A NOVEL FRAMEWORK FOR GENERATING FALSE DATA INJECTION ATTACKS IN POWER GRIDS: FDI-GENERATION

by

Krishna Vaibhav

Submitted in partial fulfillment of the requirements
for the degree of Master of Applied Computer Science

at

Dalhousie University
Halifax, Nova Scotia
Dec 2024

Table of Contents

List of Tables

List of Figures

Abstract

Acknowledgements

I would like to express my deepest gratitude to my professor, Dr.Srini Sampalli, for giving me the opportunity to work under his guidance. His invaluable support and encouragement have been instrumental in successfully completing this research.

A special thanks to my co-supervisor, Dr.Darshana Upadhyay, for her constant support; her insights and guidance have been crucial throughout this journey.

I am also grateful to Dr.Marzia Zaman for her valuable contributions and knowledge of machine learning, which significantly enhanced the quality of this research.

Finally, I extend my heartfelt thanks to my parents and sister. Their continuous support and belief in me have been my greatest strength.

Chapter 1

Introduction

1.1 Motivation for Study

The modern electric power grid is experiencing a profound transformation with the integration of advanced communication and information technologies, leading to the development of smart grids. While this integration enhances operational efficiency and allows for better resource management, it also exposes the power system to new cybersecurity threats [3]. Among these threats, False Data Injection (FDI) attacks have emerged as a significant concern due to their ability to stealthily compromise state estimation processes without being detected by conventional security measures [1].

FDI attacks can manipulate measurement data, leading to incorrect state estimation results that can cause erroneous control actions, equipment damage, and large-scale power outages [5]. The stealthy nature of these attacks makes them particularly dangerous, as they can bypass traditional Bad Data Detection (BDD) schemes and remain undetected until substantial harm is done [?]. The increasing sophistication of cyber-attack methodologies necessitates the development of advanced tools and techniques to understand, detect, and mitigate such threats.

Current research efforts have focused on developing detection algorithms and defense mechanisms against FDI attacks. Machine learning approaches, including the use of Generative Adversarial Networks (GANs), have shown promise in modeling and detecting these attacks [9]. However, there is a notable gap in accessible and flexible simulation tools that can generate realistic FDI attack scenarios for research and testing purposes [4]. Such tools are essential for evaluating the effectiveness of detection algorithms and for training machine learning models under various attack conditions.

The motivation for this study stems from the critical need to enhance the cybersecurity of power systems by providing a platform that can simulate realistic FDI attacks. By developing a simulation tool that can generate and inject false data into power system datasets, researchers and grid operators can better understand the vulnerabilities of state estimation processes and develop more robust detection and mitigation strategies. This tool will contribute to improving the resilience of power systems against cyber-attacks, ensuring the reliability and stability of the electrical grid in the face of evolving threats.

1.2 Problem Overview

The power grid, being one of the most critical infrastructures in any nation, is a target for various types of cyber-attacks, including False Data Injection (FDI) attacks. FDI attacks are a sophisticated form of cyber-attack in which false data is injected into the power system's measurement data, misleading the state estimation process, which is crucial for maintaining grid stability and security [1]. Unlike traditional cyber-attacks that directly target hardware or software components, FDI attacks exploit the inherent vulnerabilities in the data processing algorithms used in power grid monitoring and control systems.

The state estimation process in power systems relies on measurement data from various points in the grid, such as bus voltages, power flows, and load demands [3]. By manipulating this data, attackers can cause incorrect state estimations, leading to erroneous control decisions, power imbalances, equipment damage, or even large-scale blackouts [?]. One of the key challenges in detecting FDI attacks is their stealthiness; they can be designed to evade conventional Bad Data Detection (BDD) mechanisms, making them extremely difficult to identify and mitigate in real-time [5].

Given the increasing complexity and interconnectivity of modern power grids, the potential for large-scale, coordinated FDI attacks is becoming a critical concern for grid operators and cybersecurity experts. This highlights the urgent need for advanced tools and techniques to simulate, detect, and mitigate FDI attacks. Current solutions are limited in their ability to simulate realistic attack scenarios, especially for testing and evaluating machine learning-based detection algorithms [9]. This gap in research and technology underscores the importance of developing a flexible and scalable simulation tool capable of generating FDI attacks under various conditions and testing them against state estimation algorithms.

1.3 Objectives of the Research

The primary objective of this research is to develop a simulation tool that can generate realistic False Data Injection (FDI) attacks for power systems, specifically targeting the state estimation process. The key objectives are as follows:

- To develop a tool that can simulate a variety of FDI attack scenarios by manipulating normal grid operation data, such as bus voltages, power flows, and generator outputs.
- To provide a platform for testing and evaluating the effectiveness of machine learning-based FDI detection algorithms by generating realistic, customizable attack datasets.
- To analyze the impact of FDI attacks on state estimation accuracy, grid stability, and control decisions, particularly under different types of stealth and sparse attack scenarios [4].
- To contribute to the body of knowledge on power system cybersecurity by exploring new methods for generating, simulating, and mitigating FDI attacks, including the use of advanced machine learning models like Generative Adversarial Networks (GANs) [9].

By achieving these objectives, this research aims to enhance the understanding of how FDI attacks can compromise power system operations and develop better tools for mitigating their effects. The simulation tool will be designed to be flexible and scalable, allowing users to input various power system datasets and customize the attack parameters to simulate real-world scenarios.

1.4 Relevance of FDI Simulation in Power Systems

Simulating FDI attacks in power systems is crucial for understanding their impact and developing robust detection and mitigation strategies. The modern power grid is becoming increasingly vulnerable to cyber-attacks as it incorporates more digital and communication technologies. FDI attacks, in particular, exploit the data processing layer of power systems, making them difficult to detect using traditional cybersecurity measures [**?**].

By simulating FDI attacks, researchers can analyze how these attacks affect the accuracy of state estimation, the stability of the grid, and the control decisions made by operators. Simulation tools are essential for generating datasets that can be used to train machine learning algorithms to detect and respond to FDI attacks. Additionally, simulation allows for the testing of various detection and mitigation strategies under controlled conditions, providing insights into their strengths and weaknesses [5].

The development of a flexible simulation tool that can generate realistic FDI attacks will enable power system operators and researchers to better prepare for potential cyber-attacks. It will also facilitate the development of more effective detection mechanisms, ensuring the resilience and security of the grid. This research is particularly relevant as it contributes to the ongoing efforts to secure critical infrastructure from evolving cyber threats, including stealth and sparse FDI attacks [9].

1.5 Project Outline

This thesis is organized into to-do chapters, each addressing key aspects of the research objectives and contributions.

Chapter 1, *Introduction*, provides a comprehensive overview of False Data Injection (FDI) attacks in power systems. It discusses the motivation for developing an FDI simulation tool, highlights the importance of detecting and mitigating FDI attacks, and outlines the objectives of this research. Additionally, the chapter identifies research gaps in FDI simulation and discusses the potential contributions of the project.

Chapter 2, *Background*, presents the fundamental concepts of power systems and state estimation. It explains the role of buses, generators, and loads in grid reliability. Additionally, it defines and classifies FDI attacks and explores their impact on power system operations and stability. Various methods for generating FDI data and the challenges associated with detecting such attacks are also discussed.

Chapter 3, *Literature Review*, examines existing research on generating and detecting FDI attacks. It highlights different approaches, such as stealth FDI, sparse attacks, and machine learning methods like GANs for simulating attacks. The chapter also reviews data augmentation techniques used in preparing datasets for machine-learning-based cyber-attack detection.

Chapter 4, *Power System Components Susceptible to FDI Attacks*, focuses on the specific components within a power grid that are vulnerable to FDI. These include bus voltages, power flows, state estimation variables, generator outputs, and load demand. The chapter explains how FDI manipulates these components and their role in grid operations.

Chapter 5, *Methodology for FDI Simulation Tool*, outlines the technical approach used to develop the FDI simulation tool. It discusses the selection and preprocessing of the IEEE 118-Bus System dataset, the calculation of the Jacobian matrix, and the generation of attack vectors. The process of injecting false data into the dataset and validating the results to ensure realistic attack behavior is detailed.

Chapter 6, *Design of the FDI Simulation Tool*, presents the architecture of the simulation tool, implemented using Python and the Pencil framework. It describes the user input parameters, how the dataset is loaded, and how the tool generates the modified dataset with injected FDI data. The chapter also includes the step-by-step process of attack generation, the integration of machine-learning models for automated attack generation, and customization options for attack scenarios.

Chapter 7, *Results and Evaluation*,provides a performance evaluation of the FDI simulation tool. It presents the results obtained from applying the tool to the IEEE 118-Bus System dataset, comparing normal and FDI-altered scenarios. Key performance metrics, such as the impact of FDI on state estimation, are analyzed. The chapter also evaluates the stealthiness of the attacks generated by the tool and their resistance to detection by existing monitoring systems.

Chapter 8, *Discussion*,discusses the challenges faced in detecting FDI attacks in real-time power system operations. It explores the limitations of current state estimation methods and suggests potential improvements to the FDI tool. The chapter also addresses future directions for integrating larger datasets and real-time data into the simulation tool.

