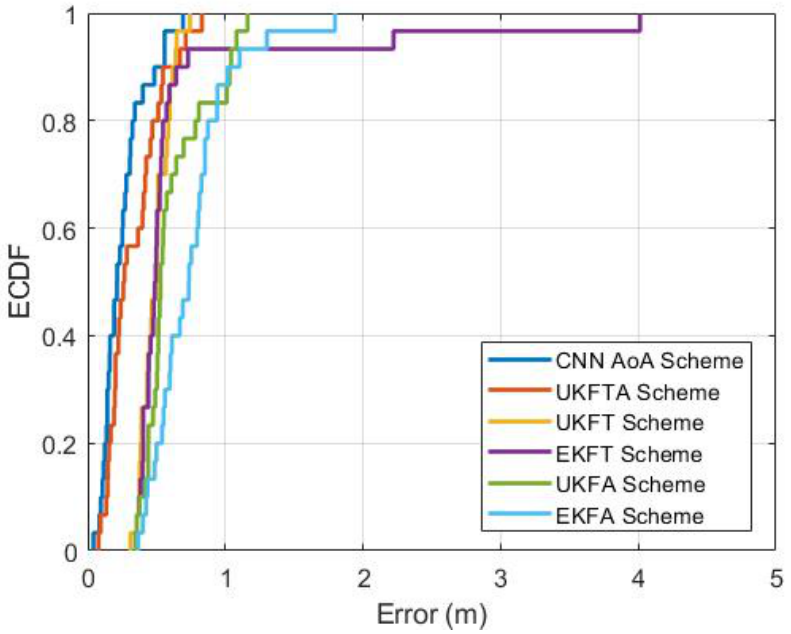


Figure 5.11: Location error ECDF.

Each user can be localized through a set of BLE beacons if the corresponding user is located in the receptive field of that BLE. The localization dataset is generated based on three different channel models: (*i*) AWGN model, where $10 \leq \text{SNR} \leq 20$ dB; (*ii*) Rayleigh fading channel, and; (*iii*) Rayleigh fading channel affected by noise in a 3-D indoor environment. We also consider the destructive effect of the elevation angle on the special spectrum of the AoA measurements in this dataset. We use $76,545$ angle images labeled by their location for training. Fig. 5.8(b) illustrates a typical angle image, containing 724 sets of AoA measurements generated by the subspace-based algorithm. The original angle image is of size 4×181, reshaped to be a square angle image of size 28×28 by zero padding. For the validation and test set, we use $15,309$ and $10,206$ angle images, respectively, which are all previously unseen and randomly chosen.

It should be noted that the proposed CNN-based AoA framework is a classification model, where the environment is divided into square zones each with dimension of (0.5×0.5) m^2. In such a case that the predicted zone and the actual zone that the mobile user is located, are not the same, a classification error will occur, where the localization error is defined as the distance between the exact zone and the predicted one. Following Reference [60], we converted the classification error to the localization error. The proposed framework tracks mobile users with 87% sub-meter accuracy in the presence of noise, Rayleigh fading, and elevation angle. While the accuracy of CiFi [57], MLP-RSS [60], Capon [216] and phase difference-based [45] frameworks are 40%, 81%, 74%, and 59%, respectively. Furthermore, the location error in Reference [58], introducing a CNN-based localization approach for the 2-D AoA estimation in the presence of noise, is 0.54 m. In a 3-D indoor environment, however, Reference [59] achieved 17° azimuth error, which is equivalent to a 1.2 m location error. It should be noted that although the location error in Reference [58] is lower than the proposed CNN-based AoA framework, the most important novelty of our study is that we consider the worst-case scenario, i.e., a dense indoor environment full of obstacles with multiple mobile users. More precisely, despite the existing research works [57–60] that are only capable of coping with noise, we consider a highly dense indoor environment without the presence of LoS links, where there are multiple mobile users. Another design option is the location resolution, i.e., the number of discretized points in the indoor environment (the zone size). Although the location resolution is

proportional to the location accuracy, it results in higher number of distinct labels for classification, where generating dataset would be more time consuming. Despite the recent DRL-based localization works [10], where the environment is divided into a grid of (5×5) m^2 and (3×3) m^2 cells, respectively, we assume higher resolution of (0.5×0.5) m^2 to improve the location accuracy. To evaluate the effect of location resolution, we consider another case study, where the location resolution is (1×1) m^2. In such a case study, the average location error is 1.53, where it is much higher than that of (0.5×0.5) m^2 location resolution.

The accuracy and the loss of the proposed CNN-based AoA framework versus the number of epochs are represented in Fig. 5.10. As it can be seen from Fig. 5.10, the model accuracy increases and the loss decreases in each epoch, which means the model is well trained. Moreover, the accuracy of the proposed localization platform over the test set is 87%, which is a high location accuracy in the presence of noise, Rayleigh fading, and elevation angle. To illustrate the superiority of the proposed approach, we compare the Empirical Cumulative Distribution Function (ECDF) for location error. As it can be seen from Fig. 5.11, CNN-based AoA framework is compared with Extended Kalman Filter AoA (EKFA), EKF Time Difference of Arrival (TDoA) (EKFT), Unscented Kalman Filter AoA (UKFA), UKF ToA (UKFT) [217], and UKF TDoA AoA (UKFTA) frameworks. According to the results, the location error of the CNN-based AoA framework is significantly lower than that of its counterparts. By leveraging the proposed CNN-based AoA localization framework implemented and embedded in our application, the proposed TB-ICT framework can positively prevent the spreading of the COVID-19.

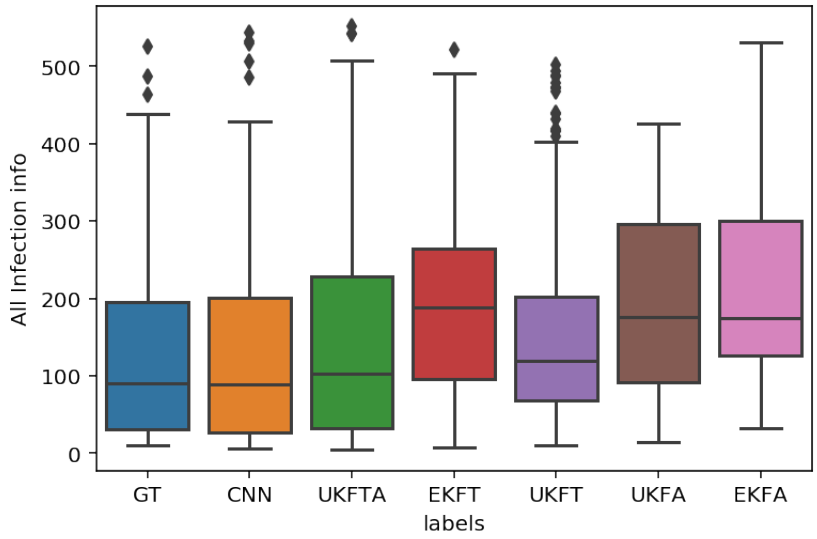


Figure 5.12: Number of users infected in the test indoor environment because of one infection case when a different proximity recognition algorithm is used.

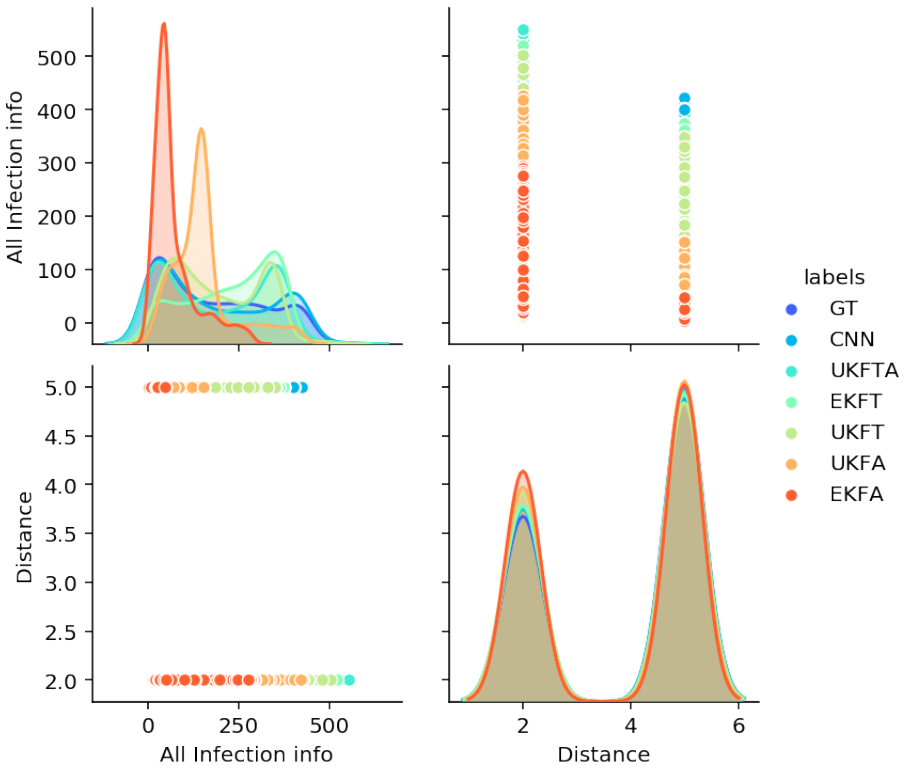


Figure 5.13: Infection ratio based on two different social distance measures, i.e., 2 meters and 5 meters.

Table 5.4: Average number of interactions of infections out of 1,000 users in indoor environment.

