CHAPTER I
INTRODUCTION

1.1 Rationale

The development of ITS, especially in Urban Rail Signaling Systems such as CBTC has opened great opportunities to improve the operational reliability. As shown in Figure 1. ITS interconnection and operation systems open opportunities for the implementation of new technologies such as IoT, AI, Digital Twin, and T2T. This is expected to improve the operational efficiency of the urban train network [1], [2], as well as be able to strengthen cybersecurity and create an adaptive and sustainable system to be able to support the reliability of the railway system on the urban train network in the future [1], [3].

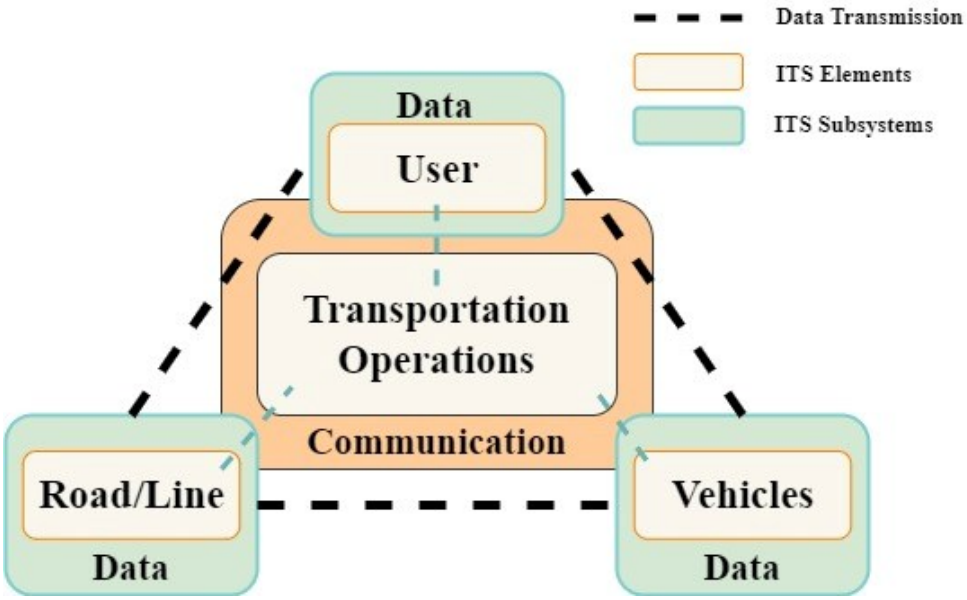

Figure 1. 1. 1. Interconnection and Operation on ITS.

As an important component of the Intelligent System in the Urban Railway Network System, CBTC plays a very important role in the Urban Rail System by having automatic train control and utilizing real-time communication data exchange, so that CBTC technology can optimize the capacity of the crosswalk and ensure more optimal, reliable and real-time travel operations. T2G systems in CBTC usually utilize IEEE 802.11 protocol Radio Frequency technology to support data exchange between infrastructure on trains and Wayside Equipment. The use of a reliable CBTC system can support the control and coordination of train operations with precision, so that it is able to meet the needs of urban train services that are safe, timely and efficient [4].

The development of innovations in the CBTC signaling system continues to bring various advancements, including integration with new technology features such as IoT, AI, and Digital Twin in urban railway systems. In addition, the development of T2T communication in the CBTC signaling system allows trains to communicate directly with each other, which can improve coordination skills and improve train operational safety through real-time response to misbehavior findings during operation. However, these technological advances can introduce cybersecurity vulnerabilities, posing new challenges for urban railway system operators, especially in CBTC signaling systems [5], [6].

The implementation of good security measures, including the security of communication channels and the protection of the integrity of the system is very important [7], [8]. In dealing with these threats, an effective and efficient cyber-attack detection mechanism is needed. Currently, various studies related to Ensemble Learning based misbehavior detection systems in the ITS system have been carried out [9]. The application of Ensemble Learning-based early detection systems has been proven to provide a high level of accuracy, allowing a system to respond quickly to system anomalies. This is important to maintain operational sustainability and ensure safety in the transportation system [10], [11].

This study proposes a Lightweight MDS for CBTC that focuses on classifying attack types and detecting misbehavior conditions during train operations. The system integrates CBTC data communication with a Machine Learning-based MDS, enabling the identification of misbehavior patterns in large datasets, including the detection of new misbehavior cases with high accuracy [12], [13]. One of the families in Machine Learning that has been proven for MDS is Ensemble Learning, which can be used for modelling in the detection and classification of cybersecurity attacks [14], [15]. By implementing this approach, the system is expected to enhance cybersecurity, safety, and the reliability of urban rail network infrastructure [4], [16], [17].