Finally,**Chapter 9**, *Summary and Conclusion*,summarizes the key findings of the research and its contributions to power grid security and FDI research. The chapter concludes with recommendations for enhancing the simulation tool and outlines the next steps for continuing research on FDI attacks.

Chapter 2

Background

2.1 Power Systems and State Estimation

Power systems are complex infrastructures that consist of various interconnected components such as buses, generators, transformers, loads, and transmission lines. These elements work together to ensure that electricity is generated, transmitted, and distributed efficiently to meet consumer demands. The operational stability of power systems relies on monitoring the state of the system, which includes determining the voltages at buses, the power flows through transmission lines, and the loads on the system [2].

2.1.1 Buses, Generators, and Loads

A bus in a power system refers to a node where one or more components, such as generators, transformers, and loads, are connected. Buses act as key points for measuring the electrical properties like voltage magnitude and phase angle. Generators at different buses supply electrical power, while loads represent the demand or consumption of electricity. The balance between power generation and load demand is crucial for ensuring stable grid operations. Any imbalance may lead to voltage instabilities or even system-wide blackouts [3].

2.1.2 State Estimation in Power Systems

State estimation is a critical function in the operation of power systems, as it provides a real-time snapshot of the grid's operational state. It involves collecting measurement data from sensors, such as Phasor Measurement Units (PMUs) and Supervisory Control and Data Acquisition (SCADA) systems, to estimate the voltage magnitudes and phase angles at various buses. These estimations enable grid operators to make informed decisions, ensuring grid reliability and preventing potential faults [5]. Accurate state estimation is essential because the control and optimization of power flow, load balancing, and fault detection heavily depend on the precise knowledge of the system's current state.

However, state estimation is vulnerable to data manipulation attacks such as False Data Injection (FDI) attacks, where attackers inject manipulated data into the measurement readings to disrupt grid operations. Detecting these attacks is crucial for maintaining the integrity and security of power systems [1].

2.2 False Data Injection (FDI) Attacks

2.2.1 Definition and Classification of FDI Attacks

False Data Injection (FDI) attacks are a type of cyber-attack that target the measurement data used in power system state estimation. In these attacks, an adversary injects manipulated data into the system with the goal of disrupting the state estimation process and potentially causing incorrect operational decisions. These attacks are particularly dangerous because they can be crafted to bypass conventional Bad Data Detection (BDD) mechanisms, making them difficult to detect in real-time operations [1].

FDI attacks can be classified into two main categories:

- **Stealth FDI Attacks**: These attacks are designed to evade detection by traditional BDD systems. The attacker strategically manipulates measurement data such that the errors introduced do not trigger any alarms, making the attack "stealthy" and difficult to detect. Research by Liu et al. [1] highlighted how such stealth attacks can bypass state estimation without being flagged by BDD mechanisms.

- **Sparse FDI Attacks**: Sparse attacks involve manipulating only a small subset of measurement data to reduce the risk of detection. By altering a limited number of data points, the attacker can still cause significant disruption to power system operations, but with a reduced chance of triggering alarms. Research by Wei et al. [5] and Narang et al. [2] has demonstrated how sparse attacks can be optimized to remain undetected while causing maximum disruption.

Both types of attacks represent significant threats to modern power systems, as they can lead to incorrect decision-making by grid operators, potentially destabilizing the grid. The ability of attackers to carefully craft FDI attacks to evade traditional detection mechanisms has led to an increased focus on the development of more advanced detection and prevention strategies.

2.2.2 Impact of FDI on Power System Operations and Grid Stability

FDI attacks pose a significant threat to the operation and stability of power systems. By injecting false data into the state estimation process, attackers can cause the system to operate under incorrect assumptions, leading to potentially catastrophic consequences. These attacks can manipulate critical state variables such as bus voltages, power flows, and generator outputs, leading to incorrect operational decisions and instability in the grid.

The impact of these attacks can be categorized as follows:

- **Operational Disruptions**: FDI attacks can cause the state estimation process to report incorrect bus voltages and power flows, leading to operational mismanagement. Grid operators may be unaware of the attack and could incorrectly dispatch generators, overload transmission lines, or initiate unnecessary load shedding, resulting in cascading failures [3].

- **Economic Losses**: FDI attacks can disrupt the optimal power flow (OPF) process, which is used to minimize operational costs while maintaining grid stability. By manipulating the data used for OPF, attackers can force the grid to operate in a suboptimal state, leading to increased operational costs. Fioretto et al. [10] demonstrated how FDI attacks could be used to disrupt the OPF process, resulting in significant economic losses for power utilities.

- **Grid Instability and Blackouts**: In extreme cases, FDI attacks can lead to widespread grid instability and even blackouts. By altering key measurements related to generator outputs and transmission line flows, attackers can cause cascading failures, where the initial disruption spreads throughout the system, potentially leading to a large-scale blackout [9].

2.2.3 Examples of FDI Attack Scenarios

Several studies have provided real-world and simulated examples of FDI attacks to illustrate their potential impact on power systems. The IEEE 118-Bus System has been widely used in these studies to simulate the effects of FDI attacks. In research conducted by Shohan et al. [9], generative models were employed to create synthetic FDI attack data that closely mimics real-world measurement data. This work highlights how machine learning techniques, such as Generative Adversarial Networks (GANs), can be used to generate stealthy FDI attacks, making detection even more challenging.

Another prominent example is the use of optimization techniques to generate sparse FDI attacks. Wei et al. [5] showed how sparse attack points can be optimized to minimize detection while still causing significant operational disruptions. Similarly, Narang et al. [2] explored how attackers could manipulate the state estimation process with minimal network information, further demonstrating the stealthy nature of these attacks.

The multi-objective optimization approach, as explored by Li et al. [4], investigates how attackers can balance between stealthiness and impact. Their work on optimizing FDI attacks in smart grids revealed that attackers can tailor their strategies to maximize disruption while minimizing the likelihood of detection.

2.2.4 Challenges in Detecting FDI Attacks

One of the primary challenges in detecting FDI attacks is their ability to evade traditional detection mechanisms, such as BDD systems. These attacks are carefully crafted to ensure that the residuals calculated during state estimation remain within the normal range, making them appear as legitimate data to grid operators [1].

Another challenge lies in the detection of sparse FDI attacks. Sparse attacks are designed to manipulate only a small subset of the data, making them difficult to detect using traditional anomaly detection methods. Sparse attack vectors can be optimized to minimize their impact on the overall system while still causing significant operational disruptions [5].

In addition, machine learning models used for detecting FDI attacks face the challenge of adaptability. As attackers evolve their strategies, pre-trained models may become ineffective, requiring continuous retraining and updating to stay ahead of new attack techniques [9]. Developing robust real-time detection systems that can adapt to evolving attack patterns remains a critical area of research.

2.3 Methods for Generating FDI Data

2.3.1 Overview of Methods Used to Generate FDI Data

The generation of False Data Injection (FDI) data is a complex process that requires a deep understanding of both the power system's state estimation mechanisms and the techniques used by attackers to manipulate measurement data. Broadly, there are three main approaches to generating FDI data: 1) Mathematical Modeling, 2) Simulation-Based Approaches, and 3) Machine Learning Models.

Mathematical Modeling approaches involve manipulating the power system's measurement data using predefined attack vectors. These attack vectors are calculated based on the power system's Jacobian matrix, which maps the relationship between the measurements and the system state variables. By altering the measurement values using these attack vectors, an attacker can generate false data that disrupts the state estimation process without triggering Bad Data Detection (BDD) mechanisms [1]. This method is frequently used in research to simulate stealth FDI attacks that evade detection [2].

Simulation-Based Approaches use specialized simulation tools, such as MATPOWER or PSS/E, to simulate real-world power system operations under attack conditions. These tools allow researchers to simulate various attack scenarios by injecting manipulated data into the power flow equations or state estimation algorithms. Such simulations help in understanding the impact of FDI attacks on the power grid's stability, reliability, and control mechanisms [3].

Machine Learning Models are increasingly being used to generate FDI data. Techniques such as Generative Adversarial Networks (GANs) have been applied to create synthetic FDI attack data that closely resembles real-world measurement data. GANs can learn the patterns of legitimate grid data and then generate false data that mimics these patterns, making detection even more challenging. These models are particularly effective in creating stealth attacks, where the injected data is indistinguishable from legitimate data [9]. Additionally, other machine learning techniques, such as Random Matrix Theory (RMT) and ADASYN, are used for augmenting and generating datasets that contain both normal and anomalous events, facilitating the training of detection models [5].

2.3.2 Challenges in Detecting FDI Attacks

Detecting FDI attacks presents several challenges due to the sophisticated nature of these attacks and their ability to bypass traditional detection mechanisms like Bad Data Detection (BDD) systems.

Stealthiness of the Attacks: One of the primary challenges is the stealthiness of well-crafted FDI attacks. By carefully selecting the attack vectors, an attacker can inject malicious data into the system without triggering any alarms. The stealthiness is achieved by ensuring that the residuals—calculated during the state estimation process to detect errors—remain within the acceptable range [2]. This makes it incredibly difficult for grid operators to differentiate between legitimate and manipulated data [1].