Metric	CNN	UKFTA	EKFT	UKFT	UKFA	EKFA
Avg. No. Interactions	**76,306.118**	75,770.248	75,803.404	75,640.436	72,247.544	72,811.806
Avg. Gained Credit	**57,111.127**	49,716.447	43,178.33	46,985.936	2,995.165	26918.931

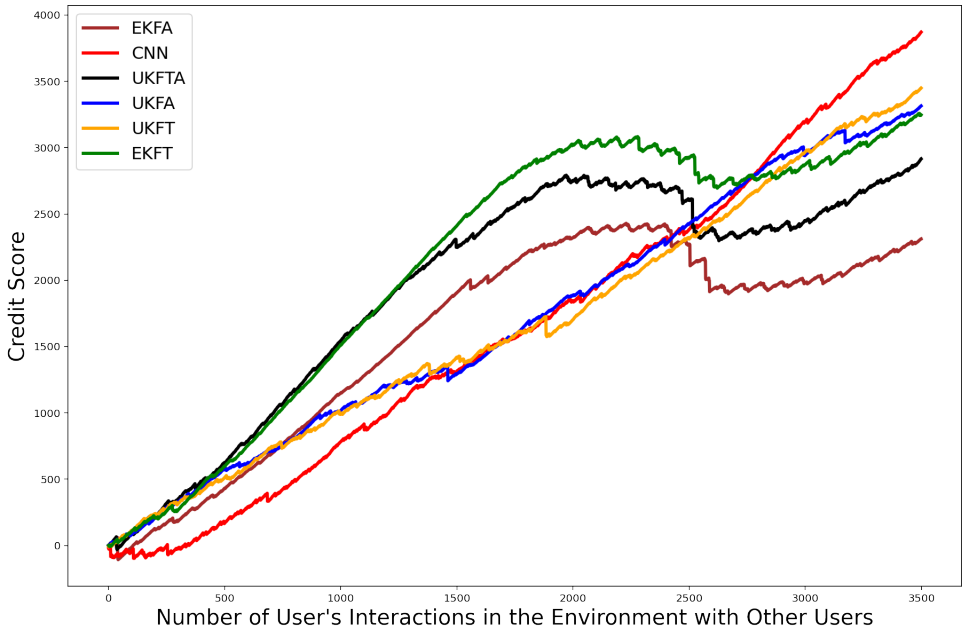


Figure 5.14: Credit Score Gained by Users in the Network Considering Different Localization Algorithms: Average Gained Credit Scores in First 3500 Interactions.

To evaluate the feasibility of the trustworthy blockchain-enabled system in terms of the noise level, we have computed the location accuracy of the TB-ICT framework (after a specific epoch where the CNN model is well trained) versus different SNR values. Note that the common value of SNR in wireless networks is in the range of $15 \leq SNR \leq 40$ dB [218], where the SNR greater than 40 dB and lower than 15 dB are considered excellent and unreliable connections, respectively. Here, the localization dataset is affected by noise, which is modeled by AWGN with $10 \leq \text{SNR} \leq 20$ dB. Fig. 5.16 illustrates the impact of the SNR on the location accuracy. According to the results in Fig. 5.16, increasing the SNR value leads to decreasing the location error. By considering the effect of the low SNR in the range of $10 \leq SNR \leq 20$ dB on the train dataset of the proposed TB-ICT framework, it can be observed from Fig. 5.16 that the TB-ICT framework is robust against noise.

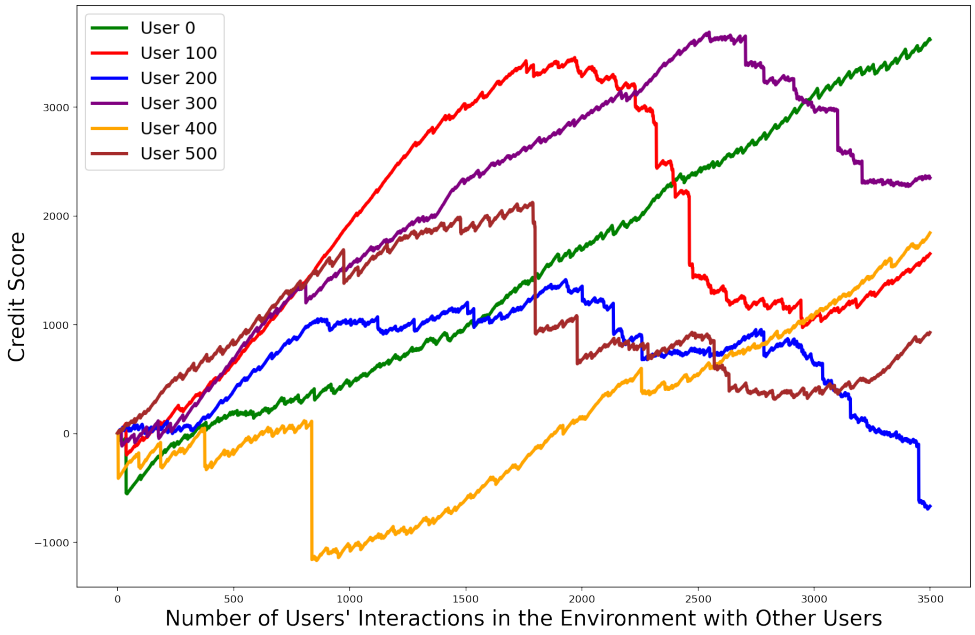


Figure 5.15: Credit Score Gained by Users in the Network Considering Different Localization Algorithms: Credit Scores Gained By Six Sample Users in CNN-Based Localization Approach.

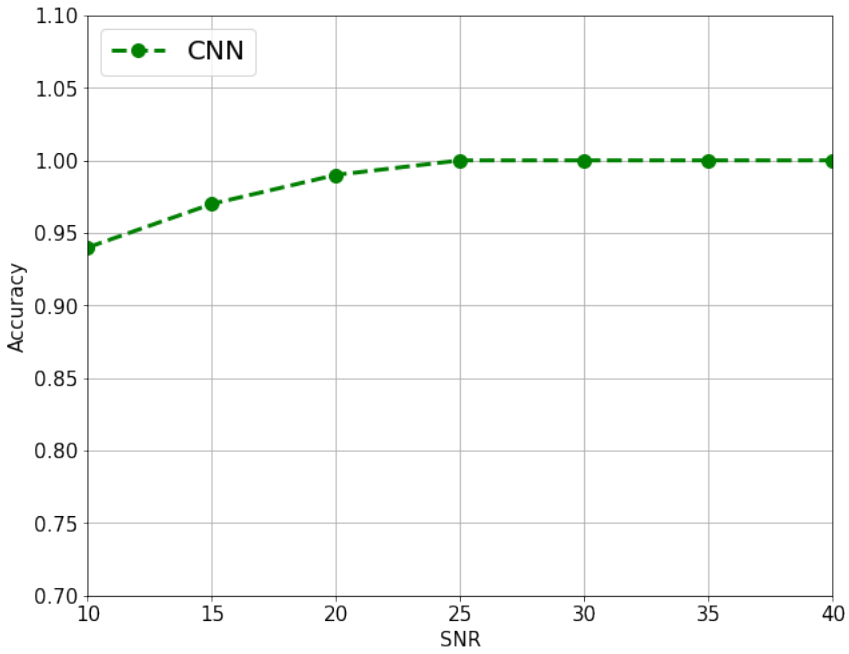


Figure 5.16: Location accuracy versus different values of SNR (dB).

Infection Spreading Nature:

Given the location of users during their movements and to illustrate the performance of the proposed TB-ICT framework, first, we investigate how one infected user can infect other users in its close contact. Fig. 5.12 evaluates the disease spreading trend. According to the results, one infected user can infect at least 341 users (out of 1000) by considering 2m as the required social distance criteria. Each of these users themselves can infect a large number of users in the environment (second cycle of infection). The contagiousness can be worsened by considering 5 m as shown in Fig. 5.13).

5.5.2 Evaluation of the Credit-based Mechanism

The credit-based approach can assist with management of the pandemic promoting users to act rationally. As stated previously, abnormal behaviors can be classified into two main categories, i.e., (i) Disease-Related Violations such as breaching the physical distancing rule and the rules related to the disease mostly bond to the proximity-based localization results, and (ii) Anomalous Behavior in Blockchain Network including different attacks or wrong claims, e.g., false claims of infection to the COVID-19 or violating the rule of sharing the last 14 days contact information in case of being diagnosed as infected. In our experiments, $1,000$ users are moving in an indoor venue having interactions with each other. Table 5.4 describes these interactions and related credit points exchanged between users based on different algorithms. It can be seen that TB-ICT has better results in gaining credits while the interactions are almost the same. In Fig. 5.14, average credit scores gained in the first $4,000$ interactions between one random user in the network considering different localization algorithms is illustrated. Agents following the TB-ICT approach are gaining more credit in the environment. This calculation of the credit is based on Eq. (104), while the $\lambda^{-} = 12$ and $\lambda^{+} = 2$ are considered as the credit calculation coefficients. Alternative coefficient values can be considered based on the pandemic situation and the community reactions. In Fig. 5.15, credit scores of six sample users out of all $1,000$ users are shown.

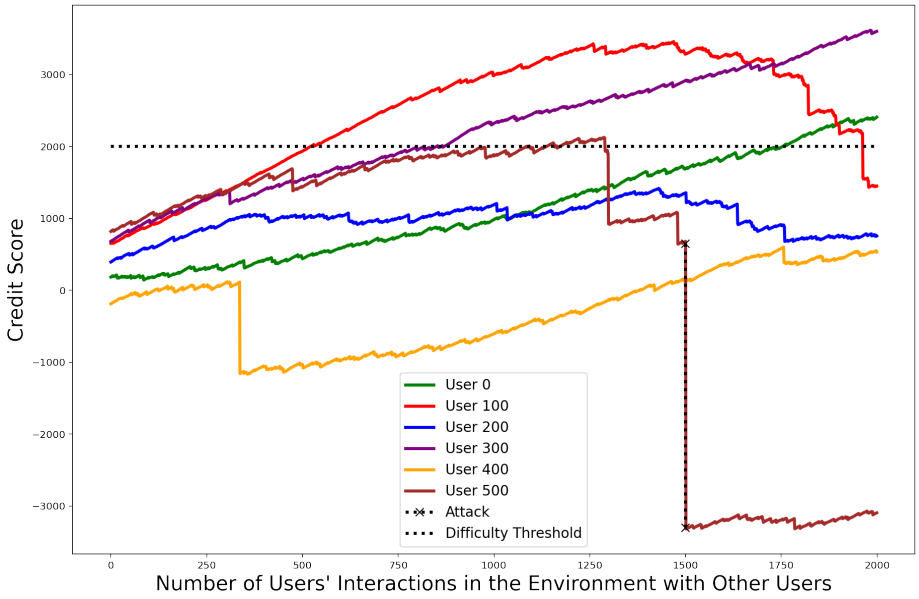


Figure 5.17: Network's behaviour to punish an attacker (User 500).