Key contributions from this study include:

i. Development of an MDS Algorithm for CBTC Data Communication: This research aims to design a lightweight and efficient misbehavior detection system specifically for data communication in CBTC. The study focuses on analyzing and modelling ensemble learning algorithms to achieve high

accuracy in classifying various types of anomalies encountered in the CBTC system.

ii. Detection Accuracy: By leveraging ensemble learning, the proposed MDS is expected to enhance detection accuracy and accelerate the misbehavior identification process. The system can automatically recognize misbehavior patterns, ensuring reliable cybersecurity measures against evolving threats.

iii. Model Performance Evaluation: The proposed detection system undergoes a comprehensive evaluation using key performance metrics such as accuracy, precision, recall, F1-Score, Processing Time, and Fit Status Model. This thorough assessment ensures the robustness and effectiveness of the model in addressing challenges within CBTC environments.

With the rapid advancement of ITS and their integration into urban rail networks, CBTC has significantly improved operational efficiency and safety. However, these advancements also expose the system to increased cybersecurity risks, particularly misbehavior in data communication, which can compromise train operations and passenger safety. To address these challenges, this study develops a lightweight MDS utilizing Ensemble learning algorithms, specifically Bagging (Random Forest and kNN with Bagging) and Boosting (AdaBoost, XGBoost, and LightGBM). By incorporating these techniques, the system enhances misbehavior detection accuracy while maintaining computational efficiency, ensuring a scalable, adaptive, and reliable solution for CBTC environments.

The remainder of this paper is structured as follows: Chapter 2 provides an overview of the CBTC signaling and security systems, along with a review of prior research on Ensemble Learning-based Misbehavior Detection Systems in ITS, highlighting existing research gaps and future challenges. Chapter 3 details the research methodology and framework adopted in this study. Chapter 4 presents the evaluation and performance analysis of the proposed system, while Chapter 5 concludes the study by summarizing key findings and offering recommendations based on the research results.

1.2 Theoretical Framework

This research focuses on Urban Rail Transportation Systems, Cybersecurity, and Ensemble Learning to address the challenges of detecting misbehavior in CBTC

signaling systems. The Urban Rail Transportation System Theory provides an in-depth understanding of CBTC components and operational mechanisms, ensuring a comprehensive analysis of their functionality. The Cybersecurity Framework underscores the importance of securing train communication channels, particularly as the reliance on wireless communication in urban rail systems increases exposure to cyber threats, which can compromise operational continuity and passenger safety. Given these vulnerabilities, an effective misbehavior detection system is crucial to maintaining the integrity and reliability of CBTC networks.

The Ensemble Learning approach in this study is designed to develop an adaptive and lightweight MDS. By utilizing Ensemble learning techniques such as Bagging (Random Forest and kNN with Bagging) and Boosting (AdaBoost, XGBoost, and LightGBM), the system enhances its ability to detect both known and emerging threats while optimizing processing time for real-time detection. The lightweight aspect of this approach is emphasized through efficient processing time, ensuring that model training, fine-tuning, and anomaly detection occur rapidly without imposing excessive computational burdens. Additionally, Ensemble learning modeling enables the system to process large volumes of data efficiently, improving detection accuracy and adaptability to dynamic conditions. By integrating transportation theory, cybersecurity, and Ensemble learning, this study provides a comprehensive and efficient framework for mitigating misbehavior in CBTC signaling systems, ultimately enhancing the safety, security, and efficiency of data communication in urban rail networks.

1.3 Conceptual Framework

The conceptual framework of this study centers on the development of a misbehavior detection system that leverages Ensemble Learning to analyze data from cyber-attack simulations previously conducted within the CBTC signaling system. This framework consists of three key components: Input, Process, and Output. The framework includes three main components:

1. Input: The input stage involves the utilization of essential data, including train position, train speed, and communication logs between train infrastructure and the wayside, as conducted in previous studies. These data serve as the foundation for detecting potential misbehavior.

2. Process: In the Process stage, Ensemble learning algorithms are applied to analyze this data and classify different types of cyber threats, ensuring high detection accuracy and adaptability to evolving attack patterns. A key aspect of this stage is processing time, which ensures that real-time anomaly detection can be performed without introducing significant computational delays, maintaining both high performance and lightweight processing.
3. Output: The Output stage focuses on utilizing the analysis results to trigger real-time automatic alerts, enabling rapid response mechanisms within railway systems.

By integrating advanced Ensemble Learning techniques, this framework enhances cybersecurity within CBTC signaling networks while maintaining scalable, lightweight processing capabilities. Ultimately, it aims to improve the operational safety and resilience of modern urban railway systems by enabling faster and more accurate detection of misbehavior in data communication, all while ensuring efficiency through optimized processing time.

1.4 Statement of Problem

CBTC signaling systems face significant challenges in cybersecurity, as cyberattacks can disrupt operational reliability and compromise passenger safety. T2G communications within CBTC infrastructure, which rely on wireless radio, are vulnerable to attacks like False Data Injection, often executed through a MitM attack. In such an attack, the MitM masquerades as a legitimate party, intercepting T2G wireless communications and undermining the integrity and timeliness of crucial data exchanges. As a result, these attacks present a serious risk to the operational safety of trains, potentially leading to disruptions or accidents.

