Politecnico di Torino

Department of Electronics and Telecommunications

Static and Dynamic False Data Injection Attacks on State Estimation in Smart Grids

Author: Abubacarr Ceesay
Student ID: s310138
Supervisor: Professor Tao Huang

ICT for Smart Societies Master's Degree Programme

Date: March, 2025

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The power grid, in short, is a network installed to distribute electrical energy from production plants to consumers efficiently. From its inception, however, it has undergone significant transformations to enhance how electricity is distributed among communities. Initially, it was designed as a centralized system where electricity flowed in a unidirectional manner, from large power plants to consumers. This led to a centralized structure that clearly distinguished producers from consumers. Technological advancements in electrical energy production led to a bidirectional flow of energy. Consumers in the modern era can also act as producers, generating electricity, for example, through Solar Panels and feeding it to the grid, leading to a more complex architecture.

With the rapid advancement in technology and increasing demand for more efficient and sustainable energy systems, the grid transitioned towards a modernized system integrating digital communication technologies, enabling a bidirectional flow of electricity.

Unlike the traditional grid, the Smart Grid allows consumers to act as "Prosumers", producing and consuming electricity. An example is households or businesses with solar panels that can generate electricity and feed surplus power into the grid. This bidirectional energy transfer improves grid efficiency, supports the integration of renewable energy sources, and provides more flexibility in balancing supply and demand.

While this evolution has brought about many benefits, including improved reliability, energy savings, and enhanced sustainability, it also introduces new challenges. The increased reliance on digital communication and interconnectedness has made the Smart Grid more vulnerable to cyber threats. As information systems become deeply rooted in the management and operation of the grid, the risk of cyber-attacks increases. These cyber threats, which target the data and control systems, can severely undermine the operational efficiency of the power grid, leading to substantial financial losses and degraded service quality for consumers.

One of these forms of threats to power grid infrastructure is physical attacks targeting field equipment. Grid operators generally install devices across a wide geographical area and manage them remotely from centralized control centers. While basic physical security measures are often in place, many of these sites such as substations, are not thoroughly secured, rendering them vulnerable to unauthorized physical access. Once on-premises, adversaries may exploit these devices to launch cyberattacks against the

broader grid infrastructure. In this context, there have been documented incidents of physical intrusions into substations with some driven by motives to steal and resell valuable components. However, malicious actors could use this physical access as an entry point to compromise system integrity, manipulate data, or disrupt operations. This convergence of physical and cyber vulnerabilities highlights the need for integrated security strategies that address both dimensions of threat exposure[20].

Another prevalent form of threat to power grid infrastructure is the Denial-of-Service (DoS) attack, which specifically targets the availability of network services. The primary objective of such an attack is to disrupt or degrade system functionality, thereby denying legitimate users access to critical services. DoS attacks are often employed as cover, masking more sophisticated intrusions to amplify the overall impact of other forms of attack. These attacks typically exploit vulnerabilities in the communication layer by overwhelming essential devices with excessive traffic. Using various network tools, attackers generate and transmit large volumes of forged or malicious data to targeted components, exhausting computational or bandwidth resources thereby making a service temporarily or permanently inaccessible. A more advanced variant of this threat is the Distributed Denial-of-Service (DDoS) attack, where multiple compromised sources are put together simultaneously to flood a single target, significantly increasing the attack's scale and effectiveness. Concerningly, several legacy communication protocols still in use within power grid infrastructures have been identified as vulnerable to such exploits, highlighting the urgent need for enhanced resilience and modernized infrastructure[24].

One of the most critical cyber threats to the Smart Grid is the False Data Injection (FDI) Attack. An FDI attack is a type of data integrity attack that exploits vulnerabilities in the grid's measurement and communication systems. During such an attack, the attacker maliciously alters or injects false data into the measurement readings (such as power injection data at nodes) sent to the grid's control center, generally known as Supervisory and Data Acquisition(SCADA). These compromised measurements lead to incorrect state estimation, which refers to the process through which the control center determines the current operating condition of the grid based on measured data received from sensors. Since state estimation is crucial for grid stability, wrong estimates can have dangerous consequences, such as improper decisions on power dispatch, load balancing, or even triggering blackouts. As a result, FDI attacks can severely disrupt grid operations, degrade service quality, and cause significant financial and reputational damage to utilities[7].

In summary, while the shift from a traditional centralized grid to a more dynamic and intelligent Smart Grid has offered numerous advantages, it also comes with heightened risks from cyber threats. Among these threats, False Data Injection Attacks are particularly dangerous because they manipulate the grid's data flow, leading to incorrect system operations and decisions that threaten the grid's stability and reliability.

1.2 History and Motivation

In recent years, notable attacks have been made on the electrical grid infrastructure. A well-known one is the 2015 attack on the Ukraine power grid, which led to a blackout lasting for several hours, with about 225,000 customers affected. The attack was im-

plemented by hijacking the SCADA systems through phishing emails and subsequently shutting down power stations across various regions. The timeline of the attack started with spear phishing to harvest credentials through the Black Energy Malware. Remote access through a Virtual Private Network(VPN) to the Industrial Control Systems(ICS) followed, allowing the attackers to send commands to connected stations digitally. To delay restoration efforts, the logs were consistently erased with KillDisk a malware crafted to erase the master boot record and thus leading to a device being unable to reboot. Finally, a denial of service on the call center followed, further delaying restoration. Subsequently, upon analysis of the attack, the Electricity Information Sharing and Analysis Center (E-ISAC) gave a couple of recommendations for future defence against possible attacks. These included the segmentation of networks, ensuring consistent logging to all connected devices, limiting remote connections and implementing an event monitoring system over the SCADA systems[23].

A similar incident occurred in the United States in 2018, as hackers tampered with the electrical grid network in California. It was not as severe as the Ukraine attack and did not lead to a blackout. The Department of Energy(DOE) handled the situation and avoided any severe impact on the services of the grid.

Another significant incident occurred in 2022 when approximately 30,000 satellite communication (SATCOM) terminals were compromised in a coordinated cyberattack. Among the affected devices were SATCOM modems deployed in wind turbines operated by ENERCON, a prominent German wind energy company. The attack had wide implications, as the impacted turbines collectively accounted for an estimated 10 gigawatts of electricity generation capacity. It was estimated that around 5,800 ENERCON turbines relied on these modems for communication.

Although the turbines continued to operate, the cyberattack rendered it impossible to remotely monitor and manage them using the Supervisory Control and Data Acquisition (SCADA) system, which is vital for the efficient and secure operation of modern energy infrastructure. The initial breach targeted a subsidiary of the U.S.-based SATCOM provider Viasat. This attack indirectly had effects on ENERCON's wind farms, which depended on the KA-SAT satellite for connectivity, particularly in remote areas lacking reliable mobile network coverage[13].

In 2021, a separate incident occured with Vestas Wind Systems A/S, an energy solutions company when they experienced a significant cybersecurity breach. Vestas specializes in the development, manufacturing, installation, and maintenance of both onshore and offshore wind turbines, with over 145 gigawatts of combined installed capacity worldwide. The attack involved unauthorized access to the company's core IT infrastructure, through which the attackers were able to infiltrate internal systems and extract a considerable amount of confidential and proprietary data. The attackers upon retrieval of such sensitive information proceeded to issue a ransom threat, of publishing the stolen data. This form of extortion poses huge financial and reputational risks to the company along with raising broader concerns about the vulnerability of critical infrastructure in the renewable energy sector. Vestas subsequently responded by shutting down affected IT systems and initiating forensic investigations. The incident underscored the growing cybersecurity challenges faced by energy companies operating in increasingly digitalized and interconnected environments as well as highlighting the potential consequences of cyber intrusions[27].

In light of these frequent attacks, the motivation of the thesis is to assess the vulnerabilities of the power grid system and find ways to enhance its reliability. In this work, we specifically study the False Data Injection attack and its impact on the state estimation, a method widely used in the control center to effectively monitor the state of the grid and maintain reliable operating conditions. State estimation, in conjunction with bad data detection algorithms, continues to be a widely employed method for ensuring data integrity and facilitating the reliable operation of the power grid. Electricity is the backbone of the technological infrastructure as every moment of an outage significantly impacts all other sectors such as industries, transportation networks and the like. The community reaches a standstill, leading to substantial financial loss, system failures and public safety risks. Some of these impacts from a successful impact could take ample time to recover, taking us a step back as a society and hindering technological advancements. With the recent advancements in AI and modern tools, we assess the vulnerability by leveraging classical attack models such as the stealthy attack. By coupling these with Generative AI, we explore the effect of such modern crafted attack models against state estimation.

1.3 Literature Review

Studies on FDI attacks are generally classified into two main groups. One faction usually plays the role of an attacker in assessing the vulnerability of the current systems in place to mitigate FDI attacks. The others are focused on the detection and mitigation strategies, validating them with well-known attack models.

One of these analyzed the effect of FDI attacks on market operations, examining the financial risks involved. First, a malicious attack against state estimation leading to monetary irregularities was crafted, and strategies for profitable attacks were implemented. By leveraging the day-ahead and ex-post real-time prices, an evaluation of the economic impact of the attacks was carried out. The study leveraged the Ex-Ante Real-Time Market by conducting security-constrained economic dispatch(SCED) at 10 to 15-minute intervals. This allows the optimal amount of power needed to be generated for a given load to be obtained. This process results in a dispatch order representing generators that directly impact the pricing structure. In the attack model, the attacker compromises sensors in such a way as to buy and sell virtual power at specific locations for a certain price in the day-ahead market. After this, the attacker injects malicious data, modifying the pricing structure in such a way as to sell power again at the compromised locations and thereby obtain profits as a difference between the initial and later prices[30].

An alternative approach of modelling FDI attacks leverages Long Term Short Term memory networks(LSTM). An LSTM is a type of recurrent neural network(RNN) in terms of deep learning. The RNN was leveraged as it had the capability of handling dynamic data. An RNN is different from other forms of neural networks in that the output is not only dependent on the current input but that of the previous input as well. In the context of FDI attack detection, this temporal dependency is especially beneficial, as measurement data used to craft successful attacks often show strong correlations over time. This inherent architecture allows the RNNs to thrive well with FDI attack detections, given that the measurement data used in crafting successful attack vectors is commonly

correlated. Among various RNN architectures, LSTM models indicate enhanced performance, particularly in tasks related to data evolving dynamically, as demonstrated in the research[32].

Recurrent neural networks extend conventional feedforward architectures by incorporating feedback loops that allow internal state information to persist across time steps. At time t, the network processes input x_t through a hidden layer A, producing output h_t. The hidden state encapsulates context from previous time steps, enabling the model to learn features dependent on the time varying nature of the data. However, this architecture leads to huge overfitting, which negatively affects model performance. Overfitting is the situation where the model performs extremely well with training data but has less accuracy when exposed to novel data. Instead of learning patterns of characteristics of the data, the model rather memorises the training data which is an undesired situation. To overcome this limitation, the use of dropout has been proposed. However, as noted in [26], applying dropout directly can lead to excessive propagation of input noise, which can hinder convergence. Therefore, the study suggests restricting dropout to the neuron outputs to maintain training stability and ensure convergence. The LSTM is structured slightly differently from the normal RNNs. The RNNs with their basic structure are subject to the problem of vanishing gradient. This phenomenon arises because, during the propagation of gradients across many time steps, repeated multiplication can cause the gradients to progressively diminish. As a result, the influence of earlier inputs on later outputs reduces, leading to a decaying effect through the network. Consequently, standard RNNs struggle to capture long-term dependencies in sequential data. To address this limitation, the LSTM architecture was developed by slightly altering the internal design of the hidden layer. These adjustments allow the network to retain important temporal information over extended sequences, mitigating the vanishing gradient problem that afflicts traditional RNNs[15].

The LSTM architecture is set up differently by its inclusion of four interdependent layers, compared to the single-layer design used in traditional RNNs, as shown in Figure X. These layers consist of three sigmoid (σ) gates—namely, the forget gate, input gate, and output gate—along with one tanh layer. The σ gates regulate the modification and retrieval of memory content. An essential feature of the LSTM is the presence of a specialized pathway, which allows data to flow through the network without being processed by the neurons, known as the "cell state." This bypassing mechanism mitigates the vanishing gradient problem, facilitating the retention of long-term dependencies within the network. In the context of a False Data Injection (FDI) attack, initial measurement data is obtained using Matpower, to which random noise is added to simulate realistic operating conditions. The attack vector a is generated according to the equation $a = Hc$, where H is known as the Jacobian matrix and c is the modification on the actual state vector. This attack vector is then incorporated into the measurement data, producing the attacked measurement data z_a. Additional details on the construction of attack vectors are provided in subsequent chapters. From here, both the original and the attacked measurement data are labelled, thus forming the dataset for model training. The dataset is split into training and testing subsets, with the training data used to fine-tune the hyperparameters of the LSTM network, incorporating dropout regularization. Once the training process is complete, the model's performance is evaluated on the test set, using the ground truth labels to assess accuracy and effectiveness.

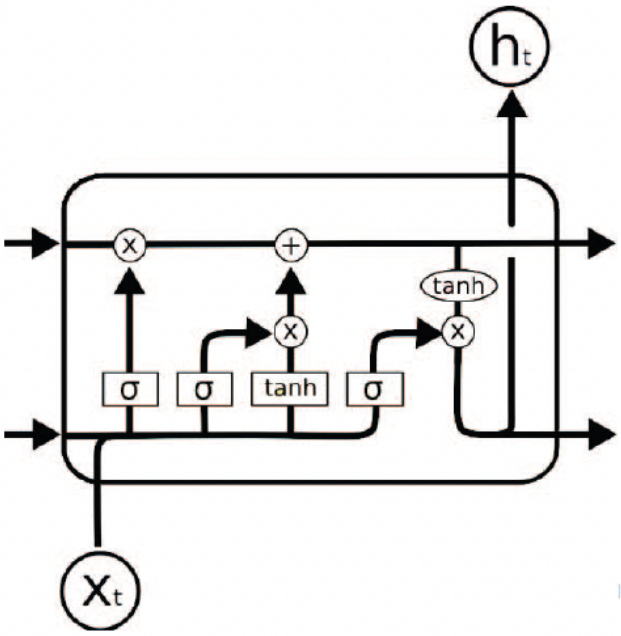


Figure 1.1: LSTM Model Architecture[32]

Another approach to FDI attacks is known as zero parameter information. A study demonstrated the possibility of an attacker to craft stealthy FDI attacks with zero information of line parameters, hence the name zero parameter information FDI attack[31]. An important concept in this model of FDI attacks is a "cut line" defined as a transmission line that can divide the power network into two disjoint islands. The adversary exploits the vulnerability exposed by these cut lines to craft stealthy attacks against the power system without actual knowledge of line parameters. In the study, the FDI attacks were implemented considering three different scenarios. One of them was a case where there was a one-degree bus, the second with a one-degree super-bus and finally a bus with multiple cut lines. One-degree refers to when a bus is connected to the external by only a single cut line. Also, a super-bus refers to a group of buses. The study claimed to prove that an adversary can generate stealthy attacks with only limited information as in if the bus or super bus is connected externally by virtue of cut lines. This attack model with incomplete information was initially designed as a flexible load redistribution but still requiring power network and line parameters info around the attack region. The zero-parameter model instead demonstrates the possibility of successfully crafting a stealthy attack with zero knowledge of line parameters. In terms of the details, we take a look at the different attack scenarios with the zero parameter model with the first being the attack on a one-degree bus. It builds on the analysis that the state variable of this type of bus can be modified by leveraging the system topology. Given that in this scenario, the bus is connected to the external by only a single branch, which here is the cut line, an adversary constructs a successful attack vector as shown below[31].

$$a_i = \begin{cases} \alpha, & \text{if i} = i_d \text{ or } d+n \\ -\alpha, & \text{if i} = j_d \text{ or } d+n+l \\ 0, & \text{otherwise} \end{cases} \tag{1.1}$$

Here α is a random value and the cut line is defined as $t_d = \{i_d, j_d\}$ representing a link from bus i to bus j and d is the index of the transmission line. The attack is crafted by the adversary injecting inverse errors in the measurements of the bus power injections at nodes i and d. This results in compromised measurements.