Any anomalous behavior will also be punished heavily in the TB-ICT network. As Fig. 5.17 shows, in our experiment, each malicious activity, e.g., attacks in the network, will be punished with a high negative credit score. The nodes that are punished based on these attacks lose the chance to be an honest node in the network or act as a miner. Moreover, Fig. 5.17 represents the difficulty level (α_d). Nodes with credit values above the threshold line can be considered as honest nodes and participate in the mining process. Difficulty level 1 (DL_e) has the lowest difficulty with considering the first one significant digit of the block hash string. The first level of difficulty is applied to the legal, authorized nodes and the nodes in the network with a credit bigger than a threshold (α_d). As can be seen, malicious behavior directly affects the credit values and can push the nodes' credits below the threshold line. Both nodes 0 and 500 experience this situation and losing their chance to be above the threshold line.

5.5.3 Scalability and Throughput of the proposed TB-ICT

Table 5.5: Mining time for different Difficulty Levels (DL) and different W-Hash values in the proposed dPoW.

		Mining Time		
		Min. Time	**Max. Time**	**Ave. Time**
W-Hash=0	DL_e	0.0025	0.4241	0.2063
	DL_h	3.823	920.2215	478.3168
W-Hash=20	DL_e	0.0018	0.5102	0.2413
	DL_h	12.8635	4186.3987	1512.8123
W-Hash=40	DL_e	0.0059	0.64967	0.3262
	DL_h	32.6557	2131.1228	976.6592
W-Hash=60	DL_e	0.0077	0.8576	0.5521
	DL_h	8.7308	2593.4276	1056.4611
W-Hash=80	DL_e	0.0049	0.5123	0.2897
	DL_h	15.7664	3037.4596	1601.2458
W-Hash=100	DL_e	0.0091	0.9287	0.6399
	DL_h	10.5398	2857.8972	1277.8963

Table 5.6: Transactions Per Second (TPS) rates for different DLs and different W-Hash values.

		TPS		
		Max. TPS	**Ave. TPS**	**Min. TPS**
W-Hash=0	DL_e	80000.0	969.46	471.57
	DL_h	52.31	0.42	0.22
W-Hash=20	DL_e	111111.11	828.84	392.0
	DL_h	15.55	0.13	0.05
W-Hash=40	DL_e	33898.31	613.12	307.85
	DL_h	6.12	0.2	0.09
W-Hash=60	DL_e	25974.03	362.25	233.21
	DL_h	22.91	0.19	0.08
W-Hash=80	DL_e	40816.33	690.37	390.4
	DL_h	12.67	0.12	0.07
W-Hash=100	DL_e	21978.02	312.55	215.35
	DL_h	18.98	0.16	0.07

Intuitively speaking, the maximum number of users who can send transactions at the same time in the TB-ICT platform is equal to the total number of users in that indoor venue. As an example, an indoor environment can be an office, a hotel, or a university building. These indoor locations, considering their maximum capacity, can not have several thousands of users at the same time. This means that the total number of transactions required to be handled every second is inherently not a large number. Considering this condition, below, we analyze latency, throughput, and scalability of the proposed TB-ICT:

• ***Latency***: As explained earlier, in our testing scenario, 120 blocks with a mean of 200 transactions for each are mined. Two different difficulty levels (i.e., DL_e and DL_h) and six different W-Hash values are defined and examined during this evaluation. As it is expected, when the W-Hash value is equal to zero, and we have a default PoW algorithm without W-Hash value, the average time needed to mine a block is minimum. By adding different W-Hash values, i.e., 20, 40, 60, 80 and 100, the required time to mine a block increases, however, when a miner is faced with an easy difficulty level (DL_e), this required time is decreased drastically. Considering the timing values mentioned in Table 5.4, for the easier difficulty level (DL_e), the average time (latency) for all different W-Hash values, required to mine a block is 0.0018 seconds. Latency (minimum, maximum, and average time) required to mine a block by solving the dPoW puzzle considering different difficulty levels and W-Hash values are shown in Table 5.5. Fig. 5.9 shows the required mining times as well. In the higher difficulty level, DL_h, the average time needed to mine a block considering all W-Hash values, is 1286.6703 seconds, which is drastically higher than that associated with the DL_e.

• ***Throughput***: To evaluate throughput of the TB-ICT framework, we look into Transactions Per Second (TPS). For comparison purposes and to have an insight on the number of incoming inputs and transactions, we can look at a classical payment platforms, e.g., VISA. Such platforms can handle about $56,000$ TPS considering its millions of users. On the other hand, the current TPS rate for Bitcoin and Ethereum networks are $5-7$ and $15-30$, respectively, which are not high ratios [219]. Fig. 5.18 shows the resulted TPS values in the experiments performed based on the proposed framework. As can be seen from this figure, approximately 580 TPS is mined, which is a much higher rate compared to well-known blockchain platforms and proves the

high throughput of the network. The TPS is only 0.155 in the higher difficulty level, DL_h, which successfully proves higher security against low credit nodes. The TPS rates for different DLs and different W-Hash values are shown in Table 5.6.

• ***Scalability***: As a final note, intuitively speaking, the number of inputs increases when a higher number of transactions is sent due to a hike in the number of infected users in the indoor venue. Considering that the main transactions that will be sent to the network is trace transactions in case of an infection, in the worst-case scenario all the users in an indoor environment are infected at the same time. In this regard, we consider a scenario where there are $50,000$ users in the network sending transactions at the same time (i.e., there are $50,000$ infected users at the same time). Considering the 580 TPS ratio, it only takes 1.43 minutes for the nodes to mine these requests and add them to the blockchain. This means that the system can efficiently handle such high peaks of the inputs pushed to the network. In such an extreme scenario with $50,000$ infected users at once, sending alerts in less than 2 minutes is significant, illustrating scalability of the TB-ICT framework.

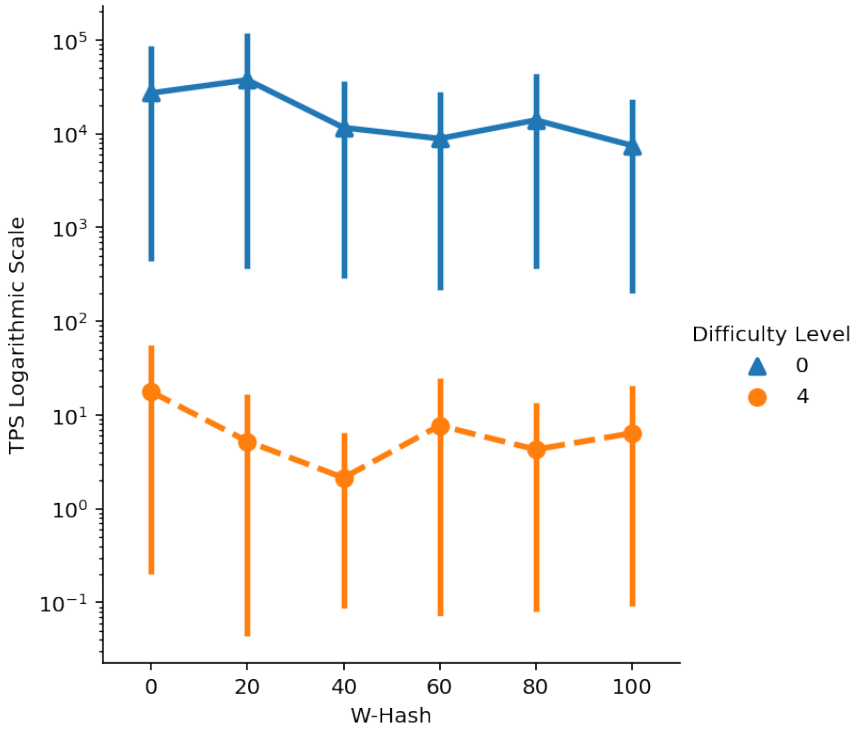


Figure 5.18: TPS for different W-Hash values: Difficulty Level 0 is DL_e the easy DL applied for high credit nodes and Difficulty Level 4 is DL_h the harder DL applied for low credit nodes (e.g., Malicious Nodes).

5.6 Conclusion

In this chapter, we present a robust and secure blockchain-based solution for indoor Contact Tracing (CT) to combat the spread of pandemics such as COVID-19. The proposed system, referred to as the TB-ICT, utilizes advanced AI-based localization techniques based on Bluetooth Low Energy (BLE) sensor measurements, as discussed in Section 5.2.1. To ensure the security of the communication between the BLE nodes, we have implemented a temporary signature generation scheme based on environmental features.

In addition, we address the security concerns related to CT platforms by proposing the Randomized Hash Window (W-Hash) Dynamic Proof of Credit (WDPoC) solution in Section 5.3. This platform employs a dynamic difficulty level through dPoC to distinguish between high credit and low credit nodes and address scalability issues without compromising the security of the network. The W-Hash mechanism is also employed to securely generate the hash of the following block, thereby increasing the complexity of the platform and making fraudulent behavior more costly. An incentive-based structure is introduced to calculate and maintain the users' credit score based on their behavior in the network and their adherence to physical distancing related to infectious diseases, such as COVID-19.

In Section 5.4, we thoroughly examine the security, complexity, and scalability of the proposed solution to emphasize its practicality. The results of our analysis demonstrate the robustness and effectiveness of the proposed system in providing an efficient, secure, and trustworthy solution for indoor Contact Tracing.

Chapter 6

Conclusion and Future Direction

In recent years, the advancement of Location-Based Services (LBSs) has led to the growth of various engineering applications of utmost significance. Indoor localization, in particular, has become a crucial aspect in providing innovative IoT-based services and in supporting public health during emergencies such as the COVID-19 pandemic. The BLE technology has proven to be a reliable solution for indoor localization and tracking. In this Ph.D. thesis, the focus was on developing a reliable and accurate BLE-based indoor localization solution, with a main application in epidemic control during public crises such as COVID-19. The study started by examining the BLE-based indoor localization, proximity classification, and analyzing the factors affecting the commonly used RSSI values in localization and tracking. The research then shifted to unsupervised learning localization, specifically Successor Representation (SR)-based approaches, as an alternative to supervised approaches that require labeled data. The challenge of fusing information from different BLE-based localization approaches and IMU-based localization results was addressed by proposing a Reinforcement Learning (RL)-based information fusion strategy. The most important use case of indoor localization, contact tracing, was also covered in the thesis. To address the need for trustworthy and accurate contact tracing during the COVID-19 pandemic, an innovative Smart Indoor Contract Tracing (TB-ICT) framework was developed by integrating blockchain and IoT technologies. The accuracy of the TB-ICT model was boosted by designing an efficient data-driven Angle of Arrival (AoA)-based indoor localization framework. In the following, we first summarize the thesis contributions, and then discuss potential directions for future research.