1.5 Research Objective

This study aims to identify and classify patterns of cyber-attacks on T2G communication within the CBTC system, with a particular focus on developing a lightweight Ensemble Learning-based detection method for real-time anomaly detection. The research utilizes data from train infrastructure, including train speed, data sending/receiving times, and train position, to develop an efficient anomaly detection model. By emphasizing processing time optimization, the model ensures real-time detection without introducing significant computational delays, maintaining a balance between high performance and lightweight processing. The comparison of

model processing durations serves to identify the most lightweight model, ensuring the detection system is both effective and efficient. The model's performance will be evaluated using key Ensemble learning metrics, such as accuracy, precision, recall, F1-score, Processing Time, and Fit Status to assess the effectiveness, reliability, and scalability of the proposed detection method within the CBTC signaling system.

1.6 Hypothesis

The proposed Ensemble Learning-based Misbehavior Detection System is expected to be able to effectively detect and classify cyberattacks such as such as false train position, false speed information, and delayed T2G communication log on the CBTC signaling system. This approach is believed to improve the defense capability, data integrity and overall performance of the CBTC system.

1.7 Research Methodology

This study will use quantitative methodology, combining experimental design and data-driven approach to develop and evaluate a MDS for CBTC system. The data used in this study will be sourced from cyber-attack simulation conducted in previous studies. After data pre-processing, Ensemble Learning model will be trained to detect and classify cyber-attacks, such as false train position, false speed information, and delayed T2G communication log. To obtain the best model, the model processing time will be one of the evaluation parameters or the best modelling selection to ensure efficient anomaly detection without significant computational delay. Additionally, the training results will be evaluated using metrics such as accuracy, precision, recall, F1-score, and Fit Status, while the model's performance will be tested using a separate test dataset. The effectiveness of the proposed Ensemble Learning algorithm in MDS will be assessed by comparing the F1-score of each model, as well as the data processing time, ensuring the efficiency and scalability of the method for real-time detection in the CBTC system.

1.8 Research Method

This research method consists of the following steps:

1. Dataset Validation: The dataset used will be validated by experts in the field of CBTC signaling systems and related lecturers to ensure that the data is relevant and appropriate to the research scenario.

2. Data pre-processing: Data will be processed through stages of cleaning, reordering, and labelling to ensure its quality before being analysed.
3. Development of Ensemble Learning Models: Models such as Random Forest, kNN with Bagging, XGBoost, LightGBM, and AdaBoost will be applied to detect cyberattacks, including false train position, false speed information, and delayed T2G communication logs. In addition to detecting threats, the development will focus on processing time to ensure lightweight and efficient model performance without introducing significant computational delays.
4. Model Performance Evaluation: Each model will be tested using labelled data and test data to evaluate performance based on matrices such as accuracy, precision, recall, F1-score and Fit Status. The processing time of each model will also be considered as a key factor in selecting the most efficient model for real-time detection.
5. Model Effectiveness Comparison: The F1-score of each model will be compared to determine the best approach in improving cybersecurity in CBTC systems, with an emphasis on the lightweight nature and optimization of processing time to ensure scalability and efficient anomaly detection in real-time applications.

Through these steps, this research aims to develop an MDS system that can improve operational reliability and safety in the Urban Rail System, while maintaining an optimal balance between accuracy and processing efficiency.

1.9 Scope and Delimitation

This study focuses only on modelling and evaluating the Misbehavior Detection System based on Ensemble Learning for the CBTC signaling system in the context of cyberattacks, using attack simulation data that includes false train position, false speed information, and delayed T2G communication logs. The delimitation of this study includes testing the detection model only on attacks related to these aspects and using 5 Ensemble Learning models such as Random Forest, kNN with Bagging, XGBoost, LightGBM, and AdaBoost. The study also emphasizes optimizing processing time to ensure the lightweight nature of the detection system, which allows for real-time anomaly detection without significant computational delays.

The study is limited to implementation in a simulation environment, without direct application to operational CBTC systems, and does not include exploration of other Ensemble Learning models or testing with real-world data from systems that are already in operation. However, the models and findings from this research are expected to offer valuable insights for future deployment in operational settings, provided that the processing time optimization and scalability of the models are confirmed.

1.10 Importance of the Study

This research aims to improve the security, reliability, and operational safety of urban railway systems in the face of technological advancements and evolving cyber threats. The research develops a Misbehavior Detection System based on Ensemble Learning to detect anomalies such as false train position, false speed information, and delayed T2G communication logs in real-time, while emphasizing the importance of processing time optimization. This lightweight approach ensures that the system can detect misbehaviors efficiently without imposing significant computational delays, maintaining high performance.

The results of this research are expected to strengthen the security of data communication in the CBTC signaling system, offering a scalable and efficient solution for real-time misbehavior detection. The findings also support the creation of a safer, more reliable urban railway system, improving its resilience to cyber threats and ensuring operational continuity in the face of emerging challenges.