Sparse Attacks: Another challenge lies in detecting sparse FDI attacks, where only a few data points are manipulated. These attacks are designed to have minimal impact on the overall system, making them harder to detect while still causing significant operational disruptions [5]. Traditional detection methods that rely on anomaly detection techniques often fail to identify sparse attacks due to the limited number of manipulated data points.

Lack of Real-Time Detection Mechanisms: Many power grids rely on batch processing of data for state estimation and anomaly detection, which delays the detection of FDI attacks. The lack of real-time monitoring tools for identifying suspicious changes in the grid's state makes it difficult to respond to FDI attacks before they cause damage [4]. Real-time detection methods, such as machine learning-based systems, are still under development, and their deployment is not widespread in modern power grids.

Adaptability of Machine Learning Models: While machine learning models have shown promise in detecting FDI attacks, they face challenges in adapting to new types of attacks that were not present in the training data. Attackers can modify their strategies, rendering pre-trained models ineffective. This adaptability gap highlights the need for continuous retraining and updating of detection models to stay ahead of evolving attack strategies [9].

2.3.3 Overview of Methods Used to Generate FDI Data

The generation of False Data Injection (FDI) data is a complex process that requires a deep understanding of both the power system's state estimation mechanisms and the techniques used by attackers to manipulate measurement data. Broadly, there are three main approaches to generating FDI data: 1) Mathematical Modeling, 2) Simulation-Based Approaches, and 3) Machine Learning Models.

One of the primary resources used in this research is the IEEE 118-Bus System Dataset, which provides a comprehensive view of power grid operations under normal conditions. This dataset has been extended with the generation of FDI attack data through simulation methods, offering both normal and manipulated datasets for analysis.

2.3.4 Normal and FDI Datasets: IEEE 118-Bus System Resource

The IEEE 118-Bus System Dataset is commonly used in power system research for simulating grid operations. In this project, the normal dataset provided by this resource is augmented by generating synthetic FDI attacks through mathematical modeling and simulation. The FDI dataset is created by injecting false data into key state variables such as bus voltages, power flows, and generator outputs, which are crucial for the state estimation process.

Below is a summary of the normal and FDI datasets provided by the resource:

Table 2.1: Normal and FDI Datasets from the IEEE 118-Bus System Resource

Dataset	**Key Features**	**Description**
Normal Dataset	Bus Voltages (V), Power Flows (P, Q), Generator Outputs	Contains normal operational data for the IEEE 118-Bus system.
FDI Dataset	Manipulated Bus Voltages, Power Flows, Generator Outputs	Generated via simulation by injecting false data into the normal dataset.

The normal dataset provides the baseline operational data, which is then used to generate the FDI dataset by introducing anomalies in the form of false data. This process is crucial for testing the effectiveness of detection algorithms and assessing the resilience of power grids against FDI attacks [11].

2.3.5 Challenges in Detecting FDI Attacks

Detecting FDI attacks presents several challenges due to the sophisticated nature of these attacks and their ability to bypass traditional detection mechanisms like Bad Data

Detection (BDD) systems.

Stealthiness of the Attacks: One of the primary challenges is the stealthiness of well-crafted FDI attacks. By carefully selecting the attack vectors, an attacker can inject malicious data into the system without triggering any alarms. The stealthiness is achieved by ensuring that the residuals—calculated during the state estimation process to detect errors—remain within the acceptable range [2]. This makes it incredibly difficult for grid operators to differentiate between legitimate and manipulated data [1].

Sparse Attacks: Another challenge lies in detecting sparse FDI attacks, where only a few data points are manipulated. These attacks are designed to have minimal impact on the overall system, making them harder to detect while still causing significant operational disruptions [5]. Traditional detection methods that rely on anomaly detection techniques often fail to identify sparse attacks due to the limited number of manipulated data points.

Lack of Real-Time Detection Mechanisms: Many power grids rely on batch processing of data for state estimation and anomaly detection, which delays the detection of FDI attacks. The lack of real-time monitoring tools for identifying suspicious changes in the grid's state makes it difficult to respond to FDI attacks before they cause damage [4]. Real-time detection methods, such as machine learning-based systems, are still under development, and their deployment is not widespread in modern power grids.

Adaptability of Machine Learning Models: While machine learning models have shown promise in detecting FDI attacks, they face challenges in adapting to new types of attacks that were not present in the training data. Attackers can modify their strategies, rendering pre-trained models ineffective. This adaptability gap highlights the need for continuous retraining and updating of detection models to stay ahead of evolving attack strategies [9].

Chapter 3

Literature Review

3.1 Overview of False Data Injection (FDI) Attacks

False Data Injection (FDI) attacks have become a significant concern for modern power systems, especially as the reliance on automated state estimation continues to grow. These attacks target the data collected and processed by the state estimation algorithms used to monitor and control power grids, leading to misleading information being fed into the system. As a result, the power grid can make erroneous operational decisions, affecting grid stability, reliability, and security.

FDI attacks are unique because they bypass traditional security mechanisms. Rather than directly tampering with physical infrastructure, attackers inject falsified data into the system's measurement streams, altering the grid's understanding of its operational state. The seminal work by [1] laid the foundation for understanding how such attacks exploit vulnerabilities in the state estimation process, demonstrating that an attacker could introduce arbitrary errors without being detected by conventional bad data detection mechanisms.

In addition to altering state estimation, these attacks can have far-reaching consequences on grid operations. As noted by [8], FDI attacks can lead to inefficient power flows, increased operational costs, and in some extreme cases, even blackouts. Such disruptions make detecting and mitigating FDI attacks crucial for maintaining grid reliability.

Moreover, FDI attacks are challenging to detect because they can be designed to appear indistinguishable from normal system measurements. This makes traditional monitoring and anomaly detection techniques less effective in identifying the subtle changes caused by the attack, as highlighted in [2]. These attacks exploit weaknesses in measurement redundancy and the structure of the power grid, requiring advanced techniques for detection and mitigation.

The growing prevalence of smart grid technologies has increased the potential attack surface for FDI. With more interconnected devices and automated systems collecting and processing data, the opportunities for attackers to exploit vulnerabilities have expanded. This has sparked significant research into better understanding the nature of FDI attacks and developing strategies to counteract them. Researchers, such as [3], have emphasized the need for more robust data integrity solutions that can both detect and mitigate these attacks in real time.

Overall, understanding FDI attacks is essential for designing more secure and resilient power systems. By studying how attackers manipulate data, researchers and engineers can develop more effective defense mechanisms, ensuring the continued reliability of our power infrastructure.

3.2 Stealth FDI Attacks

Stealth False Data Injection (FDI) attacks represent a sophisticated category of attacks that remain undetected by conventional monitoring systems in power grids. These attacks target the underlying state estimation algorithms, which are responsible for ensuring the grid's operational accuracy. The key characteristic of stealth FDI attacks is their ability to manipulate data while evading detection by bypassing bad data detection mechanisms that rely on statistical anomalies.

The foundational study by Liu et al. (2009) introduced the concept of stealth FDI attacks, demonstrating that such attacks could be constructed to bypass bad data detection (BDD) methods used in state estimation processes [1]. Their work showed that by carefully selecting the attack vector based on the system's Jacobian matrix, an attacker could inject erroneous data that appears statistically normal, thereby eluding detection systems. This research highlighted the importance of securing the state estimation process, as stealth FDI attacks can induce serious errors in power grid operation without being flagged by traditional security measures.

Further analysis of stealth FDI attacks was conducted by Narang and Bag (2019), who proposed a method for launching stealth attacks with minimal network information. Their approach reduced the amount of information needed by attackers, making it feasible to launch FDI attacks even with limited access to the power system's configuration [2]. This work emphasizes that attackers no longer need comprehensive knowledge of the system, as partial information is sufficient to execute successful stealth FDI attacks. This finding broadens the potential attack surface and increases the risk to power systems.

In addition, Wei et al. (2020) explored sparse stealth FDI attacks, where the attack vectors are strategically chosen to affect only a small number of measurement points, further reducing the chances of detection [5]. These sparse attacks demonstrate that it is possible to achieve significant disruption with minimal resource expenditure, making them an attractive option for attackers. Sparse attacks are particularly challenging to detect, as they involve minimal data manipulation, leaving fewer traces for monitoring systems to pick up on.

The effectiveness of stealth FDI attacks lies in their ability to blend into normal system operations, altering state estimates without introducing observable irregularities. The work of Li et al. (2021) examined multi-objective optimization strategies for stealth FDI attacks, where attackers optimize their attack vectors to achieve both stealth and operational impact [4]. Their study showed that stealth attacks could be optimized to cause significant damage to grid operations, such as overloading transmission lines or disrupting power flow, while remaining undetected by traditional monitoring systems.

Researchers have proposed several techniques for defending against stealth FDI attacks, but the inherent complexity of the attacks makes them difficult to counter. The use of advanced machine learning models, as explored by Shohan et al. (2021), offers some promise in detecting and mitigating stealth FDI attacks [9]. By analyzing patterns in system data, machine learning models can potentially identify subtle deviations caused by stealth attacks, although these models must be trained on large, representative datasets to be effective.