6.1 Summary of Contributions

The thesis contributions can be summarized as follows:

- **IoT-based Indoor Localization**: In Chapter 3, we discussed the development and evaluation of IoT-based indoor localization solutions. The IoT Tracking Dataset (IoT-TD) is introduced as a real-world dataset to evaluate the performance of different localization methods. The IoT-TD dataset contains synchronized data from AoA, RSSI, IMU, and the ground truth of the user's trajectories. This dataset covers three different indoor environments and provides millimeter-level accuracy. To enhance the performance of indoor localization, a hybrid localization framework that combines RSSI, PDR, and AoA is proposed. Additionally, a comprehensive evaluation of the proposed indoor localization solutions is performed using the IoT-TD dataset in three different environments. To minimize the impact of certain parameters in RSSI-based localization, an Indoor Localization SDK is developed that employs an ML-based model. The SDK is designed as a REST API, enabling seamless integration with various sensors in an indoor environment to provide real-time user proximity estimation.

- **Multi-Agent Adaptive Solutions and RL-based information solution for Indoor Localization Challenges**: In Chapter 4, two novel solutions for enhancing the accuracy and reliability of RL-based solutions were proposed. The first solution, Multi-Agent Adaptive Kalman Temporal Difference (MAK-TD) framework, integrates the Kalman filter into a Q-learning algorithm to address the challenge of unknown measurement noise covariance in MARL which leads to more accurate predictions and faster learning of the optimal policy. The second solution, Multi-Agent Adaptive Kalman SR (MAK-SR), extends the MAK-TD framework by incorporating the SR learning process into the filtering problem using the KTD formulation. The integration of SR and KTD results in a more reliable and robust algorithm with reduced computational costs and improved sample efficiency. Finally, a cutting-edge RL-based information fusion approach is introduced to enhance the accuracy and reliability of RL-based indoor localization systems. This hybrid indoor localization solution combines the strengths of various technologies through a RL-based

Fusion Framework (RL-IFF). This hybrid framework improves the overall accuracy of the localization system and reduces the impact of errors caused by any single method.

- **Blockchain-Enabled System for Indoor Contact Tracing in Epidemic Control**: The TB-ICT framework is a cutting-edge solution for secure and accurate indoor CT discussed in Chapter 5. To address the challenges of trustworthy indoor CT, the framework uses BLE beacons to accurately track close contact information and a secure and privacy-preserving communication protocol based on a blockchain-based distributed ledger. The Randomized Hash Window Dynamic Proof of Credit (WDPoC) solution is introduced to address the security risks in blockchain platforms by implementing a randomized hash window and credit score system to monitor and regulate node behavior. The inner indoor localization method in the TB-ICT model is designed to increase the precision and accuracy of indoor localization by using a Convolutional Neural Network (CNN) approach, which is efficient and reliable even in challenging indoor environments. The framework takes into account the impact of the elevation angle of the incident signal on the localization process to ensure robustness in the worst-case scenario.

6.2 Future Direction

Below, we present few fruitful directions for future research:

- Integration with other technologies: Future research can focus on integrating the proposed RL-IFF framework with other technologies such as Wi-Fi and Ultra-Wideband (UWB) to provide a more accurate and robust indoor localization solution. By combining the strengths of multiple technologies, the proposed hybrid indoor localization solution can provide improved accuracy, reliability, and security for CT purposes. Additionally, the RL-based information fusion scheme can be applied to effectively fuse the outputs of these different localization paths.

- Expanding the scope of performance evaluation: The proposed information fusion strategy, as demonstrated in this study, has been evaluated in three distinct

environments. However, to further validate its effectiveness and robustness, future research can aim to expand the scope of performance evaluation by testing it in a wider range of environments, including diverse building structures and environments with varying levels of noise and interference. This will provide a comprehensive understanding of the performance of the information fusion strategy in different indoor localization scenarios, allowing for its optimization and improvement.

- Exploring Real-World Applications of the Information Fusion Strategy: The information fusion strategy presented in this study has been tested in a controlled environment, but future research can aim to explore its practicality and effectiveness in real-world applications. This could include smart contact tracing or context-aware services, where the ability to accurately track and locate individuals is crucial. Integrating the information fusion approach into the proposed TB-ICT framework can further improve the accuracy of the localization phase and enhance the reliability of the entire application. This can provide valuable insights into the feasibility and performance of this information fusion strategy in practical settings.

- Further Integration of RL in Indoor Localization: Future research can focus on further integrating RL in indoor localization by incorporating new techniques and methods to enhance the accuracy and performance of the system. This could include the integration of different types of sensors, such as LiDAR or RADAR, and utilizing advanced algorithms, such as advanced deep RL methods, to improve the overall performance of the system.

- Multi-Agent Reinforcement Learning: Future research can focus on advancing MARL techniques, such as the multi-agent SR framework proposed in this thesis, to enhance the accuracy of indoor tracking and localization scenarios. This could include incorporating new techniques for agent coordination, such as game theory, and exploring new ways to address challenges such as non-stationarity and partial observability in MARL problems.

- Information Fusion Strategies: The development of advanced information fusion strategies, such as the RL-based information fusion framework (RL-IFF) proposed in this thesis, can be further explored and improved. This can be achieved

by incorporating new techniques for fusing different sources of information, such as probabilistic graphical models, and exploring new ways to address challenges such as data integration and uncertainty quantification.

Bibliography

[1] M. Salimibeni, Z. Hajiakhondi-Meybodi, A. Mohammadi and Y. Wang, "TB-ICT: A Trustworthy Blockchain-Enabled System for Indoor Contact Tracing in Epidemic Control," *IEEE Internet of Things Journal, 2022*, doi: 10.1109/JIOT.2022.3223329. In Press.

[2] M. Salimibeni, P. Malekzadeh, A. Mohammadi and K.N. Plataniotis, "MAAKF-SR: Multi-Agent Adaptive Kalman Filtering-based Successor Representation," *Sensors,* vol. 22, no. 4, 2022.

[3] M. Salimibeni, A. Mohammadi, Z. Hajiakhondi, M. Atashi, P. Malekzadeh, and K.N. Plataniotis, "Reinforcement Learning-based Information Fusion for Multiple Model BLE-based Indoor Localization/Tracking," *Submitted to APSIPA Trans. on Signal and Information Processing (ATSIP)*, 2023.

[4] M. Salimibeni, A. Mohammadi, N. Plataniotis, "RL-IFF: Reinforcement Learning based Information Fusion for Indoor Localization"," *Submitted to IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[5] M. Salimibeni, Z. HajiAkhondi-Meybodi, A. Mohammadi, "TB-ICT: A Trustworthy Blockchain-Enabled System for Indoor Contact Tracing in Epidemic Control," *IEEE 8th World Forum on Internet of Things (WF-IoT)*, 2022.

[6] M. Salimibeni, P. Malekzadeh, A. Mohammadi, A. Assa and K. N. Plataniotis, "MAKF-SR: Multi- Agent Adaptive Kalman Filtering-based Successor Representations," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8037-8041.

[7] M. Salimibeni, M. Atashi, P. Malekzadeh, Z. HajiAkhondi-Meybodi, K. N. Plataniotis, A. Mohammadi, "IoT-TD: IoT Dataset for Multiple Model BLE-based IndoorLocalization/Tracking," *28th European Signal Processing Conference (EUSIPCO)*, 2020, pp. 1697-1701.

[8] M. Salimibeni, P. Malekzadeh, A. Mohammadi and K. N. Plataniotis, "Distributed Hybrid Kalman Temporal Differences for Reinforcement Learning," *IEEE International Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 579-583.

[9] M. Salimibeni, P. Malekzadeh, M. Atashi, M. Barbulescu, K. N. Plataniotis and A. Mohammadi, "Event-Triggered Monitoring/Communication of Inertial Measurement Unit for IoT Applications," *IEEE SENSORS*, 2019, pp. 1-4.

[10] Y. Li, X. Hu, Y. Zhuang, Z. Gao, P. Zhang and N. El-Sheimy, "Deep Reinforcement Learning (DRL): Another Perspective for Unsupervised Wireless Localization," *IEEE Internet of Things Journal*, invol. 7, no. 7, pp. 6279-6287, July 2020.

[11] P. Silva, V. Kaseva, E. Lohan, "Wireless Positioning in IoT: A Look at Current & Future Trends," *Sensors*, vol. 8, no. 8, 2018.

[12] P. Spachos and K. N. Plataniotis, "BLE Beacons for Indoor Positioning at an Interactive IoT-Based Smart Museum," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3483-3493, Sept. 2020.

[13] P. C. Ng, P. Spachos and K. N. Plataniotis, "COVID-19 and Your Smartphone: BLE-Based Smart Contact Tracing," *IEEE Systems Journal*, doi: 10.1109/JSYST.2021.3055675., 2021.

[14] F. Zafari, I. Papapanagiotou, M. Devetsikiotis, T.J. Hacker, "An iBeacon based Proximity and Indoor Localization System," https://arxiv.org/abs/1703.07876, 2017.

[15] T. T. Dinh, N. Duong and K. Sandrasegaran, "Smartphone-Based Indoor Positioning Using BLE iBeacon and Reliable Lightweight Fingerprint Map," in IEEE Sensors Journal, vol. 20, no. 17, pp. 10283-10294, 1 Sept.1, 2020, doi: 10.1109/JSEN.2020.2989411.

[16] K. Zheng, *et al.*, "Energy-Efficient Localization and Tracking of Mobile Devices in Wireless Sensor Networks," *IEEE Transactions on Vehicular Technology*, 2017.

[17] P. Davidson and R. Piche, "A Survey of Selected Indoor Positioning Methods for Smartphones," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1347-1370, 2017.

[18] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp and K. Reddy, "Recurrent Neural Networks for Accurate RSSI Indoor Localization," in IEEE Internet of Things Journal, vol. 6, no. 6, pp. 10639-10651, Dec. 2019, doi: 10.1109/JIOT.2019.2940368.

[19] P. S. Farahsari, A. Farahzadi, J. Rezazadeh and A. Bagheri, "A Survey on Indoor Positioning Systems for IoT-Based Applications," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7680-7699, 15 May15, 2022.

[20] P. Malekzadeh, A. Mohammadi, M. Barbulescu, and K.N. Plataniotis, "STUPEFY: Set-Valued Box Particle Filtering for BLE-based Indoor Localization" *IEEE Signal Processing Letters*, 2019. In Press.