In a second scenario involving an attack on a one-degree super bus, we have a group of buses divided by a cut line from another part of the network. Here we define the buses belonging to this group as a set Ng (1,2,....p). The super-bus is defined as G, and the transmission lines in the cut are a set L_g.

With B defined as the symmetric admittance matrix and S the shift factor bus-branch matrix obtained from the A the branch bus incidence matrix and D the diagonal susceptance matrix susceptance matrix.

$$B = A^T DA \tag{1.2}$$

$$S = DA \tag{1.3}$$

The columns in S corresponding to N_g, when summed up, are given as

$$s_g = s_1 + s_2 + ... + s_p \tag{1.4}$$

From the equation, it is derived that the vector s_g only relates to Lg. Again, with the same procedure in B, b_g is obtained as

$$b_g = b_1 + b_2 + ... + b_p \tag{1.5}$$

From this, it is also derived that b_g only has a link with the line parameters. The study then extracts a theorem indicating that if an adversary has information about the parameters on all transmission lines connected to this super-bus, it will be possible to modify the state variables of the buses belonging to this super-bus. Building upon this theorem, along with information that this super bus is connected externally to another part of the network with only the cut line. The adversary constructs an attack vector same as indicated in equation **??**. The impact on the state variables is obtained as $c_i = \frac{\alpha}{b_{i_d j_d}}$ for all i, j in N_g. The compromised measurements are obtained by the injection of the errors into the power injection of buses i_d and j_d

The final scenario is one whereby the super bus has multiple cutlines. Again this is constructed from the notion that the super bus is connected to the external part of the network by virtue of the cut lines only. The attack vector a is constructed as $a = \sum_{k=1}^{l_c} \lambda_k a_k$ where λ_k is always equal to 1 or 0 and l_c is the size L_c. a_k represents an attack vector related to a single cut like t_k in L_c. The formulation is then given as

$$a_{k,i} = \begin{cases} \alpha_k, & \text{if i} = i_d \text{ or } k+n \\ -\alpha_k, & \text{if i} = j_k^c \text{ or } k+n+l \\ 0, & \text{otherwise} \end{cases} \tag{1.6}$$

k represents the index of the relevant cut line. α is a random value and hence $a_{k,i}$ is the ith element of the vector a_k. The modification on the state variable is then obtained as $\hat{x}^a_{i_d} = \hat{x}_{i_d} + c_{i_d}$ where c_{i_d} is obtained from the injected errors α i.e $c_{i_d} = \sum_{k=1}^{l_c} \lambda_k \frac{\alpha_k}{b_{i_d j^c_k}}$. Leveraging the information that the super bus G is connected only by a group of cut lines, an attack vector a_k is successfully constructed corresponding to the cut line t_k in L_c as

$$a_{k,i} = \begin{cases} \alpha_k, & \text{if } i = i^c_k \text{ or } k+n \\ -\alpha_k, & \text{if } i = j^c_k \text{ or } k+n+l \\ 0, & \text{otherwise} \end{cases} \tag{1.7}$$

Other studies focused on the defence mechanisms. In terms of these, a study [25] mentioned two main methods to protect against FDI attacks. One is based on hardware, which is to protect meters. It involves identifying vulnerable measurements and safeguarding them. An example was provided in that for the case of an n bus system, a minimum of 2n - 1 measurements will be necessary to perform state estimation. For this reason, the measurements needed to allow the system to be observed should be secured. The downside is, however, that if one of the meters undergoes a failure or is exploited, it leads to an unreliable system as only the safeguarded meter measurements can be trusted in the state estimation. A second method is software-based and uses robust algorithms. An example mentioned was based on the Bayesian framework on the grounds that the state of the system vector is random, having a Gaussian distribution. The algorithm uses historical information to estimate the distribution, which it then uses to cross-check the actual state. However, this method fails to detect replay attacks as the adversary can alter measurements while keeping them within the historical distribution of the data. Research by [7] has, however, been able to solve this problem by using the Kullback-Leibler Distance method. It still had some downsides, though, as it cannot detect replay attacks performed for brief time intervals.

Conversely, alternative strategies primarily relied on statistical methods, one of which is Bad Data Detection (BDD) using the Chi-Square Test. This technique, employed alongside state estimation, is used to identify anomalous data in smart grid systems and is practically used in real-world applications.

The Chi-Square Test evaluates discrepancies between actual measurements and estimated values by calculating the residuals and the differences between observed and predicted data. It sums the squared residuals and compares this statistic against a predetermined threshold. If the sum exceeds this threshold, it indicates a significant deviation from expected values, suggesting the presence of anomalous data.

While the Chi-Square Test provides a rigorous statistical approach to anomaly detection, it has limitations. Its effectiveness can be compromised in noisy environments or when the assumption of normally distributed residuals does not hold. Researchers are increasingly exploring hybrid methods integrating statistical techniques with machine learning algorithms to address these challenges. These advancements aim to enhance detection accuracy and bolster the resilience of smart grid systems against the evolving threat of FDI attacks.

In a study utilizing the IEEE 37-bus test feeder [4], the results demonstrated the high filtering capabilities of residual-based methods for detecting malicious data within

smart grid systems. These methods effectively identify discrepancies between actual and estimated measurements, enabling operators to isolate and correct anomalous data. The robust performance observed in this study underscores the potential of residual-based approaches in maintaining the integrity of smart grid operations.

However, the Chi-Square Test, a widely used statistical technique for bad data detection, has notable limitations. One significant concern is the occurrence of Type I and Type II errors, commonly referred to as false positives and false negatives. Type I errors can result in benign data being flagged as anomalous, while Type II errors may allow actual attacks to go undetected. Additionally, adversaries can intentionally inject false data in a calculated manner to evade detection. By manipulating measurements so that the resulting residuals fall within the predetermined valid thresholds, attackers can bypass the Chi-Square Test, effectively compromising system integrity without raising alarms.

Moreover, another critical limitation arises from the reliance on the measurement error covariance matrix, which may not accurately reflect the true errors in the data. If the model used to represent measurement errors is flawed or does not account for all sources of variability, it can significantly impair the performance of the Chi-Square Test. Such inaccuracies can lead to an increased likelihood of undetected anomalies, undermining the detection mechanism's overall effectiveness.

In light of these challenges, it is essential to explore complementary techniques and robust frameworks that enhance the reliability of bad data detection methods, ensuring that smart grid systems remain resilient against sophisticated FDI attacks.

Chapter 2

Theoretical Framework

2.1 Overview of Power Grid

The power grid in summary is a network system designed to effectively distribute electric power. The grid consists of various electrical components each with a specific function enabling the reliable transfer of energy from generation to consumers. The structure of the grid can be categorized into four main components known as Generation, Transmission, Distribution and finally Consumption.

The generation is where the actual production of electricity occurs by virtue of power plants. This is obtained mainly through the conversion of mechanical energy to electrical energy. The sources of mechanical energy varies from coal,thermal energy, natural gas, and oil to wind,solar, geothermal and other forms of energy. These are mainly grouped into two known as fossil fuels and non-fossil fuels respectively. With the application of synchronous generators, the mechanical energy obtained from these sources is converted to electrical energy and fed to the transmission system. The generated power is usually at low voltage ranging from about 11 to 35kV. These are generally stepped up by virtue of transformers as the power is transferred to the transmission system

The transmission system allows for the transfer of electricity to regions far away from the generation site. Considering transportation costs of raw materials needed for electricity, generation power plants are usually located close to theses sources. On the other hand, with generation through non-fossil fuels such as solar,hydroelectric and wind it is inevitable to have the generation plants where they are geographically obtained. The transmission system allows enables the transport of the electricity generated at these locations to regions further away in need of electrical power. In terms of the fundamental structure of transmission systems, it is worth knowing that power loss in a transmission line increases directly with an increase in the square of the line current ($P_{\text{loss}} \propto I^2$). For this reason, the voltage is usually stepped up in the transmission system, which reduces the current by virtue of ohm's law subsequently minimizing loss on the transmission line. These voltages could go up to about 275kV and higher. Also it is worth mentioning that transmission systems are generally designed as a mesh structure to enhance fault tolerance and allow for the transmission of energy through alternate paths if some are affected. This further enhances the reliability of the transmission network.

The distribution system is the section that handles the transfer of electrical energy finally to the consumers. These operate at a lower voltage level, usually ranging from 120V,

the lowest to about 69kV. As opposed to transmission systems, distribution systems are mainly designed in a radial structure. Also the distribution system is further subdivided into primary and secondary. Industrial customers who require a larger amount of power are fed through the primary distribution whilst residential customers are provided for by the secondary distribution[11].

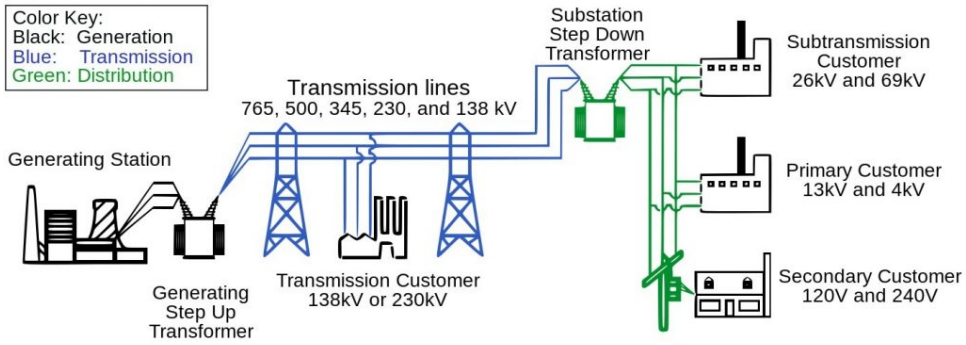


Figure 2.1: Overview of the traditional Power Grid[29]

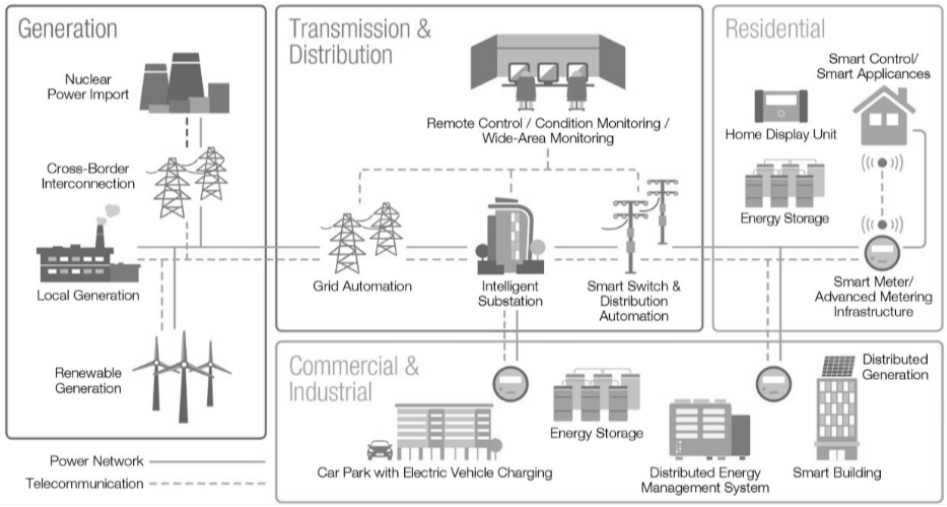


Figure 2.2: Overview of the Smart Grid[8]

2.1.1 Communication Infrastructure

A critical component of the modern power grid is its communication infrastructure, which allows for the seamless exchange of data across the network. As smart grid architectures continue to evolve—encompassing a wider range of stakeholders, devices, and services—there is an increasing demand for communication systems that are both efficient and highly reliable. The complexity and interconnectivity of these systems require robust communication channels capable of supporting real-time monitoring and control over the network at both transmission and distribution levels. Transmission networks generally require high-bandwidth, low-latency communication to ensure system-wide coordination

and reliability across vast distances. Meanwhile, distribution networks must accommodate a wider variety of devices and services at the local level which necessitates flexible and scalable communication protocols to handle data from smart meters, distributed generation units, and consumer-facing technologies.

In terms of the communication infrastructure of the power grid, it is typically segmented into two primary domains: the office network and the process control network. The office network corresponds to conventional enterprise IT infrastructure, supporting general administrative functions requiring internet connectivity. In contrast, the process control network is dedicated to the operations of the grid directly interfacing with the physical field devices of the grid. This includes the interconnection across substations and control points, enabling real-time data acquisition and control. Communication within the process control network generally adheres to standardized protocols such as those defined by the International Electrotechnical Commission (IEC), ensuring interoperability and reliable data exchange. Devices such as Phasor Measurement Units (PMUs) continuously transmit measurement and status data most often via Programmable Logic Controllers (PLCs), to centralized control centers[1]. From that point, supervisory control systems aggregate, process, and interpret the incoming data to monitor grid stability, detect anomalies, and coordinate timely operational responses.

In the context of the Smart Grid, the overall architecture is typically segmented into four distinct layers: the application layer, communication layer, power control layer, and power system layer. On the customer side, the application layer provides a range of services to improve user interaction and energy efficiency, including home automation, demand response mechanisms, and dynamic market pricing. On the utility side, the focus shifts towards grid automation and efficient real-time power distribution across the network. The primary distinguishing feature of the Smart Grid, when compared to traditional grid systems, lies in its communication layer. This layer facilitates bi-directional information exchange and enables advanced functionalities such as remote monitoring, control, and predictive maintenance. To implement this layer, a combination of wired and wireless communication technologies is leveraged. The appropriate technology is selected based on factors such as latency, bandwidth and reliability. Notably, three key network types support this communication framework: the Premise Area Network (PAN), which connects individual devices within a consumer's premises; the Neighbourhood Area Network (NAN), which aggregates data from multiple households or buildings; and the Wide Area Network (WAN), which links distributed field devices and substations to central control systems across broader geographic areas. Together, these layers and networks enable the Smart Grid to function as an intelligent, adaptive, and resilient energy infrastructure.