In conclusion, stealth FDI attacks pose a significant threat to the reliability and security of power systems. By exploiting weaknesses in the state estimation process, attackers can introduce malicious data without raising alarms, causing potentially catastrophic consequences for grid operations. The challenge of detecting these attacks underscores the need for more advanced security measures and monitoring techniques to ensure the continued resilience of power grids.

3.3 Machine Learning Approaches to FDI Generation

In recent years, the application of machine learning (ML) techniques to False Data Injection (FDI) attack generation has gained significant traction. Machine learning provides the capability to model complex patterns and learn from large datasets, which makes it highly suited for simulating and generating sophisticated FDI attack vectors. These approaches aim to exploit the inherent vulnerabilities of power systems by using data-driven methods to generate realistic, yet malicious, data that bypasses traditional detection methods.

A prominent approach for FDI generation is the use of Generative Adversarial Networks (GANs), a type of machine learning model where two neural networks, a generator and a discriminator, are trained in opposition to each other. The generator's role is to create synthetic data, such as falsified state estimation values, while the discriminator's role is to differentiate between the real data and the generated (fake) data. Over time, the generator improves its ability to create data that is indistinguishable from the real data. This method was employed by Shohan et al. (2021) in the iAttackGen model, where GANs were used to generate realistic FDI attack vectors in a power grid context [9]. This approach is particularly useful because it automates the generation of attack vectors, which can adapt to various scenarios in real time.

Supervised learning is another method often employed in FDI generation, where the machine learning model is trained on a labeled dataset. In this case, the dataset includes examples of both normal and malicious data. The model learns to classify data points as either normal or injected (malicious) based on predefined features. These features may include voltage readings, power flows, and other state estimation variables that are critical to power system operations. Once the model is trained, it can be used to simulate FDI attacks by manipulating specific features to generate new data points that mimic realistic attack scenarios. As demonstrated by Chen et al. (2021), this method is highly effective for detecting and generating data that mirrors real-world attacks [3]. By applying feature manipulation, researchers can generate diverse attack scenarios that can be used to train detection models.

Another significant machine learning approach used for FDI generation is reinforcement learning (RL). In RL, an agent interacts with the power grid environment and learns a strategy, or policy, to maximize its reward over time. In the context of FDI, the reward function is designed to reflect the success of the attack in terms of its ability to disrupt the system or evade detection. The agent iteratively learns from its actions by receiving feedback from the environment, adjusting its strategy to achieve better results. For instance, an FDI attack could be modeled as a sequential decision-making problem where the attacker incrementally injects false data and learns which modifications to make in order to maximize impact while remaining undetected. Studies like that of Wei et al. (2020) show that RL can be an effective tool for optimizing stealthy FDI attacks, particularly in scenarios where attackers have limited knowledge of the grid's structure [5].

In addition to supervised and reinforcement learning, unsupervised learning techniques like clustering have also been explored. Clustering algorithms, such as k-means or hierarchical clustering, group similar data points together based on their features. In the context of FDI generation, unsupervised learning can be used to identify patterns in historical grid data and generate attack vectors that mimic the normal behavior of these clusters. This approach is particularly useful in cases where labeled data is not available, as it does not rely on pre-classified examples of attacks. Instead, the algorithm learns the underlying structure of the data and generates new data points that fit within these learned patterns.

Another promising avenue is the use of deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are well-suited for handling sequential and high-dimensional data. CNNs are typically used for tasks involving grid data that can be represented spatially, while RNNs, particularly their variants like Long Short-Term Memory (LSTM) networks, are excellent for capturing temporal dependencies in sequential data such as power system measurements over time. Li et al. (2021) demonstrated how deep learning models could be trained to predict optimal attack vectors by learning from historical grid data [4]. By simulating past attacks and training the model on this data, they were able to generate new attack vectors that targeted vulnerabilities in the system's state estimation process.

The integration of machine learning techniques into FDI generation opens up new possibilities for simulating more realistic and adaptive attack scenarios. These models, especially when coupled with optimization techniques, allow for the creation of attack vectors that are not only effective but also harder to detect. However, a significant challenge remains: machine learning models require large, high-quality datasets to train effectively. This is particularly true for deep learning models, which are data-hungry and need extensive training data to capture subtle patterns in the grid's operation.

In conclusion, machine learning approaches to FDI generation represent a powerful toolset for attackers. By leveraging advanced models like GANs, reinforcement learning, and deep learning, attackers can craft sophisticated and adaptive attack strategies that pose serious threats to power systems. As these techniques continue to evolve, so too must the defensive mechanisms designed to detect and mitigate them. 4

3.4 Data Augmentation for FDI Detection

Data augmentation techniques are essential in generating large datasets to train machine learning models for the detection of False Data Injection (FDI) attacks. Given the stealthy nature of FDI, which often blends seamlessly into normal power system operations, traditional datasets often lack the diversity needed to train robust detection models. In this section, we discuss how various data augmentation techniques, including Random Matrix Theory (RMT), SMOTE (Synthetic Minority Over-sampling Technique), and ADASYN, have been employed in existing research to improve FDI detection.

[3] introduced a method based on RMT for FDI detection by generating synthetic datasets that mimic the normal and attack scenarios in power system behavior. This method not only increases dataset diversity but also enhances the robustness of detection models by exposing them to a wider range of potential anomalies.

On the other hand, [5] and [8] explored the use of data augmentation techniques like SMOTE and ADASYN to address class imbalances in datasets. FDI attack instances are typically far fewer than normal operations, leading to skewed datasets that hinder the performance of machine learning models. SMOTE and ADASYN create synthetic attack data, ensuring that models are better trained to recognize FDI scenarios, even in heavily imbalanced datasets.

[8] further demonstrated how synthetic data generation for FDI attack detection can enhance machine learning-based detection models' performance. By simulating a wide range of attack vectors, the datasets generated through these augmentation techniques provide a comprehensive training environment for algorithms aimed at real-time FDI detection.

Table 3.1: Overview of Data Augmentation Techniques in FDI Detection

Technique	**Description**	**Application in FDI Detection**	**Key Findings**
RMT	Generates synthetic data using random matrix theory	Used to expand dataset diversity for machine learning models	Enhanced detection performance due to diverse data [3]
SMOTE	Oversamples minority class (attack data)	Balances class distribution in imbalanced datasets	Improved model performance in detecting rare attack instances [5, 8]
ADASYN	Adaptive synthetic sampling technique	Focuses on generating synthetic examples for underrepresented attack cases	Enhanced model training for harder-to-classify cases [8]

In addition to these techniques, [8] proposed a method for analyzing vulnerabilities and assessing the consequences of FDI attacks on power systems. Their work introduced synthetic augmentation techniques that allow for detailed simulation of attack scenarios, thereby enhancing the dataset and providing a better training ground for machine learning-based detection systems.

As a result, augmentation techniques such as RMT, SMOTE, and ADASYN help address the scarcity of real-world FDI attack data and significantly enhance the accuracy and robustness of machine learning models in identifying sophisticated FDI attacks.

3.5 Impact of FDI on State Estimation and Power System Operations

False Data Injection (FDI) attacks critically threaten the stability of power systems by directly impacting state estimation, a fundamental process for grid monitoring and control. State estimation algorithms aggregate sensor data from across the grid to compute voltage levels, power flows, and load demand estimates, enabling grid operators to make informed operational decisions. However, when FDI attacks corrupt this data, state estimation becomes inaccurate, leading to incorrect control actions and system instabilities [8, 5, 6].

3.5.1 State Estimation and FDI Vulnerabilities

As [8] explains, FDI attacks deliberately target the measurement data fed into state estimation algorithms. This data, collected from smart meters, phasor measurement units (PMUs), and remote terminal units (RTUs), is crucial for ensuring that the grid operates within safe parameters. However, attackers can manipulate this data to mislead the state estimation process, injecting false information about grid conditions that may result in improper responses, such as incorrect dispatch of generation units or erroneous load shedding commands.

FDI attacks that bypass traditional Bad Data Detection (BDD) mechanisms have proven particularly problematic. [6] describes how attackers craft stealthy FDI vectors that fall within the system's tolerance thresholds, preventing BDD filters from flagging the altered data as anomalous. Consequently, system operators remain unaware of the compromised state, allowing the attacker to execute prolonged or repeated attacks without detection.

3.5.2 Operational Impact on Power Systems

FDI attacks have a profound impact on the operational efficiency and reliability of power systems. As demonstrated in [4], the consequences of FDI attacks can include:

- **Misdirection of generation resources:** Inaccurate state estimates cause the system to dispatch generation units inefficiently, leading to increased operational costs and suboptimal power flow.

- **Grid instability:** By distorting key parameters such as voltage levels and load demands, FDI attacks can lead to power oscillations, voltage collapse, and in extreme cases, blackouts.

- **Increased wear on equipment:** Constant operation under incorrect assumptions increases the mechanical and thermal stress on grid components, reducing their operational lifespan.