[21] M. Atashi, M. Salimibeni, P. Malekzadeh, M. Barbulescu, K. N. Plataniotis and A. Mohammadi, "Multiple Model BLE-based Tracking via Validation of RSSI Fluctuations under Different Conditions," *International Conference on Information Fusion (FUSION)*, Ottawa, ON, Canada, 2019, pp. 1-6.

[22] D. An and J. Lee, "Derivation of an Approximate Location Estimate in Angle-of-Arrival Based Localization in the Presence of Angle-of-Arrival Estimate Error and Sensor Location Error," *IEEE World Symposium on Communication Engineering (WSCE)*, 2018, pp. 1-5.

[23] V. Sark and E. Grass, "Modified Equivalent Time Sampling for Improving Precision of Time-of-Flight based Localization," *IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 370-374.

[24] G. Fokin, A. Kireev and A. H. A. Al-odhari, "TDOA Positioning Accuracy Performance Evaluation for ARC Sensor Configuration," *Systems of Signals Generating and Processing in the Field of on Board Communications*, 2018, pp. 1-5.

[25] S. Sadowski and P. Spachos, "Optimization of BLE Beacon Density for RSSI-Based Indoor Localization," 2019 IEEE *International Conference on Communications Workshops (ICC Workshops)*, Shanghai, China, 2019, pp. 1-6.

[26] S. Sadowski and P. Spachos, "RSSI-Based Indoor Localization With the Internet of Things," *IEEE Access*, vol. 6, pp. 30149-30161, 2018.

[27] S. Kumar, R. Ramaswami, and K. Tomar, "Localization in Wireless Sensor Networks using Directionally Information," *IEEE International Advance Computing Conference (IACC)*, 2013, pp. 577-582.

[28] P. Malekzadeh, M. Salimibeni, A. Mohammadi, A. Assa and K. N. Plataniotis, "MM-KTD: Multiple Model Kalman Temporal Differences for Reinforcement Learning," *IEEE Access*, vol. 8, pp. 128716-128729, 2020.

[29] A. Kushki, K.N. Plataniotis and A.N. Venetsanopoulos "WLAN Positioning Systems: Principles and Applications in Location-based Services," *Cambridge University Press*, 2012.

[30] C. Ranasinghe and C. Kray, "Location Information Quality: A Review," *Sensors (Basel, Switzerland)*, 2018, 18(11), 3999.

[31] P. Jeongyeup, K. JeongGil, and S. Hyungsik, "A Measurement Study of BLE iBeacon and Geometric Adjustment Scheme for Indoor Location-Based Mobile Applications," *Mobile Information Systems*, vol. 2016, Article ID 8367638, 13 pages.

[32] M.C. Cara, J.L. Melgarejo, G.B. Rocca, L.O. Barbosa and I.G. Varea, "An analysis of multiple criteria and setups for Bluetooth smartphone-based indoor localization mechanism," *Journal of Sensors*, 2017.

[33] YB. Bai, *et al.*, "A new method for improving Wi-Fi-based indoor positioning accuracy," *J. Location-based Services*, vol. 8, no. 3, pp. 135-147, 2014.

[34] S. Y. Cho and C. G. Park, "Threshold-less Zero-Velocity Detection Algorithm for Pedestrian Dead Reckoning," *European Navigation Conference (ENC)*, Warsaw, Poland, 2019, pp. 1-5.

[35] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 38-46, Nov.-Dec. 2005.

[36] R. Harle, "A Survey of Indoor Inertial Positioning Systems for Pedestrians," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1281-1293, Third Quarter 2013.

[37] W. Kang and Y. Han, "SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2906-2916, May 2015.

[38] S. Monfared, T. Nguyen, L. Petrillo, P. De Doncker, and F. Horlin, "Experimental Demonstration of BLE Transmitter Positioning Based on AOA Estimation," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Bologna, Dec. 2018, pp. 856-859.

[39] M. Cominelli, P Patras, and F Gringoli, "Dead on Arrival: An Empirical Study of The Bluetooth 5.1 Positioning System," *International Workshop on Wireless Network Test beds, Experimental Evaluation and Characterization*, pp. 13-20. Oct. 2019.

[40] P. Hu, Q. Bao and Z. Chen, "Target Detection and Localization Using Non-Cooperative Frequency Agile Phased Array Radar Illuminator," *IEEE Access*, vol. 7, pp. 111277-111286, Aug. 2019.

[41] J. Zhou, H. Zhang and L. Mo, "Two-dimension localization of passive RFID tags using AOA estimation," *IEEE International Instrumentation and Measurement Technology Conference*, May 2011, pp. 1-5.

[42] X. Zhang, W. Chen, W. Zheng, Z. Xia and Y. Wang, "Localization of Near-Field Sources: A Reduced-Dimension MUSIC Algorithm," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1422-1425, Jul. 2018.

[43] M. D. Hossain and A. S. Mohan, "Eigenspace Time-Reversal Robust Capon Beamforming for Target Localization in Continuous Random Media," *EEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 1605-1608, Jan. 2017.

[44] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-End Learning From Spectrum Data: A Deep Learning Approach for Wireless Signal Identification in Spectrum Monitoring Applications," *IEEE Access*, vol. 6, pp. 18484-18501, Mar. 2018.

[45] Z. HajiAkhondi-Meybodi, M. S. Beni, K. N. Plataniotis, and A. Mohammadi "Bluetooth Low Energy-based Angle of Arrival Estimation via Switch Antenna Array for Indoor Localization," *International Conference on Information Fusion*, July 2020.

[46] Z. HajiAkhondi-Meybodi, M. S. Beni, A. Mohammadi, and K. N. Plataniotis, "Bluetooth Low Energy-based Angle of Arrival Estimation in Presence of Rayleigh Fading," Accepted in *IEEE International Conference on Systems, Man, and Sybernetics*, 2020.

[47] Z. HajiAkhondi-Meybodi, M. S. Beni, A. Mohammadi, and K. N. Plataniotis, "Bluetooth Low Energy-based Angle of Arrival Estimation in Presence of Rayleigh Fading," *IEEE International Conference on Systems, Man, and Sybernetics*, 2020.

[48] K. Wu, W. Ni, T. Su, R. P. Liu and Y. J. Guo, "Expeditious Estimation of Angle-of-Arrival for Hybrid Butler Matrix Arrays," *IEEE Transactions on Wireless Communications* vol. 18, no. 4, pp. 2170-2185, Apr. 2019.

[49] N. H. Nguyen, and K. Dogancay, "Closed-Form Algebraic Solutions for Angle-of-Arrival Source Localization With Bayesian Priors," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, May 2019, pp. 3827–3842.

[50] L. Cheng, Y. Wu, J. Zhang and L. Liu, "Subspace Identification for DOA Estimation in Massive/Full-Dimension MIMO Systems: Bad Data Mitigation and Automatic Source Enumeration," *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 5897-5909, Nov. 2015.

[51] R. Shafin, L. Liu, J. Zhang and Y. Wu, "DoA Estimation and Capacity Analysis for 3-D Millimeter Wave Massive-MIMO/FD-MIMO OFDM Systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 6963-6978, Oct. 2016.

[52] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, Ch. Yuen, R. Raulefs, and E. Aboutanios, "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1327-1346, Secondquarter 2017.

[53] J. He, Y. Geng, F. Liu and C. Xu, "CC-KF: Enhanced TOA Performance in Multipath and NLOS Indoor Extreme Environment," *EEE Sensors Journal*, vol. 14, no. 11, pp. 3766-3774, Nov. 2014.

[54] C. Zhang, X. Bao, Q. Wei, Q. Ma, Y. Yang, and Q. Wang, "A Kalman filter for UWB positioning in LOS/NLOS scenarios," *in Proc. International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services*, Nov. 2016, pp. 73-78.

[55] R. Exel and T. Bigler, "ToA ranging using subsample peak estimation and equalizer-based multipath reduction," *in Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, Apr. 2014, pp. 2964-2969.

[56] X. Wang, X. Wang, and S. Mao. "CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi," *IEEE International Conference on Communications (ICC)*, pp. 1-6, May 2017.

[57] X. Wang, X. Wang and S. Mao, "Deep Convolutional Neural Networks for Indoor Localization with CSI Images," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 316-327, Mar. 2020.

[58] M. Comiter, and H. T. Kung, "Localization Convolutional Neural Networks Using Angle of Arrival Images," *in Proc. IEEE Global Communications Conference*, Abu Dhabi, Dec. 2018, pp. 1-7.

[59] A. Khan, S. Wang and Z. Zhu, "Angle-of-Arrival Estimation Using an Adaptive Machine Learning Framework," *IEEE Communications Letters*, vol. 23, no. 2, pp. 294-297, Feb. 2019.

[60] C. Hsieh, J. Chen and B. Nien, "Deep Learning-Based Indoor Localization Using Received Signal Strength and Channel State Information," *IEEE Access*, vol. 7, pp. 33256-33267, 2019.

[61] M. A. Ferrag, L. Shu and K. -K. R. Choo, "Fighting COVID-19 and Future Pandemics With the Internet of Things: Security and Privacy Perspectives," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 9, pp. 1477-1499, September 2021.

[62] J. Bay, J. Kek, A. Tan, C. S. Hau, L. Yongquan, and J. Tan, T. A. Quy, "BlueTrace: A privacy-preserving protocol for communitydriven contact tracing across borders," *Singapores Government Technology Agency*, Singapore, White Paper, p. 9, 2020.

[63] "Harvard College. Surveys, app. to track COVID-19," Available: https://www.hsph.harvard.edu/coronavirus/covid-19-response-public-health-in-action/surveys-apps-to-track-covid-19/, Acc. on: Dec. 27, 2020.

[64] "Exposure Notification," Apple Inc., Cupertino, CA, USA and Google LLC., Mountain View, CA, USA, May 2020.

[65] I. Levy, "The Security Behind the NHS Contact Tracing App," Accessed: May 8, 2020. [Online]. Available: https://www.ncsc.gov.uk/blog-post/security-behind-nhs-contact-tracing-app.

[66] P. Mozur, R. Zhong, and A. Krolik, "In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags," New York, NY, USA, 2020. [Online]. Available: https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html.

[67] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S.Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha, "A Survey of COVID-19 Contact Tracing Apps," *IEEE Access*, vol. 8, pp. 134 577–134 601, 2020.