Wide Area Network

The wide Area Network (WAN), also referred to as the Metropolitan Area Network (MAN), makes up the core of the communication infrastructure within smart grids. It provides essential connectivity between the transmission and distribution layers, enabling centralized monitoring and control across large geographic regions. Through the WAN, critical components such as substations, control systems, and Remote Terminal Units (RTUs) are interconnected with centralized control centers operated by utility providers. Given the extensive coverage area, typically ranging from 10 to 100 kilometers and the

need to transmit large volumes of real-time operational data, WAN implementations require high-bandwidth capabilities, generally between 10 Mbps and 1 Gbps. Measurement data collected by field devices such as Phasor Measurement Units (PMUs) and Intelligent Electronic Devices (IEDs) is transmitted via the WAN to control centers, where it is analyzed to maintain real-time observability of the grid. This continuous stream of data allows decision-making processes such as load balancing, fault detection, and dispatch of corrective actions, all of which are critical for ensuring the reliability and stability of grid operations[17]. Also, various communication technologies are utilized in WAN deployments, depending on geographic and infrastructural constraints. Fiber optic networks are preferred for their high bandwidth and low latency, while cellular technologies (e.g., LTE, 5G) offer flexibility and wider coverage in urban and suburban areas. Power Line Communication (PLC) is occasionally used for shorter WAN segments, especially where existing power infrastructure can be leveraged. In remote areas, satellite communication may be used as a fallback solution, ensuring connectivity where alternative options are limited or unavailable.

Neighbourhood Area Network

The Neighbourhood Area Network (NAN) operates within the distribution layer of the smart grid. It serves as an intermediary between the Wide Area Network (WAN) and the Premise Area Network (PAN). This layer is typically established through the installation of smart meters at the consumer level, enabling bidirectional communication and supporting advanced functionalities such as demand response, real-time energy monitoring, and the provision of high-quality electric power. Furthermore, the NAN facilitates inter-device communication between distributed Intelligent Electronic Devices (IEDs), enhancing automation and operational visibility across the distribution network.

A key role of the NAN is the aggregation of data from a large number of geographically dispersed endpoints, transmitting this information to substations or centralized data centers at a lower level. Consequently, the NAN must support moderate to high bandwidth requirements to accommodate frequent and latency-sensitive data exchange over relatively extended distances. However, modifying NAN infrastructure can pose challenges, particularly in densely built or constrained environments where changes to physical assets may be costly or impractical[19].

To address various scenarios, a range of wired and wireless communication technologies are leveraged which include fiber optics, cellular networks, Wi-Fi, and Power Line Communication (PLC), often used complementarily to ensure reliable, secure, and efficient data transmission.

Premise Area Network

The Premise Area Network (PAN) represents the communication layer closest to the end-users within the smart grid infrastructure. It serves as the interface between consumers and utility systems, allowing local data exchange. PANs can be implemented using both wired and wireless technologies, with the choice often depending on infrastructure availability, cost and user-specific needs.

The PAN is typically subdivided into three main categories: the Home Area Network (HAN), Building Area Network (BAN), and Industrial Area Network (IAN). The HAN

focuses on residential environments, facilitating communication among devices such as smart meters, home automation systems, photovoltaic panels, and Home Energy Management Systems (HEMS). This allows for enhanced monitoring and control of household energy consumption. The BAN and IAN instead are used in commercial and industrial environments, respectively, and are commonly used to manage automation systems such as Heating, Ventilation, and Air Conditioning (HVAC), lighting, and security systems at the building level.

Due to the relatively short distances covered within PANs, the bandwidth requirements are modest, typically ranging from 10 to 100 kbps. As a result, low-cost communication technologies such as Power Line Communication (PLC), Wi-Fi, and Zigbee are widely adopted in PAN installations. These technologies provide sufficient performance while maintaining affordability and flexibility, allowing for easy integration of new devices[17].

A sample practical application of the PAN is the delivery of real-time pricing signals from utilities to consumers, enabling dynamic demand-side management. This gives users the capability to adjust appliance usage based on price fluctuations which subsequently optimizes energy consumption, reduces electricity bills, and helps in handling peak demand pressures on the grid[2].

2.1.2 Challenges

Given the increasing digitalization of the power grid, as described previously, several critical challenges have emerged that must be addressed to ensure the continued reliable and efficient operation of the grid. One of the most critical issues is the dynamic nature of both power demand and generation, which fluctuates due to various factors such as renewable energy integration, grid decentralization and changing consumer needs. These challenges must be met with adaptive and resilient control strategies, leveraging relevant technologies to ensure the grid remains responsive, efficient, and resilient under varying operational scenarios.

One of the core concepts to take note of is the Confidentiality, Integrity, and Availability (CIA) Triad, a fundamental in terms of general IT security. Confidentiality ensures that information is only accessible to authorized personnel. The goal here is to prevent unauthorized access to sensitive data. Integrity on the other hand ensures that information always stays accurate and persistent. The objective is to prevent data from being corrupted or tampered with. Finally, availability ensures that information must always be accessible when needed. The main principle is that only two of the three above can be guaranteed at the same time[3]. Most times in IT infrastructures some form of availability is sacrificed to ensure integrity and confidentiality. In power grids, however, there is a huge consequence for unavailability of services as the consequences of downtime is very critical. For example, in the case of a blackout, the affected areas increase for every moment that power hasn't been restored yet. Also, there is a huge cost in switching on / off relevant electrical components such as huge plants. With such criticality of availability in power grids, measures to ensure confidentiality and integrity are generally isolated from those to guarantee the availability of power.

Another critical challenge in the power grid is to balance generation and consumption. Modern grids need to operate at a designated frequency, generally 50Hz or 60 Hz

depending on the standards used by the specific operator. The frequency, however is subject to fluctuations when there is a disparity between the power generated and the power consumed. This imbalance of power generation and consumption causes a frequency deviation proportional to the imbalance. A mechanism known as operation reserve is set up to ensure equilibrium between generation and consumption. We have the primary operation reserve, secondary and minute reserve. The primary is a continuous frequency load control and provides an instant response to frequency deviations. The concept of inertia here is important, which is the tendency of the grid to resist frequency deviation and is generally dependent on the types of plants associated with the grid. Fossil fuel-based generators, for example, generally have much more inertia than renewable sources such as solar panels. The second stage is the secondary frequency control, which operates on a slightly lower time scale than the primary, usually in minutes. This is generally achieved by dispatch systems from the control center to stabilize grid frequency[16]. Finally, the minute reserve takes over in situations where the deviation persists for a more considerable amount of time, eg about 15mins[10].

A significant challenge as well in ensuring the security of modern power systems lies in safeguarding field devices, particularly those deployed in outdoor environments such as substations. These devices are typically designed for long-term operation and hence with lifespans extending over several decades. This subsequently leads to fewer replacements or updates. As a consequence, many of the communication protocols used by these devices were developed years ago and lack the security features necessary to defend against modern cyber threats. The absence of built-in security mechanisms in some of these legacy protocols makes them particularly vulnerable to exploitation. For instance, the widely adopted IEC 60870-5-104 protocol, used by transmission and distribution system operators across the globe, has been shown to be susceptible to various forms of cyber attacks, including man-in-the-middle, replay, and spoofing attacks. Similarly, the DNP3 protocol, another commonly used standard, exhibits similar vulnerabilities, exacerbating the risks to critical infrastructure[21].

2.2 Powerflow analysis

2.2.1 Overview

Powerflow analysis, also known as load flow analysis, is a fundamental concept of the power grid systems. In a power grid, there is a continuous evolution of relevant variables, and these changes have to be constantly monitored. Powerflow analysis is leveraged for the continuous management of the grid. This is achieved by working out the steady-state equations of the grid. The relevant variables include phase angles, voltage magnitude, active, and reactive flows along the lines of the power grid. This is obtained under dynamic grid conditions. The objective of the power flow analysis is to mathematically determine relevant variables, as mentioned above, for each node, also known as a bus. This analysis involves non-linear power flow equations and is computed generally with iterative methods. Amongst the techniques used to perform power flow analysis are Newton-Raphson, Gauss-Seidel and Fast Decoupled methods. In this thesis, we leveraged the Newton-Raphson method which is faster and more reliable than the Gauss-Seidel method given available resources. In this thesis, the IEEE standard test networks were

used for experiments which didn't require heavy computational resources. In terms of history, power flow analysis was initially computed with analog boards, which were not quite reliable. However, from the 1950s onwards, digital computers were leveraged to perform the computations with better accuracy. Over the years, new methods were derived enhancing the computational efficiency whilst achieving utmost accuracy in terms of results.

Given that the power grid involves a huge interconnected network, we end up with power flows across the different branches connected to various nodes. The amount of power through each line is determined based on Kirchhoff's principles. Table 2.1 gives a description of the common variables in a power flow analysis.

Variable	**Description**
S_i	Complex power at bus i, $S_i = P_i + jQ_i$
P_i	Active power injection at bus i (MW)
Q_i	Reactive power injection at bus i (MVAR)
V_i	Complex voltage at bus i, $V_i = \|V_i\|e^{j\theta_i}$
$\|V_i\|$	Voltage magnitude at bus i (pu)
θ_i	Voltage phase angle at bus i (radians)
I_i	Complex current injection at bus i
Y_{ik}	Element of the bus admittance matrix (Y-bus) between buses i and k
Y_{ii}	Self-admittance of bus i in the admittance matrix
G_{ik}	Conductance component of Y_{ik} (real part)
B_{ik}	Susceptance component of Y_{ik} (imaginary part)
θ_{ik}	Voltage angle difference $\theta_{ik} = \theta_i - \theta_k$
P_i^{calc}	Calculated active power at bus i
Q_i^{calc}	Calculated reactive power at bus i
n	Number of buses in the system

Table 2.1: Relevant Variables in Power Flow Analysis

Firstly, we define the complex power S given as

$$S = P + jQ = VI^* \tag{2.1}$$

From ohm's law, the current is obtained as

$$I_i = \sum_{k=1}^{n} Y_{ik} V_k \tag{2.2}$$

Substituting I to the power equation yields the equation

$$S_i = V_i \sum_{k=1}^{n} Y_{ik}^* V_k^* \tag{2.3}$$

which if expanded into the imaginary and real components we have

$$P_i = \sum_{k=1}^{n} |V_i||V_k|(G_{ik}\cos\theta_{ik} + B_{ik}\sin\theta_{ik}) \tag{2.4}$$

$$Q_i = \sum_{k=1}^{n} |V_i||V_k|(G_{ik}\sin\theta_{ik} - B_{ik}\cos\theta_{ik}) \tag{2.5}$$

where P_i is the active power component at bus i, whereas Q_i represents the reactive power component for the same bus.

It is important to note that the buses are classified into 3 main categories. Firstly, we have the generator bus, which is controlled, and with this bus, we have two known variables, namely the real Power P and the voltage magnitude $|V|$. The unknown variables are Q representing reactive power and phase angle θ, which are to be determined in the power flow analysis. The second type of bus is the slack bus, otherwise known as the swing or reference bus. Here, the known variables are specified as the voltage magnitude $|V|$ and voltage angle θ whilst it is necessary to determine P and Q, which are unknown variables. Finally, we have the load bus, which is a type of bus that does not have any generator attached to it. They form the bulk number of buses in a power system, and the known variables are P and Q. The voltage magnitude $|V|$ and voltage angle θ are unknowns and have to be computed at this bus. Table 2.2 summarizes the classification of buses[14].

Type of Bus	**Known Variables**	**Unknown Variables**
Slack	$\|V\|$,θ	P, Q
Generator(PV)	P, $\|V\|$	Q, θ
Load(PQ)	P, Q	$\|V\|$, θ

Table 2.2: Classification of Buses

2.2.2 Security constrained Optimal Power flow

An important concept in power flow analysis is the security-constrained optimal power flow. This is an advanced optimization technique leveraged to obtain economic efficiency whilst ensuring the reliability and security of the grid. In this method, the capabilities of the traditional optimal power flow analysis are enhanced, considering potential contingencies such as generator failures or transmission line outages. In the real world, we have unexpected events affecting the grid that must be taken into account, eg during heavy rain with lightning, there could be equipment failures or in the case of accidents whereby a transmission line is disrupted by a vehicle in the case of overhead transmissions. How does the grid react to various kinds of failures? The objective of the security-constrained OPF is to lower operational costs while guaranteeing system stability during normal and contingency scenarios. By simulating and evaluating random failures, security-constrained OPF ensures that the power flows, voltage levels, etc, are within prescribed operational limits even during system failures.

In terms of architecture, the relationship between SCOPF and OPF is as follows.

OPF	**SCOPF**
$\min f(P)$	$\min f(P)$
subject to:	subject to:
$g(P) = 0$	$g(P) = 0$
$h_{\min} \leq h(P) \leq h_{\max}$	$h_{\min} \leq h(P) \leq h_{\max}$
	$h'_{\min} \leq h'(P) \leq h'_{\max}$

$g(P)$ represents the power flow equations, $h(P)$ the constraints under normal conditions whilst $h'(P)$ represents the constraints under contingency scenarios. It should be noted that the SCOPF differs from the OPF only when a contingency is enforced. Generally, the security-constrained OPF problem is defined as follows:

$$\begin{aligned} \min \quad & f(x_0, u_0) \\ \text{s.t.} \quad & g_k(x_k, u_0) = 0, \quad k = 0, 1, 2, \ldots, c \\ & h_k(x_k, u_0) \leq h_k^{max}, \quad k = 0, 1, 2, \ldots, c, \end{aligned} \tag{2.6}$$

Here, g_k is the power flow equation, and h_k is the branch-flow constraint. x_0 refers to the state variables indicating voltage magnitudes and angles without any contingency. k represents the k-th system configuration. The base case at $k = 0$ indicates the initial pre-contingency state and goes up to c, representing the total set of contingencies[9].

In SCOPF, there are two main types of actions: a proactive "preventive" and reactive "corrective" action. The preventive action restores the normal state from an alert state. It is implemented to avoid potential disruptions caused by a contingency. The corrective, however, shifts the system from an emergency state back to a normal phase, and in this context, the duration of the action is critical.

2.3 State Estimation

In a power grid, there is the need to have knowledge of the system state at the control center level to ensure the reliability and efficiency of the grid. The system state is constantly evolving in a power grid due to various factors. These include abnormalities such as weather conditions leading to equipment failure in some parts of the grid, sudden changes in user consumption of electricity or faults of any kind such as short circuits, overvoltage, etc. These issues must be instantaneously handled to ensure a reliable power supply.