A notable example of the operational effects of FDI attacks is illustrated in the following table, summarizing key performance metrics before and after an FDI attack:

Table 3.2: Operational Impact of FDI on State Estimation and Power Systems [1], [2], [3], [4], [5], [6], [7], [8].

Metric	Normal Operation	Under FDI Attack
Voltage Deviation (p.u.)	± 0.01	± 0.15
Generation Dispatch Error (%)	$1 - 2\%$	$10 - 12\%$
Load Mismatch (MW)	≤ 5 MW	≥ 50 MW
Operational Cost (USD/hour)	1000	5000
System Frequency Deviation (Hz)	± 0.005	± 0.05

Table 3.3: Operational Impact of FDI on Power Grid Components

Component	Normal Value Example	Typical Change (% of Value)	Example Change	Source (Paper & Page)
VGM (Voltage Magnitude at Generator)	1.0 p.u.	0.5% to 10%	0.95–1.05 p.u.	[2], p. 2
PG (Real Power Generated)	100 MW	5% to 20%	90–120 MW	[5], p. 3
PL (Real Power Load)	200 MW	10% to 30%	140–260 MW	[3], p. 5
QL (Reactive Power Load)	50 MVAR	5% to 50%	25–75 MVAR	[9], p. 4

3.5.3 Jacobian Matrix and FDI Impact

One of the most significant ways FDI attacks affect state estimation is by manipulating the Jacobian matrix used in the estimation process. The Jacobian matrix relates system states (such as voltages and angles) to the measurements, such as active and reactive power flows. By injecting false data into key measurements, attackers effectively alter the structure of this matrix, leading to incorrect power flow calculations and misinformed operational decisions [2, 3, 6].

3.5.4 Cascading Effects on Grid Operations

FDI attacks not only disrupt state estimation but also initiate cascading failures across the grid. By continuously feeding erroneous data into control systems, attackers can destabilize grid voltage levels, leading to protective relays being triggered unnecessarily, causing widespread outages or equipment damage. The work by [12] highlights that descriptor system models can capture the cascading nature of these failures, modeling the long-term consequences of FDI attacks in both centralized and decentralized grid architectures.

In conclusion, FDI attacks pose a grave threat to the accuracy of state estimation and overall power system operations. Their ability to bypass detection mechanisms and induce operational inefficiencies makes them particularly dangerous for modern smart grids, requiring advanced detection and mitigation techniques to ensure grid security and reliability.

3.6 Detection Techniques and Challenges

Detecting False Data Injection (FDI) attacks presents several challenges, particularly as these attacks evolve in sophistication and stealth. Traditional methods such as Bad Data Detection (BDD) algorithms have been widely implemented in power systems to identify anomalies in state estimation data. However, as [6] and [8] point out, FDI attacks can often bypass these detection mechanisms by injecting carefully crafted false data that remains within the expected statistical range, avoiding detection flags.

One of the primary challenges in detecting FDI attacks lies in their stealthy nature. As discussed in [13], machine learning-based approaches have emerged as a promising technique for detecting FDI, particularly deep learning frameworks. These approaches can be trained on historical and real-time data to identify patterns of anomalies caused by FDI attacks. However, the implementation of such techniques comes with its own challenges, such as the need for large datasets and the potential for high computational overhead, which may hinder real-time detection in critical systems.

Additionally, [7] demonstrates that attackers can exploit historical data with limited information to mount FDI attacks, further complicating detection efforts. The use of historical data allows attackers to predict the system's normal behavior and inject false data that closely mimics the expected measurements, making it difficult for conventional detection systems to distinguish between normal and malicious data.

A significant challenge highlighted by [13] and [14] is the integration of defense mechanisms that can both detect and mitigate the effects of FDI attacks. While machine learning models provide promising results, their deployment in live power systems requires robust real-time processing capabilities. Furthermore, the development of hybrid detection models that combine hardware-based protections, such as those discussed by [14], with advanced data-driven methods is necessary to enhance system resilience.

Lastly, [12] discusses the potential for descriptor system frameworks to provide a more formal and mathematical approach to detecting and mitigating FDI attacks. These frameworks rely on modeling the system's dynamic behavior to identify deviations caused by malicious data injections. However, the complexity of implementing such systems in real-world power grids remains a challenge due to the variability in grid architecture and operational conditions.

In summary, detecting FDI attacks requires a comprehensive approach that combines traditional detection methods with advanced machine learning models, real-time processing, and hardware-based defenses. As FDI attacks become more sophisticated, detection mechanisms must also evolve, adapting to increasingly stealthy and complex threats.

3.7 Summary of Literature Review

The literature review provided a comprehensive analysis of various aspects of False Data Injection (FDI) attacks in power systems. The key findings from the 13 papers highlight the growing sophistication of FDI attacks and the critical need for advanced detection mechanisms and mitigation strategies.

FDI attacks pose a significant threat to state estimation, as demonstrated by [1] and [2]. These attacks manipulate sensor data used by state estimation algorithms, leading to incorrect operational decisions that can compromise grid stability and security. Stealth FDI attacks, which bypass traditional Bad Data Detection (BDD) systems, have been identified as particularly dangerous, allowing attackers to remain undetected for extended periods while compromising system reliability.

The review of machine learning approaches shows that Generative Adversarial Networks (GANs) and other deep learning models are being effectively employed to generate FDI attack vectors. The work by [9] and [3] demonstrates how these models can synthesize realistic attack scenarios, making it difficult for conventional detection systems to identify malicious activity. In addition, the use of data augmentation techniques, such as ADASYN, as shown in [3], improves the accuracy of machine learning models in detecting attacks.

Sparse FDI attacks, as explored in [5], involve injecting minimal but strategic data manipulations, causing large-scale disruptions without being easily detectable. These sparse attacks leverage optimization techniques to maximize their impact while minimizing the risk of detection. Additionally, studies such as [4] and [6] highlight the cascading operational effects of FDI attacks, including increased operational costs, incorrect generator dispatch, and load imbalances that lead to voltage collapses and grid instabilities.

A critical finding across multiple papers is the vulnerability of the Jacobian matrix used in state estimation to FDI attacks. By modifying key measurements, attackers can disrupt the structure of the Jacobian matrix, leading to incorrect power flow calculations and decisions. The impact of these attacks on state estimation is particularly significant, as highlighted by [8] and [12].

In conclusion, this literature review reveals that FDI attacks are an evolving threat that requires innovative detection methods and robust grid management strategies. The integration of machine learning with traditional grid operations will be essential in mitigating the risks posed by FDI attacks and ensuring the security of modern power systems. However, one of the main challenges in addressing FDI attacks is the lack of real-world attack data and a comprehensive understanding of how these attacks manifest in power systems. To effectively develop intrusion detection methods, generating realistic FDI attack data is crucial. This gap is the central focus of our current work, where we aim to create a robust dataset that captures a variety of FDI attack scenarios, facilitating the development of more accurate detection systems and contributing to the overall security of power grids.