[68] L. Zhu, X. Tang, M. Shen, F. Gao, J. Zhang and X. Du, "Privacy-Preserving ML Training in IoT Aggregation Scenarios," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12106-12118, Aug, 2021.

[69] S. Vaudenay, "Centralized or decentralized? the contact tracing dilemma," *TIACR Cryptol*, ePrint Arch., vol. 2020, p. 531, 2020.

[70] W. Beskorovajnov, F. Dörre, G. Hartung, *et al.*, "Contra corona: Contact tracing against the coronavirus by bridging the centralized–decentralized divide for stronger privacy," *Crypt. ePrint Archive*, Report 2020/505, 2020.

[71] A. Islam, A. Al Amin and S. Y. Shin, "FBI: A Federated Learning-Based Blockchain-Embedded Data Accumulation Scheme Using Drones for Internet of Things," *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 972-976, May 2022.

[72] A. Islam, S.Y. Shin, "A Blockchain-based Secure Healthcare Scheme with the Assistance of Unmanned Aerial Vehicle in Internet of Things," *Computers & Electrical Engineering*, vol. 84, 2020.

[73] A. Islam, T. Rahim, M. Masuduzzaman and S. Y. Shin, "A Blockchain-Based Artificial Intelligence-Empowered Contagious Pandemic Situation Supervision Scheme Using Internet of Drone Things," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 166-173, August 2021.

[74] Naren, A. Tahiliani, V. Hassija, V. Chamola, S. S. Kanhere and M. Guizani, "Privacy-Preserving and Incentivized Contact Tracing for COVID-19 Using Blockchain," *IEEE Internet of Things Magazine*, vol. 4, no. 3, pp. 72-79, September 2021.

[75] K. R. Choo, Z. Yan, W. Meng, "Editorial: Blockchain in Industrial IoT Applications: Security and Privacy Advances, Challenges, and Opportunities," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4119-4121, 2020.

[76] U. Javaid and B. Sikdar, "A Checkpoint Enabled Scalable Blockchain Architecture for Industrial Internet of Things," *IEEE Trans. Ind. Informat.*, Oct. 2020.

[77] D. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, "Blockchain and ai-based solutions to combat coronavirus (covid19)-like epidemics: A survey," Preprints 2020, 2020040325.

[78] L. Zhang, T. Zhang, Q. Wu, Y. Mu and F. Rezaeibagha, "Secure Decentralized Attribute-Based Sharing of Personal Health Records with Blockchain," *IEEE Internet of Things Journal*, 2021.

[79] H. -N. Dai, Z. Zheng and Y. Zhang, "Blockchain for Internet of Things: A Survey," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8076-8094, Oct. 2019.

[80] P. Wang, C. Lin, M. S. Obaidat, Z. Yu, Z. Wei and Q. Zhang, "Contact Tracing Incentive for COVID-19 and Other Pandemic Diseases From a Crowdsourcing Perspective," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15863-15874, 1 Nov.1, 2021.

[81] D. Wang, X. Chen, L. Zhang, Y. Fang and C. Huang, "A Blockchain based Human-to-Infrastructure Contact Tracing Approach for COVID-19," *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 12836-12847, 15 July15, 2022.

[82] G. Garofalo, T. Van hamme, D. Preuveneers, W. Joosen, A. Abidin and M. A. Mustafa, "PIVOT: PrIVate and effective cOntact Tracing," *IEEE Internet of Things Journal*, 2021.

[83] M. A. Azad, *et al.*, "A First Look at Privacy Analysis of COVID-19 Contact-Tracing Mobile Applications," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15796-15806, 1 Nov.1, 2021.

[84] J. Song, T. Gu, X. Feng, Y. Ge, and P. Mohapatra, "Blockchain Meets COVID-19: A Framework for Contact Information Sharing and Risk Notification System," *IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems*, 2021, pp. 269-277.

[85] M. A. Ferrag and L. Shu, "The Performance Evaluation of Blockchain-Based Security and Privacy Systems for the Internet of Things: A Tutorial," *IEEE Int. of Things J.*, vol. 8, no. 24, pp. 17236-17260, 2021.

[86] M. Martinez, A. Hekmati, B. Krishnamachari and S. Yun, "Mobile Encounter-based Social Sybil Control," *Int. Con. Software Defined Systems (SDS)*, 2020, pp. 190-195.

[87] H. Xu, *et al.*, "Beeptrace: Blockchain-Enabled Privacy-Preserving Contact Tracing for Covid-19 Pandemic and Beyond," *IEEE Internet Things J.*, pp. 1-1, Sep. 2020.

[88] H. R. Hasan, K. Salah, R. Jayaraman, I. Yaqoob, M. Omar and S. Ellahham, "COVID-19 Contact Tracing Using Blockchain," *IEEE Access*, vol. 9, pp. 62956-62971, 2021.

[89] S. Micali, "Algorand's approach to Covid-19 tracing," Tech. Rep., 2020.

[90] W. Lv, S. Wu, C. Jiang, Y. Cui, X. Qiu, and Y. Zhang, "Towards Large-Scale and Privacy-Preserving Contact Tracing in COVID-19 Pandemic: A Blockchain Perspective," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 282-298, 2022.

[91] G. Avitabile, V. Botta, V. Iovino, and I. Visconti, "Towards defeating mass surveillance and SARS-CoV-2: The pronto-C2 fully decentralized automatic contact tracing system," IACR Cryptol. ePrint Arch., vol. 2020, p. 493, May 2020.

[92] N. Ahmed, R. A. Michelin, W. Xue, G. Dharma Putra, S. Ruj, S. S. Kanhere, and S. Jha, "DIMY: Enabling privacy-preserving contact tracing," 2021, arXiv:2103.05873. [Online]. Available: http:// arxiv.org/abs/2103.05873.

[93] K. Peng, M. Li, H. Huang, C. Wang, S. Wan and K. -K. R. Choo, "Security Challenges and Opportunities for Smart Contracts in Internet of Things: A Survey," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12004-12020, 2021.

[94] P. Esteves-Verissimo, J. Decouchant, M. Völp, A. Esfahani, and R. Graczyk, "PriLok: Citizen-protecting distributed epidemic tracing," 2020, arXiv:2005.04519. [Online]. Available: http://arxiv.org/abs/2005.04519.

[95] T. Yanagihara and A. Fujihara, "Cross-Referencing Method for Scalable Public Blockchain," *Internet of Things Journal*, Volume 15, 100419, 2022.

[96] G. Jing, H. Bai, J. George, A. Chakrabortty, Model-free optimal control of linear multi-agent systems via decomposition and hierarchical approximation. *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 3, pp. 1069-1081, 2021.

[97] M. Turchetta, A. Krause, S. Trimpe, "Robust Model-free Reinforcement Learning with Multi-objective Bayesian Optimization," In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August, 2020.

[98] Q. Liu, T. Yu, Y. Bai, C. Jin, "A sharp analysis of model-based reinforcement learning with self-play," *arXiv*, 2020, arXiv:2010.01604.

[99] R. Bellman, "The Theory of Dynamic Programming," Tech. Rep., RAND Corp: Santa Monica, CA, USA, 1954.

[100] E. Vértes,and M. Sahani, "A Neurally Plausible Model Learns Successor Representations in Partially Observable Environments," *Advances in Neural Information Processing Systems 32*, pp.13714-13724, 2019.

[101] S. Blakeman, and D. Mareschal, "A Complementary Learning Systems Approach to Temporal Difference Learning," *Neural Networks*, vol. 122, 2020, pp. 218-230.

[102] Y. Song, W. Sun, "Pc-mlp: Model-based reinforcement learning with policy cover guided exploration," In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021.

[103] J. P.Geerts, K. L. Stachenfeld, N. Burgess, "Probabilistic Successor Representations with Kalman Temporal Differences," *arXiv*, 2019, arXiv:1910.02532.

[104] M. C. Machado, A. Barreto, D. Precup, "Temporal Abstraction in Reinforcement Learning with the Successor Representation," *arXiv*,2021, arXiv:2110.05740.

[105] T. H. Moskovitz, J. Parker-Holder, A. Pacchiano, M.Arbel, M. I. Jordan, "Tactical optimism and pessimism for deep reinforcement learning," In Proceedings of the NeurIPS, Virtual, 6-14 December 2021.

[106] H. Van Hasselt, A. Guez, D.Silver, "Deep Reinforcement Learning with Double Q-Learning," AAAI: Phoenix, AZ, USA, 2016, p. 5.

[107] G. S. Babu, S. Suresh, "Meta-cognitive neural network for classification problems in a sequential learning framework," *Neurocomputing*,2012,81, 86–96.

[108] M. Riedmiller, "Neural Fitted Q Iteration-first Experiences with a Data Efficient Neural Reinforcement Learning Method," *European Conference on Machine Learning*; Springer: Berlin, Heidelberg, 2005, pp. 317–328.

[109] Y. Tang, H. Guo,T. Yuan, X. Gao, X. Hong, Y. Li, J. Qiu, Y. Zuo, and J. Wu, "Flow Splitter: A Deep Reinforcement Learning-Based Flow Scheduler for Hybrid Optical-Electrical Data Center Network," *IEEE Access*, vol. 7, pp.129955-129965, 2019.

[110] M. Kim, S. Lee, J. Lim, J. Choi, and S.G. Kang, "Unexpected Collision Avoidance Driving Strategy Using Deep Reinforcement Learning," *IEEE Access*, vol. 8, pp. 17243-17252, 2020.

[111] J. Xie, Z. Shao, Y. Li, Y. Guan, and J. Tan, "Deep reinforcement learning with optimized reward functions for robotic trajectory planning," *IEEE Access*, vol. 7, pp. 105669-105679, 2019.

[112] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M.Riedmiller, "Playing Atari with Deep Reinforcement Learning," Technical Report, Deepmind Technologies: London, 2013, arXiv:1312.5602[cs.LG].

[113] T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv*, 2015, arXiv:1509.02971.

[114] J.N. Tsitsiklis, B. Van Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Transactions on Automatic Control*, 42 (5) (May 1997) 674–690.

[115] D.P. Bertsekas, V.S. Borkar, A. Nedic, "Improved temporal difference methods with linear function approximation, Learning and Approximate Dynamic Programming," 2004, pp. 231–255.

[116] W. T. Miller, F. H. Glanz, and L. G. Kraft, "Cmas: An Associative Neural Network Alternative to Backpropagation," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1561-1567, 1990.