The main source of knowledge of the system state is measurement devices installed throughout the network. Measured values are sent to the control center through appropriate communication channels. Given that the power flow analysis is already described in the previous section, one might wonder what the purpose of state estimation is. The reality is that the measured values indicated above do not represent the true state of the system for a couple of reasons. The main one is that the measurement devices are imperfect; hence, small measurement errors are always expected. Also, measured values could undergo modification via the communication channels, leading to incorrect values being reported at the control center. The method of state estimation approximates unknown state variables from selected measured values in a way that minimizes the overall measurement error. Power flow analysis assumes perfect data and doesn't perform any

bad data correction; for this reason, state estimation is primarily used to obtain the system state of the grid in real-time, allowing continuous monitoring. Powerflow analysis is then run based on the output of the state estimation for the contingency analysis, which determines the right action needed to be performed to maintain the reliability of the grid at a specific moment, such as a generator dispatch to regulate the frequency.

In terms of the architecture, a mathematical model is leveraged to represent the whole system in the state estimation process. The main variables are explained below.

Variable	**Description**
z	The raw noisy measured values from devices
$z*$	Actual measurements that would have been obtained from a perfect device
$\bar{x}$	Estimated system state variables
$h(x)$	The non-linear equation that relates state variables to the measured variables
H	The matrix of the partial derivatives of the measurement function with respect to the state variables
R	The noise covariance matrix, in which the diagonal elements R_{ii} represent the mean noise power (or root mean square (rms) error) of each component w_i. This quantifies the statistical relationship between the noise from different components.
w	The additive noise from sensors generating measurements

Table 2.3: State Estimation Variables[18]

In a broader description, the state variables usually include Bus voltage magnitude values V and Bus voltage angles θ. In the DC state estimation, voltage magnitudes are assumed to be constant at 1 in contrast to the AC state estimation, which considers both variables mentioned above as state variables and is a more accurate estimation of the system state, albeit a bit more complex. Common algorithms used in the state estimation process include the Weighted Least Squares(WLS), Kalman Filter, Extended Kalman Filter and the Gauss-Newton Iterative Method. In our case, the WLS method was leveraged, the most commonly used for the AC state estimation, due to its robustness, effectiveness and computational efficiency.

Generally in the state estimation, the measured variables are typically modeled as linear functions of the underlying state variables. This relationship is mathematically expressed through the measurement equation:

$$z = Ax + w \tag{2.7}$$

Here A represents an m x n measurement matrix where m is the number of measurements and n is the number of state variables. w represents the measurement noise as described in table 2.3. To obtain the optimal state, J is defined as shown below to quantify the measurement residual

$$J = (z - Ax)' R^{-1} (z - Ax) \tag{2.8}$$

By minimizing the scalar from equation 2.8 above, the optimal state vector $\hat{x}$ is obtained.

This approach operates under the assumption of a linear relationship between the measured and state variables and forms the basis of linear estimation theory. Linear estimation is widely used due to computational efficiency, however, it may be limited in accuracy when nonlinearities inherent in the power system model are significant.

On the other hand, when the measurements are non-linear functions of the state which is the practical case in real-world scenarios, it is common to linearise the equations about a nominal value and an iterative process of relinearising about newly found estimates. Here the measurement vector is defined as

$$z = f(x) + w \tag{2.9}$$

where f(x) is defined as the nonlinear function relating z to the state variables. J is then defined as

$$J = (z - f(x))' R^{-1} (z - f(x)) \tag{2.10}$$

From here we take x^0 as a nominal value of the state variables defined as

$$z^0 = f(x^0) \tag{2.11}$$

and by expansion with the Taylor series, we have

$$z = z^0 + \Delta z \tag{2.12}$$

$$f(x) = f(x^0) + F\Delta x \tag{2.13}$$

F is defined the n x n Jacobian matrix given by:

$$F_{ij} = \frac{\partial f_i}{\partial x_j}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n \tag{2.14}$$

Δx and Δz are minute changes in x and z, and inserting them into the previously defined equation for J gives

$$J = (\Delta z - F\Delta x)' R^{-1} (\Delta z - F\Delta x) \tag{2.15}$$

which corresponds to the same quantity being minimized in the problem formulation given as

$$\Delta z = F\Delta x + w \tag{2.16}$$

Linear estimation equations are then applied to estimate state deviations Δx based on the measurement deviation Δz after which the newly obtained state deviation estimation $\Delta\hat{x}$ is summed to the nominal value x^0 to obtain a new state estimate $\hat{x}$. This iterative process continues until a predefined convergence criterion is satisfied[18].

2.3.1 Weighted Least Squares Algorithm

For the WLS algorithm, as described in relevant research[28], we have two main parts, namely the measurement function and the gain matrix G(x). Assuming a test system with N buses, there are 2N - 1 state variables, which are the voltage magnitudes and voltage angles at each bus, excluding the voltage angle at the slack bus, which has a value of 0 as it is the reference bus. The vector representing the state variable is then:

$$x^T = \left[V_1, V_2, \ldots, V_N, \theta_1, \theta_2, \ldots, \theta_N\right] \tag{2.17}$$

The measurement variables considered are nominally the active and reactive power injections at each bus along with the active and reactive power flows. These are given by the equations below:

$$P_i = V_i \sum_{j=1}^{N} V_j \left[G_{ij}\cos(\theta_i - \theta_j) + B_{ij}\sin(\theta_i - \theta_j)\right] \tag{2.18}$$

$$Q_i = V_i \sum_{j=1}^{N} V_j \left[G_{ij}\sin(\theta_i - \theta_j) - B_{ij}\cos(\theta_i - \theta_j)\right] \tag{2.19}$$

$$P_{ij} = V_i^2[g_{si} + G_{ij}] - V_i V_j \left[G_{ij}\cos(\theta_i - \theta_j) + B_{ij}\sin(\theta_i - \theta_j)\right] \tag{2.20}$$

$$Q_{ij} = -V_i^2[b_{si} + B_{ij}] - V_i V_j \left[G_{ij}\sin(\theta_i - \theta_j) - B_{ij}\cos(\theta_i - \theta_j)\right] \tag{2.21}$$

Equations 2.18 and 2.19 represent the bus power injections, whilst 2.20 and 2.21 indicate the power flows across the lines.

The second part deals with the Gain matrix G(x) obtained by the Jacobian matrix H, which represents the sensitivity of the measurement function with respect to the state variables and the error covariance matrix, R. We obtain G(x) as:

$$G(x) = H^T R^{-1} H \tag{2.22}$$

Going back to the WLS estimation, we then have an objective function, which is to minimize the loss between initial measurements and estimated measurements from the state variables.

$$\begin{aligned} J(x) &= \sum_{i=1}^{m} \left[\frac{(z_i - h_i(x))^2}{R_{ii}}\right] \\ &= [z - h(x)]^T R^{-1} [z - h(x)] \end{aligned} \tag{2.23}$$

From equation 2.23, z_i refers to the measurement variables, which includes the power injections and flows as previously described and $h_i(x)$ represents the measurement function. From these, we obtain the WLS equations for each time step k as follows:

$$G(x) = H^T(x^k) R^{-1} H(x^k) \tag{2.24}$$

$$\Delta x^{k+1} = [G(x^k)]^{-1} H^T(x^k) R^{-1} [z - h(x^k)] \tag{2.25}$$

2.4 Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a specific type of deep learning model tailored for generative processes. These include synthesizing of images, generation of human-like speech, enhancement of images and much more. The main objective of GANs is to generate fake data that is indistinguishable from real-world data. Widely used in unsupervised learning, applications for GANs include high-quality sample data generation, which helps train models for domains where real data is limited. GANs are also leveraged for data augmentation, allowing models to better generalize over training and avoid overfitting. In terms of architecture, GANs leverage two main models, one being a generative model and the other being the discriminator model. The generative model produces fake data and tries to fool the discriminator into accepting it as real data. On the other hand, the discriminator learns to distinguish fake data by predicting if the data given to it is from the generative model or a sample real data distribution. In this push and pull, both models try to obtain their objectives, leading to an equilibrium scenario where the real data is almost indistinguishable from the fake data by the generative model. In other words, we have a probability distribution of sample real data and a second distribution of fake data by a generative model. The goal of the GANs is to align the two probability distributions so that it is equally likely to obtain a data sample from either.

The framework is generally built on models based on multilayer perceptrons. Initially, we assume having sample data x, the distribution of the generated sample p_g, and an input noise variable $p_z(z)$. We then define a generator function $G(z;\theta_g)$ that maps to input noise an output sample.

A second model with the same multilayer perception structure $D(x;\theta_d)$ is defined, which acts as a binary classifier and is known as the Discriminator D, indicating the likelihood that x originated from the data instead of p_g. D is then trained to optimize the likelihood of correctly classifying both real data and generated samples from G. Concurrently, G is also trained to minimize $log(1-D(G(z)))$. $D(G(z))$ is the probability that the discriminator thinks the generated sample is real, whilst $1-D(G(z))$ is the probability that the discriminator correctly recognizes the fake sample as fake. The value $log(1-D(G(z))$ penalizes the generator when the discriminator correctly classifies the fake sample as fake. Hence, in this scenario, the discriminator D wants to maximize $log(1-D(G(z)))$, meaning it wants to confidently classify generated data as fake whilst the generator G wants to minimize $log(1-D(G(z)))$, meaning it wants $D(G(z))$ to be close to 1, so the discriminator gets fooled into thinking fake generated samples are real[12]. This can be likened to a minimax game with an overall objective function given as

$$\min_G \max_D V(D;G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{2.26}$$

Chapter 3

Jacobian Matrix based FDI Attack

3.1 Introduction

In this chapter, we delve into the design of stealthy attacks which leverage the Jacobian, otherwise known as the H Matrix, to craft stealthy attacks to bypass the state estimation.

In terms of the general workflow, the first step required the obtaining of the standard IEEE test networks. In our case, the pandapower library already consisted of most of the test bus systems in-built and in a friendly manner as dataframes. For this reason it was easy to manipulate directly without the need for complex data formatting so as to be able to efficiently run the simulations. The whole procedure was broken down into main smaller components, which were, namely, Power Network, State Estimation, FDI Attack, Evaluator and Visualizer. The procedure starts with the Power Network module, which loads the selected IEEE test case to be explored and runs an initial AC power flow analysis. The initial power flow analysis allows obtaining accurate results of the system state which are used as measurement data. Measurement data from the power flow analysis is then fed to the State Estimation component, which runs where the system state is obtained by virtue of the WLS algorithm. The bad data detection is internally incorporated in the state estimation process and automatically handles what is assumed to be erroneous measurements. Upon obtaining the initial results of the system state, which is what would be observed in a control center, the next phase now involves the actual attack, which leverages the measurement data and network parameters to construct a valid attack vector added to the measurement data and re-fed to the state estimation component. Due to the careful construction of the attack vector, the bad data detection algorithm is by-passed, which yields a slightly different state estimate, leading to an incorrect observation at the control center. The results of the attacked state estimate are then passed to the Evaluator component for deeper analysis on the effect of the attacked system state, which includes metrics such as frequency deviation and voltage magnitude/angle deviation.

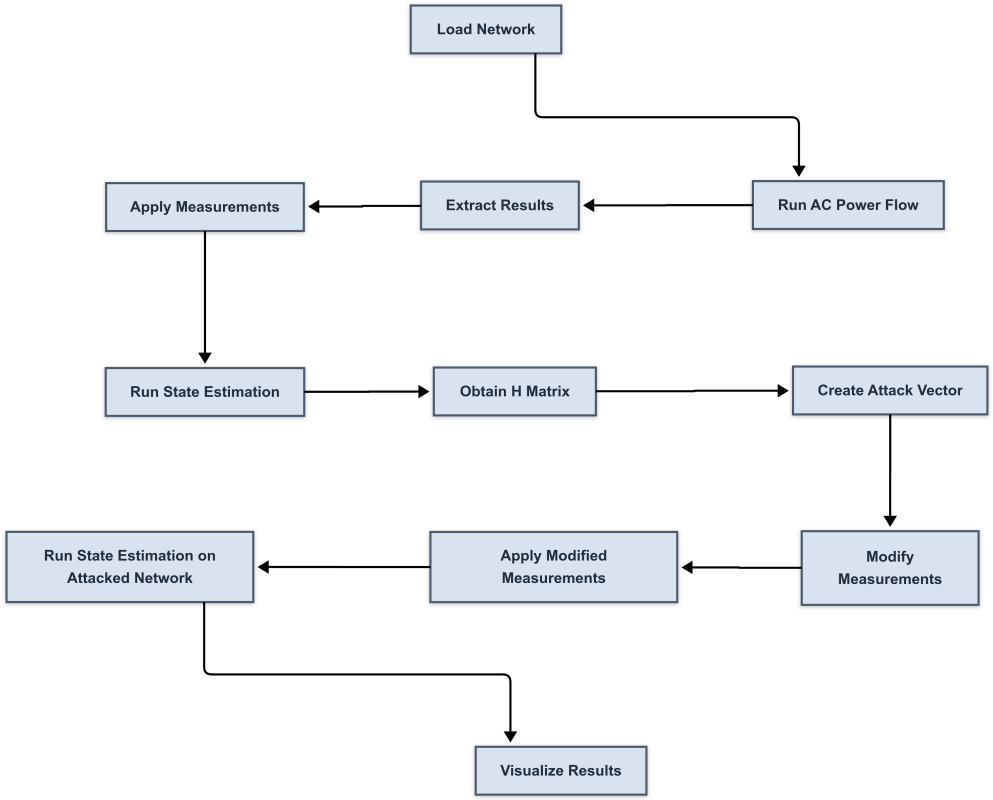


Figure 3.1: Stealthy Static Attack Flow Chart Design

3.2 Experimental Setup

In terms of tools, the main options available considering power systems analysis included MATLAB and Python. MATLAB included libraries such as matpower and pypower, whilst there are a host of python libraries in that aspect, which included PyPSA, GridCal, Matpower, pandapower, etc. Given the nature of this research and considering the main components such as the state estimation process, the Pandapower library with Python was selected as a basis to simulate the power systems. Python, being a general purpose language with a wide community support, allowed for a seamless implementation, leading to a smooth development process. The huge ecosystem of Python, which includes well-known libraries such as Pandas, NumPy, etc, allowed for a greater control of the simulation process in terms of the state estimation. Data manipulation, which is a core of the experiments, was easily applicable given the large measurement data along with output simulation data involved with the power networks needed for analysis. Another aspect is with regards to the availability of powerful visualization libraries such as Matplotlib, which is extensively used to visualize simulation results. This allowed for quick identification of key findings from the experiments which includes the FDI attacks on the state estimation accuracy along with any recognizable patterns or anomalies.