Table 3.4: Summary of Key Literature on FDI Generation and Detection

Paper Title	Authors	Year	FDI Generation Technique	How to Generate FDI Data	Key Technical Details	Key Advantages	Challenges	Relevance to FDI Tool	Dataset Details	Algorithm Used	Publication Venue	FDI Model Complexity	Data Input Requirements	Novel Contributions	Potential Applications	Security Considerations
False Data Injection Attacks against State Estimation in Electric Power Grids	Yao Liu, Peng Ning, Michael K. Reiter	2009	Stealth FDI Attacks	Manipulate state variables (voltage, phase angles) to create undetectable FDI data.	Requires grid topology knowledge and linear models.	Simple to integrate.	Requires detailed system topology knowledge.	Foundational method for creating FDI data.	Voltage, Phase Angles	Linear estimation methods	ACM CCS '09	Moderate	Voltage, phase angles, power flows	Key foundational approach for FDI simulation	Cyber-Physical System Security	Mitigation Techniques for Real-Time Attacks
A Stealth False Data Attack on State Estimation with Minimal Network Information	Narang, Jagendra Kumar and Bag, Baidyanath	2019	GAN-based FDI	Use GANs to generate attack vectors without full network knowledge.	Trains DCGAN for stealthy FDI data generation.	Limited network information required.	Complex GAN training process.	Useful for stealth attacks.	No dataset used in this paper	GAN (Deep Convolutional GAN)	Electrical Power Systems Research	High	Real-time measurements	Stealth FDI attack with GAN	Grid Security Testing	Difficulty in detecting stealth attacks
A Data Preparation Method for Machine-Learning-Based Power System Cyber-Attack Detection	Chen, Hongyu et al.	2021	ML Data Augmentation	Augment normal data with rare FDI attack scenarios.	Uses Random Matrix Theory (RMT) and ADASYN for data preparation.	Enhances detection for rare scenarios.	Requires large datasets for training.	Useful for training ML models for FDI.	IEEE 118-Bus System	RMT and ADASYN	Journal of Electrical Power and Energy Systems	High	Historical data, system measurements	Augmentation improves rare attack detection	Smart Grid Monitoring Systems	Generalization challenges
Defense of Massive False Data Injection Attack via Sparse Attack Points	L. Wei et al.	2020	Sparse FDI Optimization	Generate FDI by targeting sparse attack points in the system.	Uses sparse optimization to create efficient FDI data.	Optimizes attack points for minimal intrusion.	Limited real-world applicability.	Useful for sparse FDI in large grids.	Synthetic dataset	Sparse Attack Optimization	IEEE Trans. Smart Grid	Moderate	Grid topology, system constraints	Sparse FDI simulation	Power Grid Control Systems	Efficiency in attack detection
iAttackGen: Generative Synthesis of False Data Injection Attacks	Md. Shohan et al.	2020	Generative FDI Models	Automate FDI data generation using generative models.	Employs generative models for FDI data creation.	Automates FDI generation process.	Requires deep learning expertise.	Useful for automating FDI in tools.	No dataset provided	GAN (Generative Adversarial Network)	ACM/IEEE ICCPS	High	GAN training data	Generative approach for real-time simulations	FDI Simulators for Power Systems	Training complexity for deep models
Multi-Objective Optimization on Stealthy FDI in Smart Power Transmission	Z. Li et al.	2021	Multi-Objective FDI Optimization	Optimize FDI for stealth and impact.	Uses multi-objective optimization methods.	Highly flexible attacks.	Complex optimization process.	Useful for optimizing FDI attacks in grid.	IEEE 118-Bus Dataset	Multi-objective optimization methods	IEEE Access	High	System parameters, constraints	Multi-objective optimization	Power Transmission Optimization	Computational costs in optimization
Vulnerability Analysis and Consequences of False Data Injection Attack on Power System State Estimation	J. Liang et al.	2016	Stealth FDI Attack	Analyze vulnerabilities in state estimation for FDI.	Examines weaknesses in state estimation techniques.	Provides insights into power system vulnerabilities.	Focus on vulnerabilities with less focus on countermeasures.	Valuable for identifying FDI weaknesses.	IEEE 118-Bus Dataset	State Estimation Analysis	IEEE Trans. Power Systems	High	Voltage, state estimation variables	Vulnerability assessment in state estimation	FDI Detection in Power Systems	Focus on vulnerability analysis
A Deep Learning Framework to Identify Remedial Action Schemes Against False Data Injection Cyberattacks	Naderi, Ehsan and Asrari, Arash	2024	Deep Learning-Based	Identifies schemes to counter FDI.	Uses deep learning to identify attack points.	Advances in detecting attacks.	Requires complex ML models.	Relevant for ML-based FDI detection.	Simulated dataset	Deep learning models	IEEE Trans. Industrial Informatics	High	Power system data	Framework for detecting attack schemes	Smart Grid Cybersecurity	High computational requirements
A Descriptor System Design Framework for FDI Attack Toward Power Systems	Aibing Qiu et al.	2021	Descriptor-Based FDI	System design framework for generating FDI	Uses descriptor models for attack detection	High precision in attack detection	Descriptor systems are complex to train	Important for understanding FDI system impacts	Simulated data	Descriptor system model	Electric Power Systems Research	High	Grid data, real-time measurements	System design framework for FDIs	Power system state monitoring	Computational requirements for descriptor systems
Can Attackers With Limited Information Exploit Historical Data to Mount Successful FDI Attacks?	Zhang, Jiazi et al.	2018	Historical Data Exploitation	Uses historical data to simulate FDI	Employs methods to exploit historical information	Effective for simulating stealth attacks	Limited to available historical data	High relevance for FDI tools	IEEE 118-Bus dataset	Historical data analysis	IEEE Trans. Power Systems	Moderate	Historical grid data	Attack simulations with limited information	Intrusion detection testing	Need for diverse historical datasets
Designing False Data Injection Attacks Penetrating AC-Based Bad Data Detection	Minh N. Tran et al.	2020	AC-Based FDI	Design FDI attacks that bypass AC-based detection.	Uses AC-based techniques to generate attack data.	Effective at bypassing AC detection.	Requires AC system information.	Important for designing resilient FDI tools.	Synthetic dataset	AC-based simulation	Concurrency and Computation	High	AC system data	Focus on bypassing AC detection systems	Power Grid Attack Simulation	Challenges in AC-based system complexity
Multi-Objective False Data Injection Attacks of Cyber-Physical Power Systems	Kang-Di Lu and Zheng-Guang Wu	2022	Multi-Objective FDI	Multi-objective optimization	Uses optimization to improve attack stealth	Complex optimization requirements	Highly relevant for generating stealthy attacks	Simulated dataset	Multi-objective optimization methods	IEEE Trans. Circuits and Systems II	High	Power system parameters	Multi-objective attack optimization	Power Grid Cybersecurity	Computation costs in optimization	Stealthy attacks increase grid instability and risk of cascading failures, requiring advanced detection systems with potential computational overhead.
A Remedial Action Scheme Against FDI Cyberattacks in Smart Transmission Systems	Ehsan Naderi et al.	2022	Remedial Action Schemes	Identifies action schemes against FDI	Uses thyristor-controlled series capacitors (TCSCs) to counter FDIs	High resilience to attack	Requires advanced knowledge of TCSCs	Useful for designing defense mechanisms	Synthetic dataset	Remedial action framework	IEEE Trans. Industrial Informatics	High	Power system parameters	TCSC-based defense mechanisms	Power System Cybersecurity	High complexity for implementation

Chapter 4

Power System Components Susceptible to FDI Attacks

4.1 Introduction

Power systems rely on accurate measurements and state estimation to ensure reliable and economic operation. False Data Injection (FDI) attacks target the integrity of these measurements by injecting maliciously crafted data, leading to severe operational disruptions. This chapter explores the power system components most susceptible to FDI attacks, including voltage magnitude and angle, real and reactive power, and load measurements, with insights drawn from the relevant literature.

4.2 Voltage-Related Components

4.2.1 Voltage Magnitude at Generator (VGM)

Voltage magnitude is a crucial parameter in power system operation. As Narang et al. [2] explain, FDI attacks can alter generator voltage magnitudes by up to 10%, typically forcing values between 0.95 and 1.05 p.u. for a nominal value of 1.0 p.u. This alteration affects generator excitation systems and can lead to overloading or underutilization of equipment. The paper highlights how such manipulations remain stealthy under conventional bad data detection (BDD) mechanisms.

4.2.2 Voltage Angle at Line and Generator

Wei et al. [5] emphasize the importance of voltage phase angles in power flow calculations. By manipulating phase angle differences, attackers can mislead state estimation algorithms. For example, a deviation of $\pm 1 canresultinpowerflowchangesexceeding 5\%, potentiallydestabili$

4.3 Real Power Components

4.3.1 Real Power Generated (PG)

Real power generation is directly impacted by FDI attacks, as Wei et al. [5] discuss. The paper outlines scenarios where power generation values are altered by up to 20%, leading to incorrect dispatch signals. For instance, a generator scheduled to produce 100 MW may instead operate at 90 MW or 120 MW under attack, causing generation-load imbalances and increased operational costs.

4.3.2 Real Power Load (PL)

Chen et al. [3] demonstrate that load measurements are particularly vulnerable to synthetic data injection. Their study shows deviations ranging from 10% to 30%, which translate to operational mismatches of 50 MW or more in typical scenarios. Such changes disrupt economic load dispatch and force the grid to operate inefficiently.

4.4 Reactive Power Components

4.4.1 Reactive Power Load (QL)

Reactive power loads, essential for voltage stability, are highly susceptible to FDI attacks. Shohan et al. [9] highlight that reactive power loads can deviate by as much as ±50% under attack, significantly altering grid stability margins. For example, a reactive load of 50 MVAR can be manipulated to values as low as 25 MVAR or as high as 75 MVAR, leading to potential voltage collapses.

4.5 Operational Impact and Cascading Failures

Liang et al. [8] and Zhang et al. [7] discuss the cascading effects of FDI attacks. Manipulations in voltage and power measurements propagate through the grid, causing widespread failures. Their studies reveal that even small changes in key parameters can amplify over time, affecting multiple subsystems. For instance, altered power flow calculations can lead to incorrect overload protections, triggering line outages and blackouts.

4.6 Advanced Threat Scenarios

Shohan et al. [9] propose a generative framework, iAttackGen, to model advanced FDI attacks. Using generative adversarial networks (GANs), their method synthesizes stealthy attack vectors that evade detection while maximizing grid disruption. Similarly, Narang et al. [2] explore minimal-information attacks, showing that attackers with limited network topology knowledge can still execute effective FDI scenarios by targeting critical measurement points.

4.7 Detection Challenges

FDI attacks exploit vulnerabilities in state estimation and SCADA systems, as noted by Tran et al. [6]. Detection mechanisms must account for stealthy changes that align with expected operational patterns. The complexity of modern grids, coupled with limited computational resources, makes real-time detection particularly challenging.

4.8 Conclusion

This chapter provided an in-depth examination of power system components vulnerable to FDI attacks, including voltage magnitude and angle, real and reactive power, and load measurements. Insights from the literature reveal the far-reaching implications of these attacks, highlighting the need for advanced detection and mitigation strategies.