[117] S. Haykin, "Neural Networks: A Comprehensive Foundation," *Prentice Hall PTR*, 1994.

[118] A. d. M. S. Barreto and C. W. Anderson, "Restricted Gradient-descent Algorithm for Value-function Approximation in Reinforcement Learning," *Artificial Intelligence*, vol. 172, no. 4-5, pp. 454-482, 2008.

[119] I. Menache, S. Mannor, and N. Shimkin, "Basis Function Adaptation in Temporal Difference Reinforcement Learning," *Annals of Operations Research*, vol. 134, no. 1, pp. 215-238, 2005.

[120] D. Choi and B. Van Roy, "A generalized Kalman filter for Fixed Point Approximation and Efficient Temporal-difference Learning," *Discrete Event Dynamic Systems*, vol. 16, no. 2, pp. 207-239, 2006.

[121] Y. Engel, "Algorithms and Representations for Reinforcement Learning," *Hebrew University of Jerusalem*, 2005.

[122] S. J. Bradtke and A. G. Barto, "Linear Least-squares Algorithms for Temporal Difference Learning," *Machine Learning*, vol. 22, no. 1-3, pp. 33-57, 1996.

[123] M. Geist and O. Pietquin, "Algorithmic Survey of Parametric Value Function Approximation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 6, pp. 845-867, 2013.

[124] M. Geist and O. Pietquin, "Kalman Temporal Differences," *Journal of Artificial Intelligence Research*, vol. 39, pp. 483-532, 2010.

[125] A. Mohammadi and K. N. Plataniotis, "Distributed Widely Linear Multiple-Model Adaptive Estimation," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 3, pp. 164-179, 2015.

[126] C. Yang, A. Mohammadi, Q-W. Chen, " Multi-Sensor Fusion with Interaction Multiple Model and Chi-Square Test Tolerant Filter," *Sensors*, vol. 16, no. 11, 1835, 2016.

[127] A. Mohammadi and K. N. Plataniotis, "Improper Complex-Valued Multiple-Model Adaptive Estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1528-1542, 2015.

[128] R. Mehra, "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 175-184, 1970.

[129] A. Assa and K. N. Plataniotis, "Similarity-based Multiple Model Adaptive Estimation," *IEEE Access*, vol. 6, pp. 36 632-36 644, 2018.

[130] T. Kitao, M. Shirai, and T. Miura, "Model Selection based on Kalman Temporal Differences Learning," *IEEE International Conference on Collaboration and Internet Computing (CIC)*, 2017, pp. 41-47.

[131] C. Ma, J. Wen, and Y. Bengio, "Universal successor representations for transfer reinforcement learning," *arXiv preprint*, arXiv:1804.03758, 2018.

[132] I. Momennejad, E.M. Russek, J.H. Cheong, M.M. Botvinick, N.D. Daw, S.J.Gershman, "The successor representation in human reinforcement learning," *Nature Human Behaviour*, 1 (9) (2017) 680–692.

[133] E.M. Russek, I. Momennejad, M.M. Botvinick, S.J. Gershman, N.D. Daw, "Predictive representations can link model-based reinforcement learning to model-free mechanisms," *PLoS Computational Biology*, 13 (9) (2017) e1005768.

[134] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," *Cambridge*, MA, USA: MIT Press, 2018.

[135] S.C. Chan, S. Fishman, J. Canny, et al., "Measuring the reliability of reinforcement learning algorithms," *International Conference on Learning Representations*, 2020.

[136] P. Malekzadeh, M. Salimibeni, M. Hou, A. Mohammadi, K. N. Plataniotis, "AKF-SR: Adaptive Kalman Filtering-Based Successor Representation," *Neurocomputing*, vol. 467, no. 7, pp. 476-490, January 2022.

[137] S. Spanò, G.C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, M. Matta, A. Nannarelli, and M. Re, "An Efficient Hardware Implementation of Reinforcement Learning: The Q-Learning," Algorithm," *IEEE Access*, vol. 7, pp. 186340-186351, 2019.

[138] M. Seo, L.F. Vecchietti, S. Lee, and D. Har, "Rewards Prediction-Based Credit Assignment for Reinforcement Learning With Sparse Binary Rewards," *IEEE Access*, vol. 7, pp. 118776-118791, 2019.

[139] A. Toubman *et al.*, "Modeling behavior of Computer Generated Forces with Machine Learning Techniques, the NATO Task Group approach," *IEEE Int. Con. Systems, Man, and Cyb. (SMC)*, Budapest, 2016, pp. 001906-001911.

[140] J. J. Roessingh *et al.*, "Machine Learning Techniques for Autonomous Agents in Military Simulations - Multum in Parvo," *IEEE Int. Con. Systems, Man, and Cyb. (SMC)*, Banff, AB, 2017, pp. 3445-3450.

[141] H. Hu, S. Song and C. L. P. Chen, "Plume Tracing via Model-Free Reinforcement Learning Method," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 8, pp. 2515-2527, Aug. 2019.

[142] H. K. Venkataraman, and P. J. Seiler, "Recovering Robustness in Model-Free Reinforcement Learning," *American Control Conference (ACC)*, Philadelphia, PA, USA, 2019, pp. 4210-4216.

[143] G. Williams, N. Wagener, B. Goldfain, P. Drews, J. M. Rehg, B. Boots, and E. A. Theodorou, "Information Theoretic MPC for Model-based Reinforcement Learning," *International Conference on Robotics and Automation (ICRA)*, 2017.

[144] R. Bellman, "The Theory of Dynamic Programming," *RAND Corp Santa Monica CA*, Tech. Rep., 1954.

[145] A. Ducarouge, O. Sigaud, "The Successor Representation as a Model of Behavioural Flexibility," *Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JFPDA)*, 2017.

[146] A. Lazaric, M. Restelli, and A. Bonarini, ""Reinforcement Learning in Continuous Action Spaces Through Sequential Monte Carlo Methods," *Advances in Neural Information Processing Systems*, 2008, pp. 833-840.

[147] D. D. Castro, D. Volkinshtein and R. Meir, "Temporal Difference Based Actor Critic Learning - Convergence and Neural Implementation," *Neural Information Processing Systems Conference*, 2008.

[148] E. Keogh and A. Mueen, "Curse of Dimensionality," *Encyclopedia of Machine Learning*, Springer, 2011, pp. 257-258.

[149] Y. Ge, F. Zhu, X. Ling, and Q. Liu, "Safe Q-Learning Method Based on Constrained Markov Decision Processes," *IEEE Access*, vol. 7, pp. 165007-165017, 2019.

[150] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," *arXiv:1312.5602*, 2013.

[151] S. Sukhbaatar, I. Kostrikov, A. Szlam, and R. Fergus, "Intrinsic motivation and automatic curricula via asymmetric self-play," *arXiv*, preprint arXiv:1703.05407, 2017.

[152] S. Haykin, "Neural Networks: A Comprehensive Foundation," *Prentice Hall PTR*, 1994.

[153] R. M. Kretchmar and C. W. Anderson, "Comparison of CMACs and Radial Basis Functions for Local Function Approximators in Reinforcement Learning," *International Conference on Neural Networks*, vol. 2. IEEE, 1997, pp. 834-837.

[154] G. Konidaris, S. Osentoski, and P. S. Thomas, "Value Function Approximation in Reinforcement Learning using the Fourier Basis," *AAAI*, vol. 6, 2011, p. 7.

[155] I. Menache, S. Mannor, and N. Shimkin, "Basis Function Adaptation in Temporal Difference Reinforcement Learning," *Annals of Operations Research*, vol. 134, no. 1, pp. 215-238, 2005.

[156] K. Doya, K. Samejima, K.-i. Katagiri, and M. Kawato, "Multiple Model-based Reinforcement Learning," *Neural Computation*, 14(6), 1347–1369, 2002.

[157] D. G. Lainiotis, "Partitioning: A Unifying Framework for Adaptive Systems, i: Estimation," *Proceedings of the IEEE*, vol. 64, no. 8, pp. 1126-1143, 1976.

[158] B.Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, Volume 14, Issue 1, July 2013.

[159] C. Federico, "A review of data fusion techniques," *Sci. World J.*, 1–19, 2013.

[160] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Information Fusion*, Volume 57, 2020.

[161] S. Chen, J. Wang, H. Li, Z. Wang, F. Liu and S. Li, "Top-Down Human-Cyber-Physical Data Fusion Based on Reinforcement Learning," *IEEE Access*, vol. 8, pp. 134233-134245, 2020.

[162] X. Liu, C. Sun, M. Zhou, C. Wu, B. Peng and P. Li, "Reinforcement Learning-Based Multislot Double-Threshold Spectrum Sensing With Bayesian Fusion for Industrial Big Spectrum Data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3391-3400, May 2021.

[163] J. Guo, Q. Liu and E. Chen, "A Deep Reinforcement Learning Method For Multimodal Data Fusion in Action Recognition," *IEEE Signal Processing Letters*, vol. 29, pp. 120-124, 2022.

[164] T. Abirami, E. Taghavi, R. Tharmarasa, T. Kirubarajan, and A.-C. Boury-Brisset, "Fusing social network data with hard data," *Proc. 18th Int. Conf. Inf. Fusion (Fusion)*, 2015, pp. 652–658.

[165] F. Xiao, "A new divergence measure for belief functions in D–S evidence theory for multisensor data fusion," *Inf. Sci.,* vol. 514, pp. 462–483, Apr. 2020.

[166] C. Zhu, F. Xiao, Z. Cao, "A generalized Rényi divergence for multi-source information fusion with its application in EEG data analysis," *Inf. Sci.,* vol. 605, pp. 225-243, 2022.

[167] G. Zhao, A. Chen, G. Lu, and W. Liu, "Data fusion algorithm based on fuzzy sets and D-S theory of evidence," *Tsinghua Sci. Technol.,* vol. 25, no. 1, pp. 12–19, Feb. 2020.

[168] S. Surathong, C. Maisen and P. Piyawongwisal, "Modified Fuzzy Dempster-Shafer Theory for Decision Fusion," *13th Int. Conf. on Information Tech. and Elec. Eng. (ICITEE),* 2021, pp. 244-248.

[169] J. Li, and Q. Wang, "Multi-modal bioelectrical signal fusion analysis based on different acquisition devices and scene settings: Overview, challenges, and novel orientation," *Information Fusion,* Volume 79, 2022.