The Pandapower library, in addition to being open source, was rich in features and provided various sets of predefined functions to model the power system. On top of providing benchmark IEEE standard network test cases, it included methods to easily create or modify critical parameters such as lines, buses, transformers and generators,

which makes it easy to apply mathematical models on the network in terms of simulating the attacks on the state estimation process. Key features include the load flow analysis for both AC(Alternating Current) and DC(Direct Current) which serves as the benchmark of measurement data for the state estimation process. There is also the in-built state estimation feature, which allows for an accurate depiction of system state from measurement data along with features for bad data detection from measurement data, which have been verified, enabling rapid development and tests. These in-built features enabled a direct focus on the actual implementation and study of the effects of the FDI attack models.

3.2.1 Description of IEEE Test Cases

In power systems, IEEE test cases are generally used as a research benchmark and provide a common framework for testing and validation of tools used for power analysis, optimization, and further system analysis. These range from small networks such as the 9-bus system to huge networks with over 2000 buses and some representing portions of a real grid generally from the US and EU. Smaller networks are often used for stability studies, power flow and economic dispatch, whilst larger networks, such as with over 300 buses, are leveraged for huge-scale contingency analysis. For this reason, as with most research on power flow analysis and state estimation, only smaller networks were leveraged to better understand the effects of the proposed methodologies not going above the 300-bus system. In terms of their fundamental structure, an IEEE test case is modelled with Buses, Transmission lines, Generators, Loads and Transformers. Buses refer to nodes in the network where the load and generation are defined. At the bus, voltage levels are defined with nominal voltage. The transmission lines serve as links between buses. These are characterized by impedance Z formed as a complex number $R + jX$. R represents the resistance whilst X constitutes the reactance. The admittance Y is defined as the inverse of the impedance, hence $Y = 1/Z$. Generators provide power to the system, which includes active power P, typically in MW and reactive power Q in MVAR. The load corresponds to the demand or consumption for active and reactive power. Finally, there are transformers, which are used to step up or step down the voltage levels across buses.

Test Case	**Buses**	**Transmission Lines**	**Generators**	**Loads**
IEEE 9-Bus	9	9	3	3
IEEE 14-Bus	14	15	4	11
IEEE 24-Bus	24	33	10	17
IEEE 30-Bus	30	34	5	21
IEEE 57-Bus	57	63	6	42
IEEE 118-Bus	118	173	53	99
IEEE 300-Bus	300	304	69	193

Table 3.1: Comparison of IEEE Test Cases

Table 3.1 summarizes the characteristics of the main smaller bus systems considered in this thesis. In the smaller systems, very few generators are present, while larger systems model more distributed generation. Also, the branch density kind of increases

with respect to the system size. This usually indicates higher connectivity as transmission networks are generally designed as mesh networks for improved reliability and high quality of services.

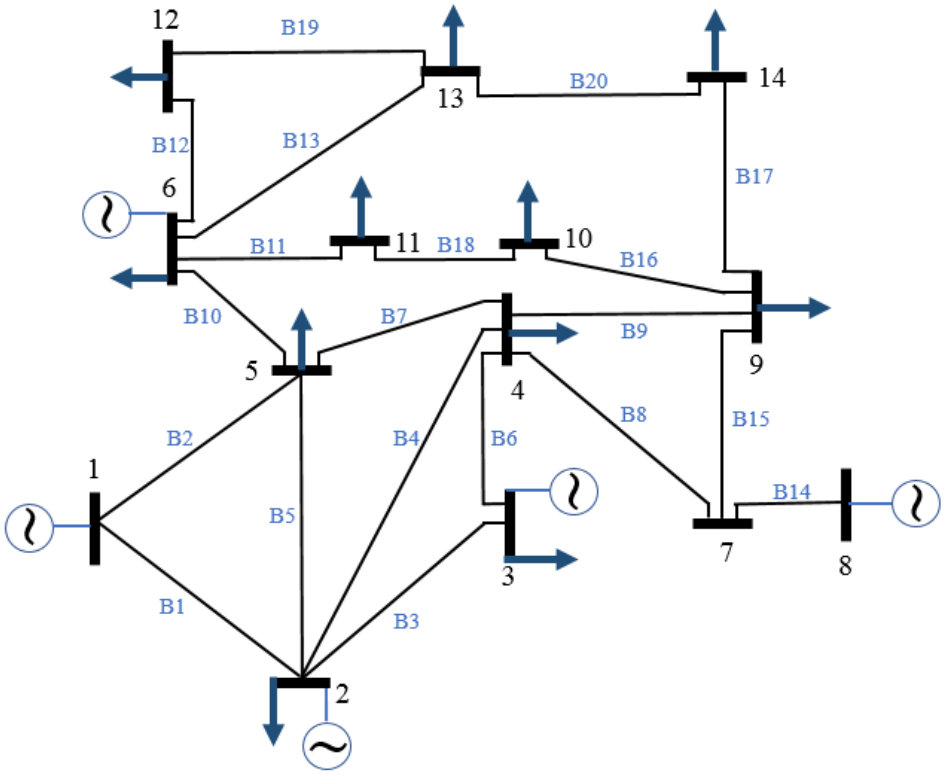


Figure 3.2: Diagram of the IEEE 14 bus system

3.2.2 Measurements

For the implementation of the state estimation, measurements are critical in obtaining the operational state of the system, which enables the continued real-time observation of the network's operating condition. The key state variables as previously defined are bus voltage angles and magnitudes, which are estimated leveraging the sensor data representing the measurements. These measurements include mainly the power flows, bus power injections and the voltage magnitude itself. The voltage magnitude measurements V_i refer to the voltage levels at various buses across the network. In terms of power flows, we have the active power flow P_{ij} representing the measured active from a node bus_i to bus_j whilst Q_{ij} is the reactive power flow along the same transmission line. In addition to these, there are bus power injections P_i and Q_i, active and reactive power injections, respectively representing the total active and reactive power produced or consumed at each bus in the network. In real-world SCADA systems, measurement values typically include the voltage magnitude and power flows with a data update of 2 to 10 seconds per update. In Wide Area Monitoring(WAM) monitoring systems, there is a faster update of data with an interval of 20ms per update.

Sources of these measurements in real-world scenarios typically include devices such as Phasor Measurement Units (PMUs), Remote Terminal Units(RTUs) and SCADA-based

sensors distributed across the network. [6]

Measurement Type	**Units**	**Accuracy (σ)**
Voltage Magnitude	pu	0.01 pu
Active Power Injection	MW	1–2%
Reactive Power Injection	Mvar	1–2%
Active Power Flow	MW	1–2 MW
Reactive Power Flow	Mvar	1–2 Mvar

Table 3.2: Measurements Used in State Estimation

An important concept to ensure a reliable and accurate state estimation process is the requirement of system operators to have sufficient measurement redundancy and ensure the observability of the network. Measurement redundancy refers to having multiple independent measurements of the same system parameters, which helps in error detection, bad data filtering, and enhancing the robustness of state estimation algorithms. Redundancy is achieved through the installation of multiple sensors across the network, enabling state estimators to cross-verify data and reduce the impact of measurement errors.

Another important concept in terms of measurements is known as observability analysis, which is a critical aspect of the state estimation process. This determines whether the available measurements provide enough information to estimate all system state variables. A power system is considered observable if the set of available measurements is sufficient to solve for all bus voltages and phase angles. If a system is partially observable, it means that some parts of the grid cannot be fully estimated, which can lead to operational inefficiencies and inaccurate decision-making.

Measurement redundancy significantly improves bad data detection and correction. It allows the Weighted Least Squares (WLS) estimation which is used in this thesis to process redundant measurements and minimize the impact of noisy or erroneous data. If an individual measurement deviates significantly from the expected value, the estimator assigns a lower weight to that measurement reducing its influence on the final state estimation results.

3.3 Baseline State Estimation

In this section, we perform a baseline state estimation without any attack to form an initial base with which to compare with attacked measurements as will be described ahead.

This baseline state estimation, performed under normal operating conditions (i.e., without cyber or physical attacks), serves as the benchmark to analyze the impact of modified state variables under attack. The estimated values are compared with ground truth data to evaluate the accuracy of the estimation process. Ground truth data is obtained by running an initial power flow analysis on the network. The results of this power flow analysis are summed up with a little random noise and leveraged as input to the state estimation process with the WLS algorithm. To quantify estimation performance, appropriate error metrics such as Root Mean Squared Error (RMSE) and Mean Absolute

Error(MAE) were leveraged to gain insight into the overall deviation from the ground truth.

Additionally, the formulation and derivation of the measurement matrix (H-matrix) used in the Weighted Least Squares (WLS) estimation approach is clearly described.

In terms of the methodology, IEEE test cases 14, 30, 39, 57 and 118 were used to analyze the baseline state estimations. For each of these test cases, the AC power flow analysis was performed with Pandapower. From the results of these, the voltage magnitude V, active power injection P, reactive power Q, and line power flows P_{ij} and Q_{ij} were extracted and applied as measurements to the state estimation with a little Gaussian noise. Note that the voltage magnitude was used as a measurement variable but also served as a state variable. The voltage angle state variable was not included in the measurements simulating the SCADA setup, which does not include it[6]. Upon convergence, the estimated voltage magnitudes V_{est} and phase angles θ_{est} were extracted and compared against the ground truth from the power flow analysis results previously obtained.

For the five IEEE test cases on which the baseline estimation, relevant error metrics were used to analyze the performance of the state estimation. The two main metrics used were the Mean Absolute Error(MAE) and the Root Mean Square Error(RMSE). The Mean Absolute Error computes the absolute difference between the ground truth and estimated values for both voltage magnitudes and voltage angles.

The MAE for voltage magnitudes, representing the average absolute error, is given by:

$$MAE_V = \frac{1}{N} \sum_{i=1}^{N} |\hat{V}_i - V_{true,i}| \tag{3.1}$$

MAE for voltage magnitudes in per-unit (pu), where N is the number of buses, $\hat{V}_i$ is the estimated voltage magnitude, and $V_{true,i}$ is the true value.

For phase angles, excluding the slack bus ($\theta_1 = 0°$):

$$MAE_\theta = \frac{1}{N-1} \sum_{i=2}^{N} |\hat{\theta}_i - \theta_{true,i}| \tag{3.2}$$

MAE for phase angles in degrees, where $N-1$ accounts for the slack bus exclusion, $\hat{\theta}_i$ is the estimated angle, and $\theta_{true,i}$ is the true value.

The RMSE, which emphasizes larger errors, is defined for voltage magnitudes as:

$$RMSE_V = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{V}_i - V_{true,i})^2} \tag{3.3}$$

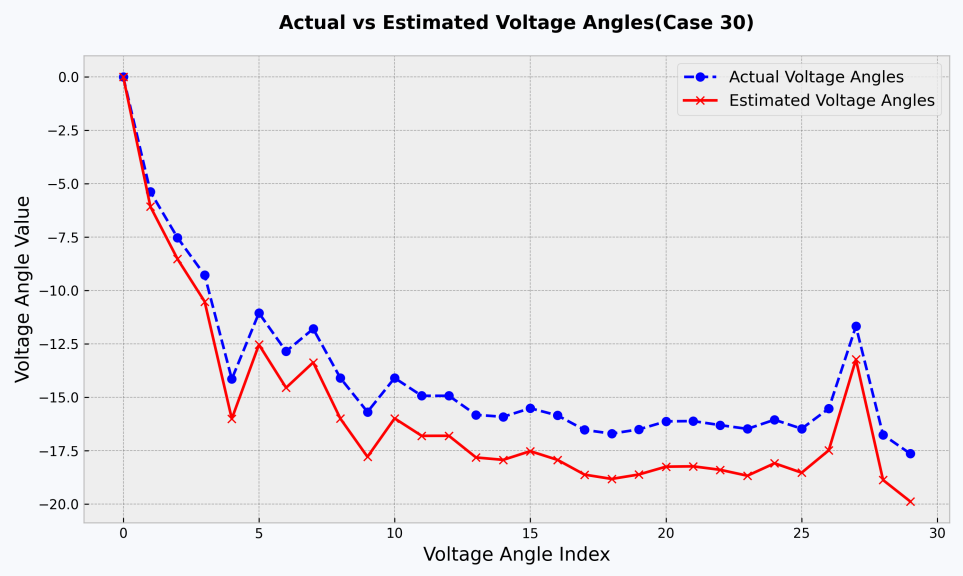
RMSE for voltage magnitudes in per-unit (pu), highlighting the root of the average squared differences.

and for phase angles as:

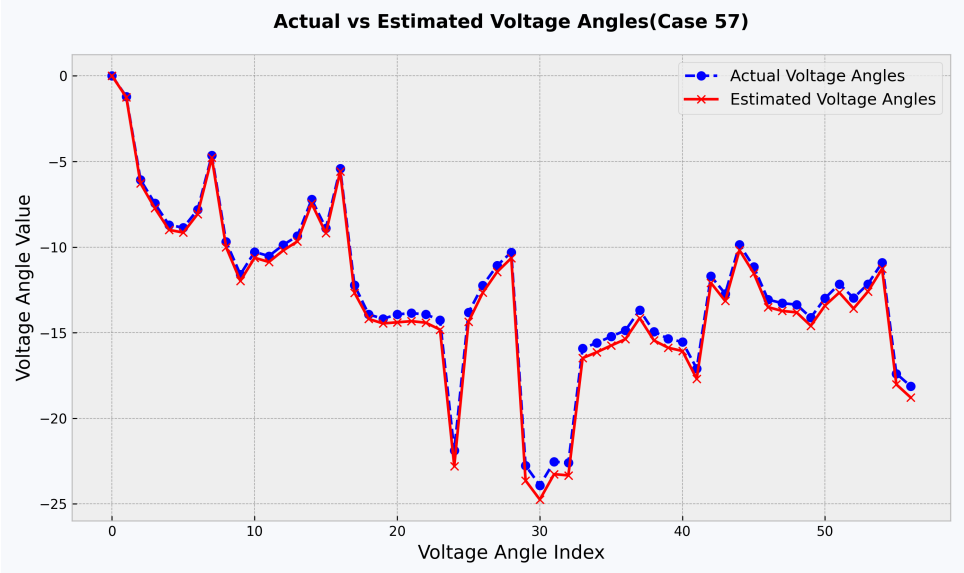
$$RMSE_\theta = \sqrt{\frac{1}{N-1} \sum_{i=2}^{N} (\hat{\theta}_i - \theta_{true,i})^2} \tag{3.4}$$

RMSE for phase angles in degrees, reflecting the root of the average squared differences excluding the slack bus.