Chapter 5

Methodology

This chapter outlines the methods employed in generating and analyzing False Data Injection (FDI) attacks using Random Matrix Theory (RMT) and Synthetic Minority Over-sampling Technique (SMOTE). We use power system attack datasets from [11] to investigate the behavior of FDI and generate synthetic attack data for further analysis. The key steps involved include RMT for anomaly detection and SMOTE for balancing the data distribution.

5.1 RMT and SMOTE Approach

FDI data generation and detection are critical aspects of maintaining grid security, as discussed in [1], [2], and [3]. Our approach to anomaly detection leverages RMT, while SMOTE is applied to create synthetic data to augment the dataset.

5.1.1 Dataset Description

The dataset used in this methodology is from the Power System Attack Datasets - Mississippi State University and Oak Ridge National Laboratory, as described in [11]. Specifically, we used features 7 to 12 from the dataset, which correspond to FDI data, including voltage phase angles, current phase angles, and magnitudes.

5.1.2 Random Matrix Theory (RMT) for Anomaly Detection

RMT has been successfully applied in various anomaly detection applications in power systems [5]. The idea behind RMT is to calculate the eigenvalues of the covariance matrix of the data and use thresholds to identify anomalous eigenvalues, which may correspond to anomalies in the data.

The steps involved in applying RMT are as follows:

```
# Standardizing the data
scaler = StandardScaler()
standardized_data = scaler.fit_transform(data)

# Computing the covariance matrix
covariance_matrix = np.cov(standardized_data, rowvar=False)

# Calculating eigenvalues
eigenvalues, _ = np.linalg.eigh(covariance_matrix)
```

The eigenvalues are then compared against the calculated RMT thresholds to identify anomalous indices.

```
# RMT threshold calculation
lambda_max = (1 + np.sqrt(p / n)) ** 2
lambda_min = (1 - np.sqrt(p / n)) ** 2

# Identifying anomalies
anomalous_indices = np.where((eigenvalues < lambda_min) |
    (eigenvalues > lambda_max))[0]
```

The full implementation of this approach is included in the appendix under the file `RMT.ipynb`.

5.1.3 SMOTE for Synthetic Data Generation

To augment the dataset, SMOTE is applied to generate synthetic FDI data [3]. This method over-samples the minority class by generating new samples between existing data points, helping balance the dataset.

After applying RMT, we observed that the eigenvalue distributions showed some anomalies, indicating that the original dataset was unbalanced. To address this, we applied SMOTE to the dataset. The specific code for applying SMOTE is as follows:

```
# Applying SMOTE to the dataset
smote = SMOTE()
X_resampled, y_resampled = smote.fit_resample(data, labels)
```

This technique increased the dataset from 9,583 rows to 17,063 rows. However, as we will discuss in the next section, the similarity between the original and generated data varied across different columns, indicating that while SMOTE was effective in balancing the data, achieving high similarity for all features remains challenging.

5.2 Similarity Metrics

To evaluate the similarity between the original and synthetic datasets, we calculated similarity metrics using the method implemented in `compare.ipynb`. In the code, the similarity between columns was computed using a combination of Euclidean distance and cosine similarity measures. This allowed us to quantitatively assess how well the generated data matched the original data across different features. The results of the similarity metrics are shown in Table 5.1.

Table 5.1: Similarity Metrics for R1, R2, R3, and R4

R1 Similarity Metrics		**R2 Similarity Metrics**		**R3 Similarity Metrics**		**R4 Similarity Metrics**	
Feature	**Similarity**	**Feature**	**Similarity**	**Feature**	**Similarity**	**Feature**	**Similarity**
R1-PA1:VH	0.9491	R2-PA1:VH	0.9438	R3-PA1:VH	0.9881	R4-PA1:VH	0.9121
R1-PM1:V	0.0	R2-PM1:V	0.0	R3-PM1:V	0.0	R4-PM1:V	0.0
R1-PA2:VH	0.9317	R2-PA2:VH	0.8657	R3-PA2:VH	0.9168	R4-PA2:VH	0.8655
R1-PM2:V	0.0	R2-PM2:V	0.0	R3-PM2:V	0.0	R4-PM2:V	0.0
R1-PA3:VH	0.9847	R2-PA3:VH	0.9426	R3-PA3:VH	0.9934	R4-PA3:VH	0.9364
R1-PM3:V	0.0	R2-PM3:V	0.0	R3-PM3:V	0.0	R4-PM3:V	0.0
R1-PA4:IH	1.0	R2-PA4:IH	0.9676	R3-PA4:IH	1.0	R4-PA4:IH	0.9819
R1-PM4:I	0.1017	R2-PM4:I	0.0497	R3-PM4:I	0.1322	R4-PM4:I	0.0372
R1-PA5:IH	0.9673	R2-PA5:IH	0.9789	R3-PA5:IH	1.0	R4-PA5:IH	0.9324
R1-PM5:I	0.213	R2-PM5:I	0.06	R3-PM5:I	0.1735	R4-PM5:I	0.0379
R1-PA6:IH	1.0	R2-PA6:IH	0.9388	R3-PA6:IH	1.0	R4-PA6:IH	1.0
R1-PM6:I	0.1711	R2-PM6:I	0.0535	R3-PM6:I	0.242	R4-PM6:I	0.0547
R1-PA7:VH	0.9565	R2-PA7:VH	0.9273	R3-PA7:VH	0.9692	R4-PA7:VH	0.9085
R1-PM7:V	0.0	R2-PM7:V	0.0	R3-PM7:V	0.0	R4-PM7:V	0.0
R1-PA8:VH	0.2745	R2-PA8:VH	0.1788	R3-PA8:VH	0.1634	R4-PA8:VH	0.2366
R1-PM8:V	0.033	R2-PM8:V	0.0157	R3-PM8:V	0.0182	R4-PM8:V	0.1234
R1-PA9:VH	0.2925	R2-PA9:VH	0.2819	R3-PA9:VH	0.2464	R4-PA9:VH	0.2312
R1-PM9:V	0.1432	R2-PM9:V	0.017	R3-PM9:V	0.0182	R4-PM9:V	0.3176
R1-PA10:IH	1.0	R2-PA10:IH	0.9641	R3-PA10:IH	1.0	R4-PA10:IH	0.9746
R1-PM10:I	0.2376	R2-PM10:I	0.0351	R3-PM10:I	0.2475	R4-PM10:I	0.0580
R1-PA11:IH	0.8878	R2-PA11:IH	0.8767	R3-PA11:IH	0.9314	R4-PA11:IH	0.8910
R1-PM11:I	0.8831	R2-PM11:I	0.4876	R3-PM11:I	0.8395	R4-PM11:I	0.4966
R1-PA12:IH	0.9331	R2-PA12:IH	0.9166	R3-PA12:IH	0.9231	R4-PA12:IH	0.8742
R1-PM12:I	0.9899	R2-PM12:I	0.5852	R3-PM12:I	0.8966	R4-PM12:I	0.5403
R1:F	0.0493	R2:F	0.0	R3:F	0.019	R4:F	0.0355
R1:DF	0.0174	R2:DF	0.0381	R3:DF	0.0377	R4:DF	0.0305
R1-PA:Z	0.26	R2-PA:Z	0.2326	R3-PA:Z	0.2749	R4-PA:Z	0.2099
R1-PA:ZH	0.0405	R2-PA:ZH	0.0008	R3-PA:ZH	0.0009	R4-PA:ZH	0.0329
R1:S	0.75	R2:S	1.0	R3:S	1.0	R4:S	1.0

As seen in the table, several features, such as `R1-PA1:VH`, `R1-PA2:VH`, and `R1-PM1:V`, show good similarity (above 90%), while others, like `R1-PA5:IH` and `R1-PM6:I`, exhibit lower similarity, often below 85%. Notably, the Voltage Phase Angle and Phase Angle features had a similarity above 0.9 most of the time and consistently above 0.85, indicating high-quality generation for these features. However, other features, such as current magnitude, showed much more random and inconsistent similarity, highlighting areas where the synthetic data could be improved.

5.3 Results and Discussion

The combination of Random Matrix Theory (RMT) and Synthetic Minority Over-sampling Technique (SMOTE) was effective in generating a larger dataset with more balanced classes. However, as shown in Table 5.1, not all features achieved high similarity between the original and generated data:

- **Voltage Phase Angle and Phase Angle Features**: The features related to voltage and phase angles consistently showed a similarity score of above 0.9 and were consistently above 0.85. This suggests that the synthetic data generation worked particularly well for these columns, which are essential for detecting changes in the power system under FDI attacks.

- **Current Magnitude Features**: Features such as current magnitude showed significant variance, with some similarity scores dropping below 0.5. This suggests that the SMOTE technique was less effective in generating accurate synthetic data for these columns, potentially due to the more complex nature of current-related data in power systems.

- **Eigenvalue Distribution**: Using RMT, the majority of eigenvalues fell within the expected range, identifying normal behavior, with a few outliers detected as potential anomalies. These outliers correspond to features with lower similarity, indicating possible system irregularities or difficulties in data generation.