[170] P. Wang, L. T. Yang, J. Li, J. Chen, and S. Hu, "Data fusion in cyber-physical-social systems: State-of-the-art and perspectives," *Information Fusion,* vol. 51, Nov. 2019.

[171] Y. Himeur, B. Rimal, A. Tiwary, and A. Amira, "Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives," *Information Fusion*, Volume 86–87, 2022.

[172] P. Saha and S. Mukhopadhyay, "Multispectral Information Fusion With Reinforcement Learning for Object Tracking in IoT Edge Devices," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 4333-4344, April, 2020.

[173] Y. Han et al., "Deep Reinforcement Learning for Robot Collision Avoidance With Self-State-Attention and Sensor Fusion," *IEEE Robotics and Automation Letters,* vol. 7, no. 3, pp. 6886-6893, July 2022.

[174] T. Zhou, M. Chen and J. Zou, "Reinforcement learning based data fusion method for multi-sensors," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 6, pp. 1489-1497, November 2020.

[175] P. Liu, F. Bao, X. Yao, C. Zhang, et al., "Multi-type data fusion framework based on deep reinforcement learning for algorithmic trading," *Applied Intelligence,* 2022.

[176] J. Yu, P. Wang, T. Koike-Akino and P. V. Orlik, "Multi-Modal Recurrent Fusion for Indoor Localization," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[177] S. -I. Sou, F. -J. Wu and W. -C. Wu, "JoLo: Multi-device Joint Localization based on Wireless Data Fusion," *IEEE Transactions on Mobile Computing*, 2022.

[178] T. -M. T. Dinh, N. -S. Duong and Q. -T. Nguyen, "Developing a Novel Real-Time Indoor Positioning System Based on BLE Beacons and Smartphone Sensors," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23055-23068, 15 Oct.15, 2021.

[179] M. Woolley, "Bluetooth Direction Finding: A Technical Overview," version 1.0, Revision Date: 20 March 2019.

[180] X. Zhong, A. Mohammadi, A.B. Premkumar, and A. Asif, "A Distributed Particle Filtering Approach for Multiple Acoustic Source Tracking using an Acoustic Vector Sensor Network," *Signal Processing*, vol. 108, pp. 589-603, March 2015.

[181] J. Fang, H. Sun, J. Cao, X. Zhang, Y. Tao, "A novel calibration method of magnetic compass based on ellipsoid fitting," *IEEE Trans. on Instrument. Meas.*, 60, 2053–2061, 2011.

[182] M. Hutter and S. Legg, "Temporal Difference Updating without a Learning Rate," *Advances in Neural Information Processing Systems*, 2008, pp. 705-712.

[183] R. S. Sutton, "Generalization in Reinforcement Learning: Successful Examples using Sparse Coarse Coding," *Advances in Neural Information Processing Systems*, 1996, pp. 1038-1044.

[184] W. Xia, C. Di, H. Guo, and S. Li, "Reinforcement Learning Based Stochastic Shortest Path Finding in Wireless Sensor Networks," *IEEE Access*, vol. 7, pp.157807-157817, 2019.

[185] C. J. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279-292, 1992.

[186] J. Li, T. Chai, F. L. Lewis, Z. Ding and Y. Jiang, "Off-Policy Interleaved Q -Learning: Optimal Control for Affine Nonlinear Discrete-Time Systems," *IEEE Transactions on Neural Networks and Learning Systems,*, vol. 30, no. 5, pp. 1308-1320, May 2019.

[187] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Annual Conference on Neural Information Processing Systems (NIPS)*, 2017.

[188] A. Singh, T. Jain, and S. Sukhbaatar, "Learning when to communicate at scale in multiagent cooperative and competitive tasks," *In ICLR*, 2019.

[189] A. Mohammadi and K. N. Plataniotis, "Event-Based Estimation With Information-Based Triggering and Adaptive Update," *IEEE Transactions on Signal Processing*, vol. 65, no. 18, pp. 4924-4939, 15 Sept. 2017.

[190] K. Zhang, Z. Yang, and T. Bacsar, "Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms," *, https://arxiv.org/abs/1911.10635*, springer, 2021.

[191] P. Hamadanian, M. Schwarzkopf, S. Sen and M. Alizadeh, "reinforcement learning in time-varying systems: an empirical study," *arXiv:2201.05560v1*, 2201.05560v1, 2022.

[192] I. Mordatch, P. Abbeel, "Emergence of grounded compositional language in multi-agent populations." *Proc. AAAI Conference of Artificial Intelligence*, 2018.

[193] Henderson, P.; Islam, R.; Bachman, P.; Pineau, J.; Precup, D.; Meger, D. Deep Reinforcement Learning that Matters. *arXiv* **2017**, arXiv:1709.06560.

[194] S.C. Chan, S. Fishman, J. Canny, et al., "Measuring the Reliability of Reinforcement Learning Algorithms," *International Conference on Learning Representations*, Addis Ababa, Ethiopia, 26–30 April, 2020.

[195] Z. HajiAkhondi-Meybodi, M. S. Beni, A. Mohammadi and K. N. Plataniotis, "Bluetooth Low Energy and CNN-Based Angle of Arrival Localization in Presence of Rayleigh Fading," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 7913-7917.

[196] S. R. Hussain, S. Mehnaz, S. Nirjon, and E. Bertino, "Secure seamless bluetooth low energy connection migration for unmodified iot devices," *IEEE Trans. Mob. Com.*, vol. 17, no. 4, pp. 927–944, 2018.

[197] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," Ethereum Project Yellow Paper, 2014.

[198] M. A. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras and H. Janicke, "Blockchain Technologies for the Internet of Things: Research Issues and Challenges," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2188-2204, 2019.

[199] M. Ganardi, D. Hucke, and M. Lohrey, "Randomized sliding window algorithms for regular languages," *45th International Colloquium on Automata, Languages, and Programming, ICALP*, 2018, Czech Republic, pages 127:1–127:13.

[200] Microsoft, "The Stride Threat Model. [Online]. Available: https://docs.microsoft.com/en-us/previous-versions/commerce-server/ee823878(v=cs.20), 2021.

[201] X. Huang, C. Xu, P. Wang, and H. Liu, "LNSC: A security model for electric vehicle and charging pile management based on blockchain ecosystem," *IEEE Access*, vol. 6, pp. 13565–13574, 2018.

[202] J. Wang, *et al.*, "A blockchain based privacy-preserving incentive mechanism in crowdsensing applications," *IEEE Access*, vol. 6, pp. 17545–17556, 2018.

[203] C. Lin, D. He, X. Huang, K.-K. R. Choo, and A. V. Vasilakos, "BSeIn: A blockchain-based secure mutual authentication with fine-grained access control system for industry 4.0," *J. Netw. Comput. Appl.*, vol. 116, pp. 42–52, Aug. 2018.

[204] L. Li, *et al.*, "CreditCoin: A Privacy-Preserving Blockchain-Based Incentive Announcement Network for Communications of Smart Vehicles," *IEEE Transactions on Intelligent Transportation Systems,*, vol. 19, no. 7, pp. 2204-2220, July 2018.

[205] S. Malik, V. Dedeoglu, S. S. Kanhere and R. Jurdak, "TrustChain: Trust Management in Blockchain and IoT Supported Supply Chains," *IEEE International Conference on Blockchain*, 2019, pp. 184-193.

[206] P. Ferrari, A. Flammini, E. Sisinni, S. Rinaldi, D. Brandão and M. S. Rocha, "Delay Estimation of Industrial IoT Applications Based on Messaging Protocols," *IEEE Internet of Things Journal*, vol. 67, no. 9, pp. 2188-2199, Sept. 2018.

[207] R. Pass, C. Tech, and L. Seeman, "Analysis of the Blockchain Protocol in Asynchronous Networks," *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Paris, France, May 2017, pp. 643-673.

[208] J. Garay, A. Kiayias, N. Leonardos, "The bitcoin backbone protocol: analysis and applications," *Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015. LNCS*, vol. 9057, pp. 281–310. Springer, Heidelberg 2015.

[209] K. Miyachi and T. K. Mackey, "hOCBS: A privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design Author links open overlay panel," *Information Processing & Management Journal*, vol. 58, no. 3, 2021, Art. no. 102535.

[210] Q. Zhou, H. Huang, Z. Zheng and J. Bian, "Solutions to Scalability of Blockchain: A Survey," *IEEE Access*, vol. 8, pp. 16440-16455, 2020.

[211] E. Lombrozo, J. Lau, and P. Wuille, "Segregated witness (consensuslayer)," *Bitcoin Core Develop. Team*, Tech. Rep., 2015.

[212] J. Poon and D. Thaddeus, "The bitcoin lightning network: Scalable off-chain instant payments," 2016.

[213] R. Norvill, B. B. Fiz Pontiveros, R. State and A. Cullen, "IPFS for Reduction of Chain Size in Ethereum," *IEEE Int. Conf. on Internet of Things (iThings)*, 2018, pp. 1121-1128.

[214] S. Delgado-Segura, C. Perez-Sola, G. Navarro-Arribas, and J. Herrera-Joancomarti, "Analysis of the Bitcoin UTXO Set," *Financial Cryptography and Data Security*, Springer Berlin Heidelberg, 2019, pp. 78–91.

[215] J. Stark, "Making sense of ethereum's layer 2 scaling solutions: state channels,Plasma, and truebit. Medium," February 2018, Available at: https://medium.com/l4-media/making-sense-of-ethereums-layer-2-scaling-solutions-state-channels-plasma-and-truebit-22cb40dcc2f4.

[216] H. Obeidat, W. Shuaieb, O. Obeidat, and R. Abd-Alhameed, "A Review of Indoor Localization Techniques and Wireless Technologies," *Wireless Personal Communications*, vol. 119, no. 1 pp. 289-327, 2021.

[217] S. Yousefi, X. Chang, and B. Champagne, "Mobile Localization in Non-Line-of-Sight Using Constrained Square-Root Unscented Kalman Filter," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 2071-2083, May 2015.

[218] Available online: https://www.increasebroadbandspeed.co.uk/what-is-a-good-signal-level-or-signal-to-noise-ratio-snr-for-wi-fi(acc. on 21-05-25).

[219] D. Vujicic, D. Jagodic, S. Randic, "Blockchain technology, bitcoin, and Ethereum: A brief overview," *Int. Symposium Inf.-Jah.*, 2018, pp. 1-6.