3.3.1 Results



(a) IEEE 30-bus system



(b) IEEE 57-bus system

Figure 3.3: Comparison of Actual and Estimated Voltage Angles

The images shown in Figure 3.3 illustrate the different comparisons of voltage angles across selected test bus systems. The first is the IEEE 30 bus system as in figure 5.1a. The state estimation was quite close to the actual values of the first five buses and then deviated slightly but was overall close to the actual values. From the values obtained during the experiments, a percentage deviation of about 12% on average was obtained based on the Mean Absolute Errors. In figure 3.3b for the IEEE 57 bus system, the

estimated voltage angles were quite close to the actual values with a percentage deviation of about 3.5% on average obtained. This significantly dropped to about 0.6% for the IEEE 118 bus system as indicated in Table 3.3. It is quite easy to observe that as the number of buses increased, the estimated state variables got closer to the actual values.

IEEE Bus System	**Mean Absolute Error**	**Root Mean Square Error**	**Percentage Deviation by MAE**	**Percentage Deviation by RMSE**
Case 14	2.3244	2.4705	19.27%	20.49%
Case 30	1.8041	1.8690	12.95%	13.42%
Case 39	0.0000	0.0000	0.00%	0.00%
Case 57	0.4330	0.4696	3.43%	3.72%
Case 118	0.1076	0.1234	0.53%	0.61%

Table 3.3: Voltage Angle Error Metrics for IEEE Bus Systems (Baseline)

IEEE Bus System	**Mean Absolute Error**	**Root Mean Square Error**	**Percentage Deviation by MAE**	**Percentage Deviation by RMSE**
Case 14	0.0824	0.0827	7.86%	7.88%
Case 30	0.0572	0.0573	5.55%	5.57%
Case 39	0.0000	0.0000	0.00%	0.00%
Case 57	0.0151	0.0155	1.70%	1.74%
Case 118	0.0054	0.0057	0.55%	0.58%

Table 3.4: Voltage Magnitude Error Metrics for IEEE Bus Systems (Baseline)

Instead, for the voltage magnitudes as detailed in table 3.4, it is worth noting that the variance of voltage magnitudes is smaller relative to the phase angles. Subsequently, the percentage deviation was relatively smaller compared to the voltage angles, as previously mentioned. For the IEEE 30 bus system, the estimated values closely follow the actual values with a deviation of about 5.5% calculated from the mean absolute error. This deviation drops to about 1.7% in the 57 bus system and about 0.6% in the 118 bus system.

3.4 Attack Model

This section discusses the model used in designing an attack against state estimation. Two main models are generally considered: attacks against DC(linear state estimation) and AC(nonlinear state estimation). As discussed previously, the DC state estimation is a simpler linear estimate that approximates the state of the system. The nonlinear AC state estimation, though more complex, provides a more accurate estimation of the system state. As much work has already been done regarding attacks on DC state estimation in

academia, I focused instead more on the attack models aimed at nonlinear state estimation in my experiments

Regarding the DC attack model, the base point assumes constant voltage magnitude. This meant that Vk is set to 1 for all k in the set S of buses. There is also an assumption of negligible resistance on the branches, which results in having the reactive power on each line being set to zero. Hence, only active power injection is considered for each bus. Now, given that the power flow equations are linear, we derive the equation that relates the measurements z to the state estimate x with the equation:

$$z = Hx \tag{3.5}$$

An attacker tries to attack the system by substituting z with z_a. This vector is obtained as $z_a = z + a$, where a is the attack vector added to real measurements z. For every non-zero element in a, the corresponding sensor reading in z has been modified based on the value in a. In some scenarios, it is assumed that some sensor readings are protected and can't be compromised; in that case, the indices corresponding to such measurements are fixed at zero in the attack vector. This allows for a more practical implementation as usually in the real world, the adversary only normally has access if at all to a subset of measurement data as opposed to the whole data which is usually at the control center. Given z_a, we end up with a state estimate $\hat{x}\_a$ being the attacked system state. What is to be noted here is the residue obtained when the measurements are re-estimated from the attacked state. This is given as $z_a - H\hat{x}_a$. The main bad data detection algorithms work by ensuring the residue never passes a certain threshold; if it does, the measurement data has been compromised. What is interesting to note is that if the attack vector a is carefully selected as a linear combination of the H matrix rows, the bad data detection algorithm fails to spot the attack as the residue is no longer affected. The equation below holds:

$$r_a = z_a - H\hat{x}_a = z + a - H(\hat{x} \ + c) = z - H\hat{x} \ = r \tag{3.6}$$

r_a represents the residue upon an FDI attack, while r indicates the base residue when the real measurements are used.

For the attacks against AC state estimation, the non-linear relationship between the state and the measurement values is given by $z = h(x)$ instead, where $h(x)$ is the non-linear function that relates the state to the measurement data. Hence, the residue r_a from an attack is given by $r_a = z_a - h(x_a)$. By substituting into the equations, as seen below, we obtain the following.

$$\begin{aligned} r_a &= z_a - h(x_a) + h(x) - h(x) \\ r_a &= z + a - h(x_a) + h(x) - h(x) \\ r_a &= r + a - h(x_a) + h(x) \end{aligned} \tag{3.7}$$

For r_a to be equal to r, it is clear that we would have to equate $a - h(x_a) + h(x)$ to zero, which forms the basis of an attack against AC state estimation. Rewriting it means always selecting an attack vector a that satisfies the equation $a = h(x_a) - h(x)$ and yields an attack that would bypass the bad data detection algorithm as the residue from a true state estimate wouldn't be affected[22].

3.4.1 Results

This section presents an analysis of the impact of falsified measurement data on the state estimation process, as defined in the attack model that exploits the structure of the network through its parameters. Building upon the methodology employed in the baseline estimation, simulations were conducted for both voltage angles and voltage magnitudes in order to assess and visualize the modifications on the state variables introduced by the attack vector. This comparative analysis allows for a clearer understanding of how false data injection (FDI) affects the accuracy and reliability of the state estimation process.

Figure 3.4 illustrates the effect of the attack on the voltage angle estimates for the 30-bus system. Notably, significant deviations were observed for buses indexed between 10 and 25. This region also exhibited increased estimation errors during the baseline analysis, suggesting a naturally higher level of uncertainty or reduced observability. The observed increase in estimation errors highlights the potential for adversaries to target structurally weak areas of the network to enhance the effectiveness of their attacks. The percentage deviation in this case reached approximately 16%, signifying a considerable divergence from the true system state.

In comparison, the 57-bus and 118-bus systems demonstrated relatively lower deviations under the same attack conditions. The average percentage deviations were recorded at around 4.1% and 0.65%, respectively. While these figures suggest improved robustness in larger systems due to better redundancy and observability, the results also confirm that targeted FDI attacks can still yield noticeable estimation errors, especially in strategically vulnerable regions of the network. These outcomes underline the importance of identifying sensitive areas within the network topology that could be exploited for stealthy attack strategies.

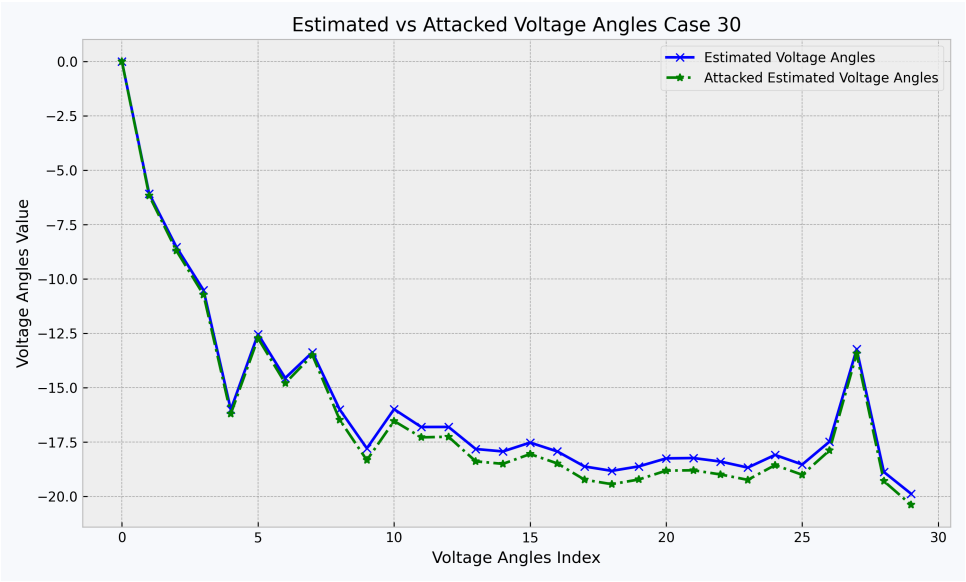
Overall, the experimental results emphasize the significance of system size, measurement placement, and network observability in determining the vulnerability of state estimation to FDI attacks. The insights gained from this analysis serve as a foundation for developing more complex attack models and subsequently robust detection and mitigation mechanisms in subsequent research.

3.5 Conclusion

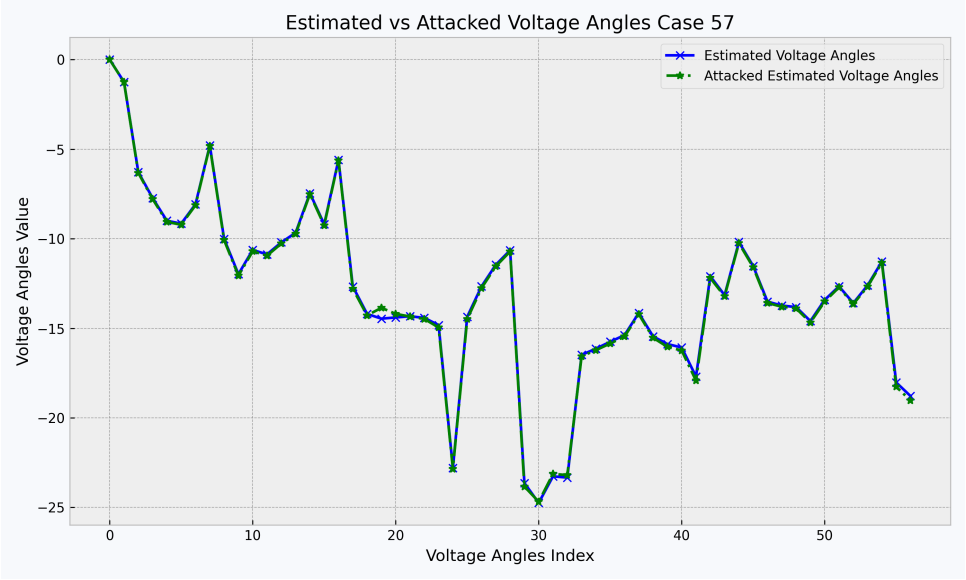
In this chapter, we designed the model for the attacks based on the Jacobian matrix which leverages full knowledge of a network. This was implemented as a starting block for the more dynamic experiment in the next leveraging the methods clearly tested in existing literature.

The experimental setup detailed the implementation of the model leveraging the Panda Power library. The simulations were then run evaluating the selected attack model and its effect on various selected networks. This gave us an insight into the effects of stealthy attacks on state variables of the power system.

Firstly an analysis of baseline estimation was done to measure the deviation of estimated state variables from a ground truth defined with some little added Gaussian noise. The impact of False Data Injection attacks on the voltage angle estimates of power systems, specifically focusing on the 30-bus, 57-bus, and 118-bus systems was subsequently analyzed. The results as described in the previous section revealed significant insights into



(a) IEEE 30-bus system



(b) IEEE 57-bus system

Figure 3.4: Comparison of Estimated and Attacked Voltage Angles

how adversaries can manipulate voltage angle estimates and the vulnerabilities inherent in different system configurations.

For the 30-bus system for example, the effect of the FDI attack was particularly pronounced, with substantial deviations in voltage angle estimates observed for buses indexed from 10 and 25 exhibiting estimation errors of up to a deviation 16%. These findings suggest that certain regions within the system, likely due to lower observability or higher inherent uncertainty, are more susceptible to attack. The relatively higher baseline errors in this part of the network further indicate its structural vulnerability, highlighting areas where attackers could strategically target for maximal impact.

In contrast, the 57-bus and 118-bus systems demonstrated better resilience to FDI attacks, with average lower percentage deviations. The lower deviation in these larger systems can be attributed to their higher levels of redundancy and observability, which enhance the system's robustness against such adversarial interventions. These results indicate that while larger systems may show an overall stronger defence against attacks, targeted FDI attacks can still lead to noticeable estimation errors, particularly in areas of the network that remain vulnerable despite the additional system size.

These outcomes are critical in understanding the broader implications of FDI attacks on power system security. The substantial impact of the 30-bus system illustrates how even small and less complex networks can be highly susceptible to attack, especially if key areas are left less observed or under-protected. On the other hand, the larger systems, though more robust, still present attack surfaces that adversaries can exploit, especially in strategically weak regions.

To conclude, we can observe that network vulnerability varies significantly across different regions within the system, with specific buses or areas being more prone to an FDI attack. Also, larger systems naturally offer better protection due to redundancy and improved observability, but they are not impervious to targeted attacks, especially in structurally weak regions.

The findings indicate the importance of not only enhancing overall system security but also identifying and reinforcing vulnerable areas within a network. Future research can be done to design advanced techniques to identify vulnerable regions of a network which can be exploited by attackers. By having specialized detection strategies for critical areas, more resilient and reliable power grid infrastructures could be implemented.

Chapter 4

Attack based on Generative Adversarial Network

4.1 Introduction

One important thing to note in stealthy attacks is the importance of the Jacobian matrix in crafting an attack model. This means, however, an assumption of an adversary having full knowledge of the network parameters, which is quite impractical in a real-world scenario. In this chapter, we advance our previous investigation by proposing a methodology to leverage GANs and the state estimation process to craft valid FDI attacks without requiring knowledge of the Jacobian matrix. Also, one thing to note is that as the static attack is for a single time step and given there is a threshold of maximum residual that can be tolerated, the state variables can only be modified to a limited extent. This makes it virtually impossible to force the control center into action, making it difficult for an adversary to achieve a meaningful objective such as influencing electricity market prices in terms of economic impacts or system disruption in the case of an attack on security. The more practical form of attack is a continuous dynamic attack on the state estimation process, which is built on top of static attacks. Since measurements are sent at very short intervals, calculating the Jacobian matrix, even with full network knowledge, would also be computationally expensive. In terms of the proposed GAN, the initial training phase can be a bit resource-intensive, but that would no longer be required at the time of an actual attack. The GAN was trained in a reinforcement learning architecture with the main objective of maximizing frequency deviation whilst minimizing detection by the bad data detection algorithm of the state estimation process. Frequency deviation is a critical metric for power system stability, and a shift beyond the threshold indicates an action on the control center, which includes generator dispatch, load shedding, etc, to restore the system to a reliable frequency. This enhancement, built on the foundational findings of the static attack, reflects a critical necessity to address advanced cyber threats that could manipulate the real-time behaviour of the system.