- **FDI Data Generation**: This method has proven useful for generating additional attack data, particularly for features with high similarity scores, which can be used in training machine learning models for detecting stealthy FDI attacks [9].

The approach was successful in generating synthetic data that closely matches the original data for several key features. However, further refinement is necessary to achieve consistent similarity across all features, especially for current magnitude, which exhibited random and inconsistent similarity scores.

5.4 Conclusion

In this chapter, we have described the methodology for applying RMT and SMOTE to FDI data, a critical step in enhancing the understanding and detection of FDI attacks in power systems. By using a combination of RMT to detect anomalies and SMOTE to generate additional synthetic data, we augmented the dataset and ensured better class balance for the detection models. The similarity metrics showed that Voltage Phase Angle and Phase Angle features achieved high similarity (above 0.9), demonstrating the effectiveness of this approach for certain features. However, other columns, particularly current magnitude features, showed random and inconsistent similarity, indicating the need for further refinement in data generation techniques.

This approach, while generating data that is mostly similar to the original, highlights the need for improvement in certain areas. Future work will explore additional data augmentation techniques and further tuning of SMOTE parameters to improve the quality of synthetic data for all features. In the next chapter, we will explore additional techniques for generating and detecting FDI attacks and compare their performance against the RMT and SMOTE-based approach.

Chapter 6

Design of the FDI Simulation Tool

Chapter 7

Results and Evaluation

7.1 Results Analysis

Chapter 8

Discussion

Chapter 9

Summary and Conclusion

This section summarizes the quintessential outcomes of our proposed methodology.

Chapter 10

Appendix A: Code for RMT and Similarity Calculation

This appendix contains the Python code used for the implementation of Random Matrix Theory (RMT) for anomaly detection and the comparison of similarity metrics between the original and synthetic datasets.

The code files are located inside the folder `code` and are described below:

- `RMT.ipynb`: This file contains the implementation of the Random Matrix Theory (RMT) approach used for anomaly detection in the dataset.
- `compare.ipynb`: This file contains the code used to calculate similarity metrics between the original and generated data using measures like Euclidean distance and cosine similarity.

The full code files can be found in the folder named `code`, attached with this report. Below are brief code snippets and explanations of key parts of the implementation.

10.1 RMT Implementation (RMT.ipynb)

The following snippet shows the key part of the RMT implementation, which involves calculating eigenvalues, identifying anomalous features, and applying thresholding to determine anomalies:

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import SMOTE

def apply_rmt(data):
    scaler = StandardScaler()
    standardized_data = scaler.fit_transform(data)
```

```
    covariance_matrix = np.cov(standardized_data, rowvar=False)


    eigenvalues, _ = np.linalg.eigh(covariance_matrix)


    n, p = data.shape
    lambda_max = (1 + np.sqrt(p / n)) ** 2
    lambda_min = (1 - np.sqrt(p / n)) ** 2

    anomalous_indices = np.where((eigenvalues < lambda_min) |
     (eigenvalues > lambda_max))[0]
    normal_indices = np.where((eigenvalues >= lambda_min) &
     (eigenvalues <= lambda_max))[0]


    print(f"\nRMT Thresholds: Lambda Min = {lambda_min},
     Lambda Max = {lambda_max}")
    print(f"\nAnomalous Eigenvalues: {eigenvalues[anomalous_indices]}")
    print(f"\nNormal Eigenvalues: {eigenvalues[normal_indices]}")

    return eigenvalues, anomalous_indices, normal_indices, lambda_min,
    lambda_max


file_path = 'filtered_data.csv'
data = pd.read_csv(file_path)

data = data.select_dtypes(include=[np.number])
data = clean_data(data)

eigenvalues, anomalous_indices, normal_indices, lambda_min,
lambda_max = apply_rmt(data.values)


data['Anomaly'] = 0
valid_anomalous_indices = [i for i in anomalous_indices if i < len(data)]
valid_anomalous_indices = [i for i in valid_anomalous_indices
    if i in data.index]  # To Ensure indices are in DataFrame's index
```

```
data.loc[valid_anomalous_indices, 'Anomaly'] = 1


X = data.drop('Anomaly', axis=1)
y = data['Anomaly']


scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)


smote = SMOTE(sampling_strategy='minority',k_neighbors=3,
    random_state=42)
X_resampled, y_resampled = smote.fit_resample(X_scaled, y)


resampled_data = pd.concat([pd.DataFrame(X_resampled,
    columns=X.columns), pd.DataFrame(y_resampled, columns=['Anomaly'])],
        axis=1)

balanced_data_path = 'balanced_data_with_rmt_smote.csv'
balanced_data = pd.read_csv(balanced_data_path)

balanced_data_multiplied = balanced_data * 100

balanced_data_multiplied.to_csv('data_rmt_smote.csv', index=False)
```

The eigenvalues are calculated from the covariance matrix of the standardized dataset. Using Random Matrix Theory, thresholds λ_{min} and λ_{max} are computed to classify the eigenvalues as normal or anomalous.

10.2 Similarity Calculation (compare.ipynb)

The following snippet shows how similarity between the original and generated datasets is calculated, using both Euclidean distance and cosine similarity:

```
from sklearn.metrics.pairwise import cosine_similarity
import numpy as np

```

```
def calculate_similarity(original_data, generated_data):
    # Euclidean distance
    euclidean_dist = np.linalg.norm(original_data - generated_data)

    # Cosine similarity
    cosine_sim = cosine_similarity([original_data],
    [generated_data])[0][0]

    return euclidean_dist, cosine_sim

# Example usage
similarity_results = []
for column in original_data.columns:
    original_col = original_data[column].values
    generated_col = generated_data[column].values
    euclidean_dist, cosine_sim = calculate_similarity(original_col,
    generated_col)
    similarity_results.append((column, euclidean_dist, cosine_sim))
```

In this code, each feature from the original and generated datasets is compared, and the results (Euclidean distance and cosine similarity) are stored in a list. This method helps quantify the similarity between the two datasets, identifying which features exhibit good similarity and which do not.

10.3 Conclusion

These code files form the core of the implementation used in this report for anomaly detection with RMT and comparison of similarity metrics. The RMT code identifies anomalous features, while the similarity calculation code measures how closely the generated data matches the original data across various features.

Bibliography

[1] Yao Liu, Peng Ning, and Michael K. Reiter. False data injection attacks against state estimation in electric power grids. In *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS '09)*, pages 21–32, 2009.

[2] Jagendra Kumar Narang and Baidyanath Bag. A stealth false data attack on state estimation with minimal network information. *Electrical Power Systems Research*, 166:105971, 2019.

[3] Hongyu Chen, Jingyu Wang, and Dongyuan Shi. A data preparation method for machine-learning-based power system cyber-attack detection. *Journal of Electrical Power and Energy Systems*, 133:107041, 2021.

[4] Z. Li, X. Hu, M. Zhang, and Y. Sun. Multi-objective optimization on stealthy false data injection attack in smart power transmission grid. *IEEE Access*, 9:17875–17885, 2021.

[5] L. Wei, P. Wang, J. Wu, and D. Choi. Defense of massive false data injection attack via sparse attack points considering uncertain topological changes. *IEEE Transactions on Smart Grid*, 11(1):283–293, 2020.

[6] Minh N. Tran, Huynh T. Le, and Van-Son Tran. Designing false data injection attacks penetrating ac-based bad data detection. *Concurrency and Computation*, 32(24):e5977, 2020.

[7] Lalitha Sankar Oliver Kosut Jiazi Zhang, Zhigang Chu. Can attackers with limited information exploit historical data to mount successful false data injection attacks on power systems? *IEEE Transactions on Power Systems*, 33(5):4775–4786, 2018.

[8] Oliver Kosut Jingwen Liang, Lalitha Sankar. Vulnerability analysis and consequences of false data injection attack on power system state estimation. *IEEE Transactions on Power Systems*, 31(5):3864–3872, 2016.

[9] Md. Shohan, M. Mohsin, S. Hasan, and A. Mahmood. iattackgen: Generative synthesis of false data injection attacks in cyber-physical systems. In *Proceedings of the 12th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS '21)*, pages 240–250, 2021.

[10] F. Fioretto, T. W. Mak, and P. Van Hentenryck. Predicting ac optimal power flows: Combining deep learning and lagrangian dual methods. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 630–637, 2020.

[11] University of Washington. IEEE 118-Bus System Data, 1993. Accessed: 2024-07-29.

[12] Shengfeng Wang Aibing Qiu, Zhou Ding. A descriptor system design framework for false data injection attack toward power systems. *Electric Power Systems Research*, 192:106932, 2021.

[13] Ehsan Naderi and Arash Asrari. A deep learning framework to identify remedial action schemes against false data injection cyberattacks targeting smart power systems. *IEEE Transactions on Industrial Informatics*, 20(2):1208–1219, 2024.

[14] Ehsan Naderi, Arash Asrari, and Saman Pazouki. A remedial action scheme against false data injection cyberattacks in smart transmission systems: Application of thyristor-controlled series capacitor. *IEEE Transactions on Industrial Informatics*, 18(4):2297–2309, 2022.