The next sections describe the architectural setup of the GAN attack model and the results obtained on selected IEEE test cases compared to those of the stealthy attacks in the previous chapter. The test cases were selected considering their moderate size and frequent use in stability studies. Consequently, they are a practical basis for assessing the attack's influence.

4.2 Attack Model

The GAN architecture was designed with the generator and discriminator constructed as deep neural networks. For the attack model with the GAN architecture, the IEEE 30, 14 and 57 bus systems were leveraged in the training phase. From the setup as described in the previous section, the initial step included preprocessing of the measurement vector. This was obtained from simulations with Pandapower whilst leveraging the stealthy attack model.

Exploring the preprocessing phase further, at each epoch, one of the networks is randomly selected. This followed an augmentation data by randomly perturbing the loads and generators. With this application, for every epoch, even with the same network selected, we have a unique set of network data which better allows the generalization of the model and avoids overfitting. The power flow analysis was run from this stage, and the measurements were extracted, which included the voltage magnitudes and active and reactive power injections. These were then fed to the state estimation process, and the output was leveraged in the training process.

The training process involved leveraging the state estimation results to produce realistic attack vectors. This was obtained by customizing the optimization function with a specialized reward function that penalized bad measurements whilst maximizing the frequency deviation. In order to feasibly train the models on a local PC with minimal resources, the number of epochs was set at 500, along with a batch size of 5.

After the training process, the GAN is deployed as a model to inject continuous attack vectors upon every steady-state time instant measurement. This is given as

$$z_t' = z_t + G(y, \theta) \tag{4.1}$$

z_t' represents a fake measurement from the generator for a single time instant. $G(y, \theta)$ is the generator function which outputs the attack vector, which, if added to real measurements at a certain time instant, leads to obtaining the attacked measurements. y represents network parameters embedded with measurement vector z. Note that θ represents the model parameters, i.e the weights and biases learned from the training phase. The testing was done on the IEEE 24 and IEEE 39 cases.

In terms of applying it to our specific setup, we recall the power system state estimation as described in chapter 2 and consider a power system with n state variables and m measurements. The standard nonlinear measurement model is given by:

$$z = h(x) + w, \tag{4.2}$$

where $z \in \mathbb{R}^m$ is the measurement vector, $x \in \mathbb{R}^n$ is the state vector, $h(x)$ is a nonlinear measurement function (e.g., power flow equations), and w is the measurement error, $w \sim \mathcal{N}(0, R)$.

An attacker introduces a perturbation vector $a \in \mathbb{R}^m$ to the original measurements:

$$z_a = z + a. \tag{4.3}$$

If a is constructed as $a = Hc$ for some $c \in \mathbb{R}^n$, then the modified measurements satisfy:

$$z_a = H(x + c) + w. \tag{4.4}$$

Such an attack is termed *stealthy* because it aligns with the column space of H, making it difficult to detect using residual-based bad data detection (BDD) schemes.

Delving deeper into the formulation to remove the dependency on the Jacobian matrix, we reintroduce the GAN framework that learns to generate stealthy measurement vectors directly from network features as follows.

It consists of two neural networks:

- **Generator** $G_\theta(y)$: Takes a feature vector $y \in \mathbb{R}^d$ representing the network configuration which is the bus parameters embedded with a valid measurement z and outputs an attack measurement vector $a \in \mathbb{R}^m$.
- **Discriminator** $D_\phi(z')$: Outputs a probability that a given attacked measurement vector obtained as a summation of the original measurements and the attack vector generated a is real (i.e., not generated) or from the GAN.

Delving into the specialized custom objective function for the GAN, recall that the standard GAN objective as explained in chapter 2 is defined as:

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}(x)} \left[\log D(x)\right] + \mathbb{E}_{z\sim p_z(z)} \left[\log\left(1 - D(G(z))\right)\right] \tag{4.5}$$

From this, we define a new loss function for the Generator as

$$L_G = \mathbb{E}_{z\sim p_z(z)} \left[\log\left(1 - D(G(z))\right)\right] + L_{custom} \tag{4.6}$$

where L_{custom} is defined as

$$L_{custom} = \mathbb{E}_{z\sim p_z(z)} \left[\alpha\left(\Delta f(G(z))\right)\right] \tag{4.7}$$

α here is leveraged as a control variable which allowed us to train separate models for different objectives i.e. a model that seeks to increase frequency deviation with a latter to decrease it. This is summarized as

$$\alpha = \begin{cases} 1, & \text{if maximizing } \Delta f \\ -1, & \text{if minimizing } \Delta f \end{cases} \tag{4.8}$$

The purpose was to provide a mechanism to train multiple different models with varying objectives. These could be maximizing voltage deviation or targeting specific vulnerable buses. With multiple models, a more adaptive attack mechanism could be crafted further enhancing the capabilities of an adversary.

4.2.1 Transfer Learning and Input Design

To promote generalization across networks, the generator is trained on multiple IEEE test cases with a unified input dimension, set at 57 in our case study. Therefore, for a network with $n < 57$ buses, the input vector is zero-padded:

$$y = [y_1, y_2, \ldots, y_n, 0, \ldots, 0] \in \mathbb{R}^{57}. \tag{4.9}$$

The generator outputs a measurement vector for each bus, consisting of:

- Voltage magnitude V_i
- Active power injection P_i
- Reactive power injection Q_i

Thus, for a system with n buses, the output dimension is:

$$\hat{z}_a \in \mathbb{R}^{3n}. \tag{4.10}$$

The discriminator is trained on these generated measurements to detect fakes and thereby guide the generator to improve its stealthiness.

The proposed GAN-based approach enables the construction of stealthy, transferable FDI attacks without requiring access to the system Jacobian matrix. By learning from data across multiple networks, the generator captures generalized attack strategies applicable even to unseen systems.

4.3 Experimental Setup

In this section, we describe the experimental setup of the GAN-based FDI attacks. Building upon the stealthy attack design, we leverage Pandapower with the standard IEEE test cases. The state estimation and bad data detection blocks are the same and were used exactly in this setup. The GAN implemented with the Pytorch library was then introduced to the setup, extending the stealthy attack model design and leveraging the measurements from the state estimation and power flow analysis. A unique approach was intended with the GAN in that the training was done on three IEEE test cases with data augmentation through modification of the network load and generation to produce more datasets. The objective was to obtain some kind of transfer learning in which case the GAN model learns a general attack strategy from known networks but is able to apply it to unknown networks and hence not require the H matrix to successfully craft an attack as in the stealthy attack model.

As for the GAN architecture, the Generator and Discriminator were designed as two neural networks, each with five layers. The initial input layer of the generator with Dimension D was set at 57 as previously mentioned. This number represented the maximum number of buses we would consider in this experiment. For each network, the number of buses is fed to the input layer, and the remaining values are padded with zeros. So in the case of training with the IEEE 30 bus system, the first 30 values are set whilst the remaining are padded with zeros. The hidden layers refined learnt representations, each with a non-linear activation function, and finally, the output layer was 3 times the number of buses. The reason for this was that for each bus in a network, three measurements were considered, which were voltage magnitudes, active power injections, and reactive power injections. The discriminator, on the other hand, had an input layer with the same dimension as the output of the generator, which generated fake measurements. The hidden layers learned the representations, whilst the final output layer was a binary 0 or 1 indicating if the passed input was a valid measurement. In each of the layers, Rectified Linear Unit (ReLU) activation was implemented, enhancing convergence.

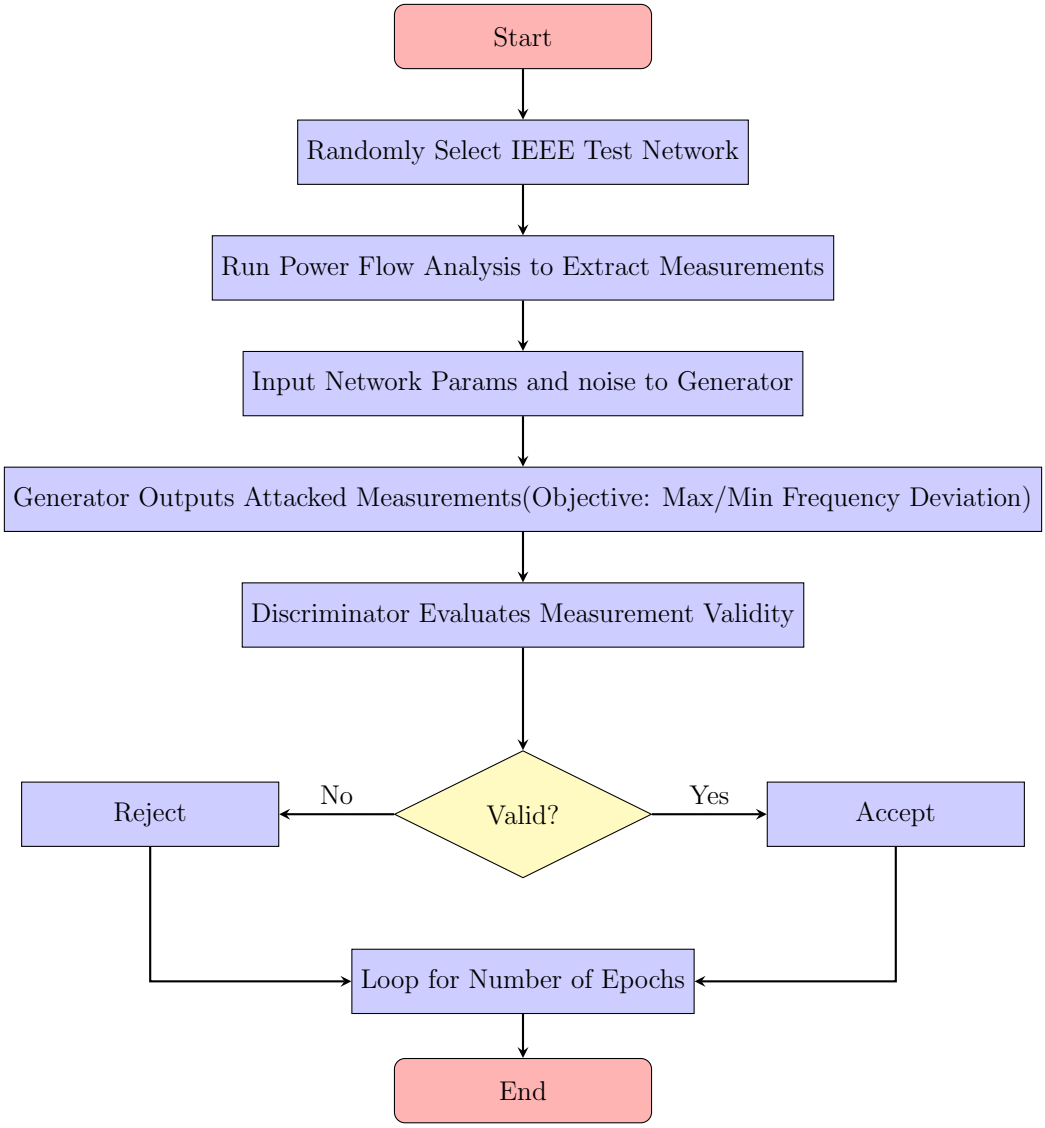


Figure 4.1: Flowchart of the GAN Attack Training Process

Chapter 5

Discussion

5.1 Introduction

In this chapter, we delve deep into the two main approaches of the study and analyze the findings. As discussed in the previous chapters, the first method involved the static attack leveraging the Jacobian matrix obtained from the state estimation process. This is then followed by the GAN-based dynamic attacks.

5.2 Static H matrix Attack

The static attack with the Jacobian matrix served as the baseline experiment. As discussed in chapter 3, the simulations were run on selected five standard IEEE test cases and the results of these attacks were analyzed. The ability to bypass the traditional detection methods employed by the WLS state estimation process was examined and the effects of the distortion of measurements on the state variables.

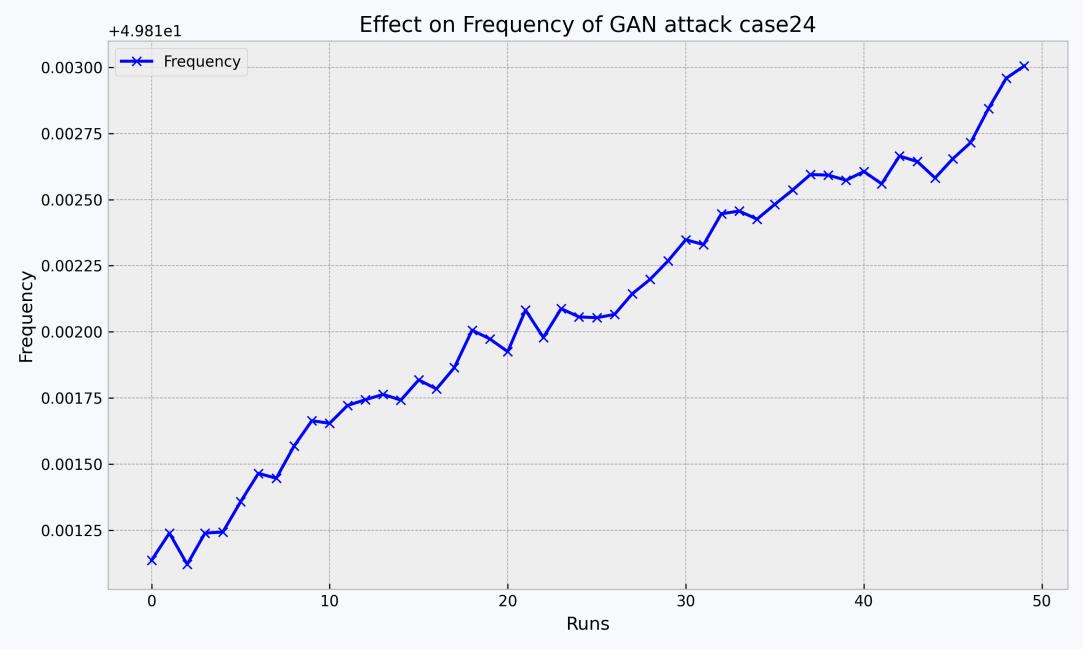
The effect on the attacks was observed with minor changes in the state variables.

The major limitation of the static attack is its impact on the overall power system. Given the small residual modifications that could be allowed, a significant modification on the output of the state variables whilst evading the bad data detection process was almost impossible to achieve. Limitations were also due to the Pandapower library in terms of determining which of the actual measurements were flagged as malicious data. The output of the detection algorithm only mentioned if there were any malicious measurements and the number of malicious measurements. Being able to have had access to the exact measurement could have helped in better analyzing and understanding the properties of the rejected which can be leveraged to optimize the attack strategy.

5.3 Dynamic Attack

Given the limitations of the static attack, the more extensive dynamic attack was studied as discussed in chapter 4, where the attacks based on GANs were implemented.

Firstly we analyze the effect of an attack on the IEEE 24-bus system with a model that had its objective function modified to raise the frequency of the system obtained from the state variables. This 24 bus-system was not included in the training process and yet



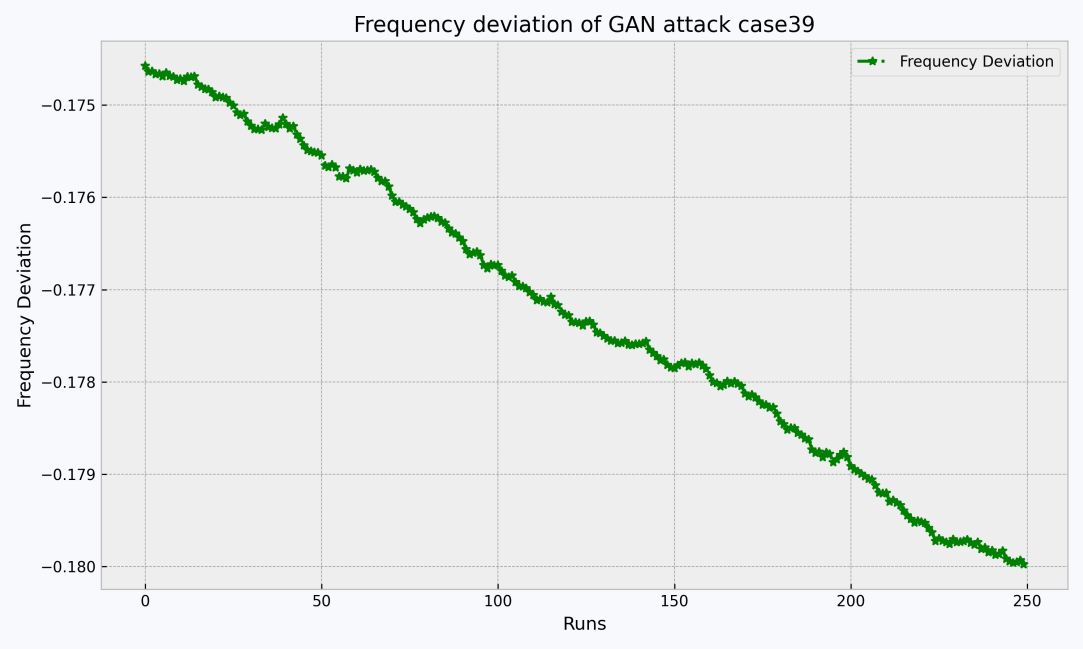
(a) IEEE 24-bus system

Figure 5.1: Analysis of Frequency Deviation with a GAN attack model trained to increase Deviation on IEEE 24-bus system

still as shown in Figure 5.1, we could observe a steady increase in system frequency over the course of the simulation. A shift of about $0.002Hz$ was obtained in about 50 steps of continuous attack on the state variables. The margin albeit small but with a longer duration of attack clearly indicates a potential to disrupt the grid network. By standards, a shift of about only $0.02Hz$ in a power grid requires a dispatch action to maintain the frequency and keep the power grid reliable which could be a generator dispatch or load shedding. Given that measurements are sent from PMU devices at millisecond intervals for WAMS or less than 10-second intervals in terms of SCADA systems, the system can definitely be affected within a short period of time within minutes if no further security constraints are embedded in the state estimation process at the control center. This is because at each of the time steps, the GAN model only slightly tweaks the measurement data and hence not exceeding the tolerated residuals of the state estimation. These modifications though minute subtly succeeds over time in forcing the control center to an action intended by the adversary due to the fact that the GAN was trained with an objective.

As for figure 5.2, a second GAN model which was trained by customizing its objective function to decrease the grid frequency was leveraged to perform the attacks on the measurement variables on the IEEE 39 bus-system. Again, this was a test network not included in the training phase and it could be observed that in about 250 timesteps, there was an overall frequency deviation of about $0.005Hz$ following a nearly linearly declining trend.

From the observed results it can be inferred that there is potential for GAN-based FDI attacks to cause huge disruptions in the power system. It can be observed that the trained GAN could mimic the statistical properties of valid measurement data and



(a) IEEE 39-bus system

Figure 5.2: Analysis of Frequency Deviation with a GAN attack model trained to decrease Deviation on IEEE 39-bus system

dynamically adapt them whilst achieving a predefined objective. One main aspect of the effect as discussed in the literature review is the economic and operational consequences. A successful frequency shift causing a generator dispatch could have a huge impact on the market operation as demonstrated in relevant research[30]. This can affect the whole pricing structure with heavy consequences for relevant stakeholders leading to huge losses. In comparison to traditional FDI attacks, which leverage the Jacobian matrix [5], the GAN model could prove to be a more feasible attack in a black-box scenario i.e little knowledge of the underlying attacked system. This is as a result of the GAN learning the distribution of relevant measurement data from historical real-time samples and can formulate attacks that could dynamically adapt to varying conditions of the system state. This makes it a more feasible form of approach. Also, attacks formulated with the Jacobian matrix are static and can be detected when the system conditions change as they are based on fixed parameters. The GAN-based approach produces attacks evolving over time imitating the real measurement data distribution hence bypassing the typical bad data detection algorithms and possibly slightly more sophisticated detection techniques.

The main challenge in the implementation of the GAN-based attack was in computational resources and the availability of high-quality data. Even with the data augmentation applied as described in the attack model of the GAN implementation, it wasn't easy to obtain enough data to be able to efficiently train the Generator and Discriminator. This led to training on fewer epochs along with a small batch size that affected the overall performance of the model. Another challenge was in the output of the state estimation results. Given the nature of the dynamic attacks, output state variables from the state estimation process were continuously leveraged as measurement data for the subsequent timestep. In the experiments, only four state variables namely the voltage magnitudes,

phase angles, active power injections and reactive power injections were available from the state estimation output. This limited running experiments on slightly larger networks such as the IEEE 300 bus system as it was difficult to have a more accurate output of the state estimation process with more redundant measurements.

5.4 Future Work and Research

From the case studies demonstrated it could be observed that the GAN attack models can prove to be a feasible model to craft successful attacks on the state estimation leading to grid instability. Nonetheless, there are gaps still required to be examined to further solidify the hypothesis. In the experimental setup, there are only a few standard IEEE test networks available and the model would have to be tested across a wide variety of different to be able to gauge the model's overall performance. Hence more networks would be required to further enhance the GAN generalizability in generating measurements for networks with varying parameters and conditions. In short, there is a need to have a larger dataset of real power grid networks for both training and testing to validate the methodology. Another issue was with the computational resources. All experiments were done on a local PC which had a limited capacity of Random Access Memory. For this reason, it was not feasible to train the model with a larger batch size and a huge number of epochs. Dedicated high computing resources can be leveraged to train a more powerful and efficient model. Furthermore, in terms of the objectives, only frequency deviation was considered. However, there are various other metrics to consider such as overall voltage deviation, line overloading etc. With multiple models trained for each of these objectives, an adaptive strategy could be implemented with a specified trained model selected at each timestep for a larger objective such as cascading failures. This can be research built on top of the experiments and methodology examined in this thesis.

With all these, however, is our goal to cause harm and create disruptions in the power grid infrastructure? Certainly not. Our goal in this thesis was to assess the vulnerability of the traditional defence mechanism of the power grid with respect to innovative modern tools currently available and fix the loopholes prior to being exploited by a malicious adversary. The end goal is then to build modern adaptive defence mechanisms to detect and mitigate such complexly crafted attacks. In building such modern defence mechanisms, we anticipate guaranteeing the reliability and efficiency of the smart grid enabling technological advancement with economic prosperity.

Chapter 6

Conclusion

This study involved assessing the vulnerability of traditional state estimation in power systems to False Data Injection Attacks. We explored attacks formed on the basis of stealthy models along with a novel approach with deep learning methods. The assessability is done to explore the vulnerabilities and find ways to mitigate them to avoid them being exploited by adversaries with malicious intentions.

In the first part of the thesis, an attack model based on network topology was explored. It was noticed that this is the most effective form of attack having a considerable effect on the state estimation outcome. Due to the nature of the attack model, full knowledge of network parameters is assumed, which makes it a less practical form of attack. The second part involved an attack model built on top of deep learning frameworks. Compared to the previous method, these had a less significant impact on a single static attack on the state estimation process. However, for a continuous form of attack, these deep learning models after training with defined objectives were able to achieve considerable impact over a period of time on the outcome of the state estimation. Also, it is worth knowing that during the attack phase, full knowledge of the system parameters is not required, which makes a practical attack model in real-world scenarios.

However, it is important to note that this new method was tested on only a few selected datasets available as standard networks. Initial results from the tested networks indicated the ability of such novel methods to have an impact on the state estimation. However, more experiments will be required to further validate the approach. There were also some limitations regarding the library used for the state estimation process. Reimplementing the experiments with more advanced libraries may provide a more comprehensive analysis of the impact of deep learning methods on state estimation.

Furthermore, the goal of this research was directed towards subsequently improving the operational reliability of modern power systems. The preliminary findings from the experiments highlight critical vulnerabilities in the state estimation framework, underscoring the need for the development of advanced countermeasures. Consequently, the next phase of this work should involve the formulation of advanced defensive strategies or learning-based models that can proactively detect and mitigate such threats. Evaluating these models against a variety of adversarial scenarios would further ensure their robustness and contribute to building more secure and reliable power systems.

Bibliography

[1] Distributed network protocols. *Information Theory, IEEE Transactions on*, 29:23 – 35, 02 1983.

[2] Fredrik Ege Abrahamsen, Yun Ai, and Michael Cheffena. Communication technologies for smart grid: A comprehensive survey, 2021.

[3] Jason Andress. *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*. Syngress, Amsterdam, The Netherlands, 2014. [DOI] [Google Scholar].

[4] Jam Angulo-Paniagua and Jairo Quirós-Tortós. Comparing chi-square-based bad data detection algorithms for distribution system state estimation. In *2020 IEEE PES Transmission & Distribution Conference and Exhibition - Latin America (T&D LA)*, pages 1–5, 2020.

[5] Adnan Anwar, Abdun Mahmood, and Mark Pickering. Data-driven stealthy injection attacks on smart grid with incomplete measurements. pages 180–192, 04 2016.

[6] Christoph Brosinsky, Dirk Westermann, and Rainer Krebs. Recent and prospective developments in power system control centers: Adapting the digital twin technology for application in power system control centers. In *2018 IEEE International Energy Conference (ENERGYCON)*, pages 1–6, 2018.

[7] Gu Chaojun, Panida Jirutitijaroen, and Mehul Motani. Detecting false data injection attacks in ac state estimation. *IEEE Transactions on Smart Grid*, 6(5):2476–2483, 2015.

[8] ElProCus. Smart grid technology working operation and applications, 2025. Image retrieved from article.

[9] Iowa State University Engineering. Security constrained optimal power flow.

[10] ENTSO-E. Entso-e operation handbook. `https://eepublicdownloads.entsoe.eu/clean-documents/pre2015/publications/entsoe/Operation_Handbook/Policy_1_final.pdf`.

[11] Francisco Gonzalez-Longatt. Chapter 1. introduction to power systems, 12 2019.

[12] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[13] Jonathan Greig. German wind turbine maker shut down after cyberattack, April 2022. Accessed: 2025-03-29.

[14] Anthony Hill*, Penrose Cofie, John Fuller, Justin Foreman, Kelvin Kirby andd Emmanuel Dada, Olatunde Adeoye, and Adeyemi Taylor. The power flow analysis for electric power network. *World Journal of Advanced Engineering Technology and Sciences*, 2023.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–1780, 11 1997.

[16] T. Krause, R. Ernst, B. Klaer, I. Hacker, and M. Henze. Cybersecurity in power grids: Challenges and opportunities. *Sensors*, 21(18):6225, 2021.

[17] Murat Kuzlu, Manisa Pipattanasomporn, and Saifur Rahman. Communication network requirements for major smart grid applications in han, nan and wan. *Computer Networks*, 67:74–88, 2014.

[18] Robert E. Larson, William F. Tinney, and John Peschon. State estimation in power systems part i: Theory and feasibility. *IEEE Transactions on Power Apparatus and Systems*, PAS-89(3):345–352, 1970.

[19] Weixiao Meng, Ruofei Ma, and Hsiao-Hwa Chen. Smart grid neighborhood area networks: A survey. *Network, IEEE*, 28:24–32, 01 2014.

[20] Paul W. Parfomak. *Physical Security of the U.S. Power Grid: High-Voltage Transformer Substations*. Congressional Research Service, Washington, DC, USA, 2014.

[21] Panagiotis Radoglou-Grammatikis, Panagiotis Sarigiannidis, Ioannis Giannoulakis, Emmanouil Kafetzakis, and Emmanouil Panaousis. Attacking iec-60870-5-104 scada systems. In *2019 IEEE World Congress on Services (SERVICES)*, volume 2642-939X, pages 41–46, 2019.

[22] Md. Ashfaqur Rahman and Hamed Mohsenian-Rad. False data injection attacks against nonlinear state estimation in smart power grids. In *2013 IEEE Power & Energy Society General Meeting*, pages 1–5, 2013.

[23] Tim Conwa Robert M.Lee, Michael J. Assante. Analysis of the cyber attack on the ukrainian power grid, 2016.

[24] Christian Rossow. Amplification hell: Revisiting network protocols for ddos abuse. In *Proceedings of the 21st Annual Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, USA.

[25] J. G. Sreenath, A. Meghwani, S. Chakrabarti, K. Rajawat, and S. C. Srivastava. A recursive state estimation approach to mitigate false data injection attacks in power systems. In *2017 IEEE Power & Energy Society General Meeting*, pages 1–5, 2017.

[26] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks, 2014.

[27] Vestas Wind Systems A/S. Third update on cyber incident, December 2021. Press Release, 12:00 CET, Aarhus. Accessed: 2025-03-29.

[28] T. P. Vishnu, Vidya Viswan, and A. M. Vipin. Power system state estimation and bad data analysis using weighted least square method. In *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, pages 1–5, 2015.

[29] Wikipedia contributors. Electricity grid simple - north america, 2024. Image retrieved from the Wikipedia article "Electrical grid".

[30] Le Xie, Yilin Mo, and Bruno Sinopoli. Integrity data attacks in power market operations. *IEEE Transactions on Smart Grid*, 2(4):659–666, 2011.

[31] Zhenyong Zhang, Ruilong Deng, David K. Y. Yau, Peng Cheng, and Jiming Chen. Zero-parameter-information fdi attacks against power system state estimation. In *2020 American Control Conference (ACC)*, pages 2987–2992, 2020.

[32] Yi Zhao, Xian Jia, Dou An, and Qingyu Yang. Lstm-based false data injection attack detection in smart grids. In *2020 35th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 638–644, 2020.