Real-Time Detection of GPS Spoofing Attack with Hankel Matrix and Unwrapped Phase Angle Data

IMTIAJ KHAN

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Electrical Engineering

Virgilio A. Centeno, Chair
Chen-Ching Liu
Vassilis Kekatos

11/15/2021
Blacksburg, Virginia

Keywords: FDIA, GPS-spoofing, PMU, Unwrapped, Hankel matrix

`

Real-Time Detection of GPS Spoofing Attack with Hankel Matrix and Unwrapped Phase Angle Data

Imtiaj Khan

ABSTRACT

Cyber-attack on synchrophasor data has become a widely explored area. However, GPS-spoofing and FDIA attacks require different responsive actions. State-estimation based attack detection method works similar way for both types of attacks. It implies that using state-estimation based detection alone doesn't give the control center enough information about the attack type. This scenario is specifically more critical for those attack detection methods which consider GPS-spoofing attack as another FDIA with falsified phase angle data. Since identifying correct attack type is paramount, we have attempted to develop an algorithm to distinguish these two attacks. Previous researchers exploited low-rank approximation of Hankel Matrix to differentiate between FDIA and physical events. We have demonstrated that, together with angle unwrapping algorithm, low-rank approximation of Hankel Matrix can help us separating GPS-spoofing attack with FDIA.

The proposed method is verified with simulation result. It has been demonstrated that the GSA with 3 second time-shift creates a low-rank approximation error 700% higher than that of normal condition, whereas FDIA doesn't produce any significant change in low-rank approximation error from that of normal condition. Finally, we have proposed a real-time method for successful identification of event, FDIA and GSA.

Real-Time Detection of GPS Spoofing Attack with Hankel Matrix and Unwrapped Phase Angle Data

Imtiaj Khan

GENERAL AUDIENCE ABSTRACT

Cyber-attack on synchrophasor data has become a widely explored area. However, GPS-spoofing and FDIA attacks require different responsive actions. State-estimation based attack detection method works similar way for both types of attacks. It implies that using state-estimation based detection alone doesn't give the control center enough information about the attack type. This scenario is specifically more critical for those attack detection methods which consider GPS-spoofing attack as another FDIA with falsified phase angle data. Since identifying correct attack type is paramount, we have attempted to develop an algorithm to distinguish these two attacks. Previous researchers exploited low-rank approximation of Hankel Matrix to differentiate between FDIA and physical events. We have demonstrated that, together with angle unwrapping algorithm, low-rank approximation of Hankel Matrix can help us separating GPS-spoofing attack with FDIA. The simulation result verifies the next chapter discusses our proposed algorithm on GPS-spoofing attack detection and its ability to distinguish this type of attack from conventional FDIA.

The proposed method is verified with simulation result. It has been demonstrated that the GSA with 3 second time-shift creates a low-rank approximation error 700% higher than that of normal condition, whereas FDIA doesn't produce any significant change in low-rank approximation error from that of normal condition. Finally, we have proposed a real-time method for successful identification of event, FDIA and GSA.

Acknowledgements

I am thankful to all of those who have been with me along the journey. Special thanks to my academic advisor Dr Virgilio Centeno who provided me with the opportunity to work on synchrophasor data provided by Pacific Northwest National Laboratory (PNNL). He also gave me the perfect learning environment along with the setup and guidance, which was pivotal for my work.

Moreover, I would love to thank Dr Vassilis Kekatos and Dr Chen-Ching Liu for their guidance and course works, which strengthened my foundation on optimization technique and cyber physical systems. Furthermore, I thank my lab partners, specially Alok Kumar, who helped me in my understanding of synchrophasor systems.

Last but not the least, I would love to show my gratitude toward my friends and family, especially my elder brother Imran Khan.

And finally, I cannot thank my beloved wife Monica enough, her support kept me focused during the whole research project.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

"Smart Grid" has become a buzzword since the early 21st century. The revolutionary development of electronic communication system paved the way for the smart grid concept. Due to the increasing demand of power in the utility side as well as the advent of renewable energy resources, conventional centrally controlled power grid is being replaced by the distributed grid system. Distributed grid topology along with smart metering devices, demand response and demand response constitute the concept of smart grid.

According to the "*Energy Independence and Security Act of 2007 (EISA-2007)*" [1], a smart grid must follow the following criteria: utilizing the digital information and controls technology for the improvement in reliability and efficiency of the grid, ensuring cyber-security, dynamic grid optimization, integrating distributed and renewable resources, deploying smart metering devices with the ability of performing real-time monitoring and protective actions, inclusion of plug-in electric vehicles, and the last but not the least, establishing two-way communication.

Another important component of smart grid is the electricity market. An electricity market is like conventional stock market, with energy being the product which is bought and sold continuously. An electricity market can either be regulated or deregulated, where deregulated market is becoming a dominant factor in the distributed grid system.

In short, smart grid corresponds to an interconnected system of communication technology, electronic devices, physical power system and cyber system.

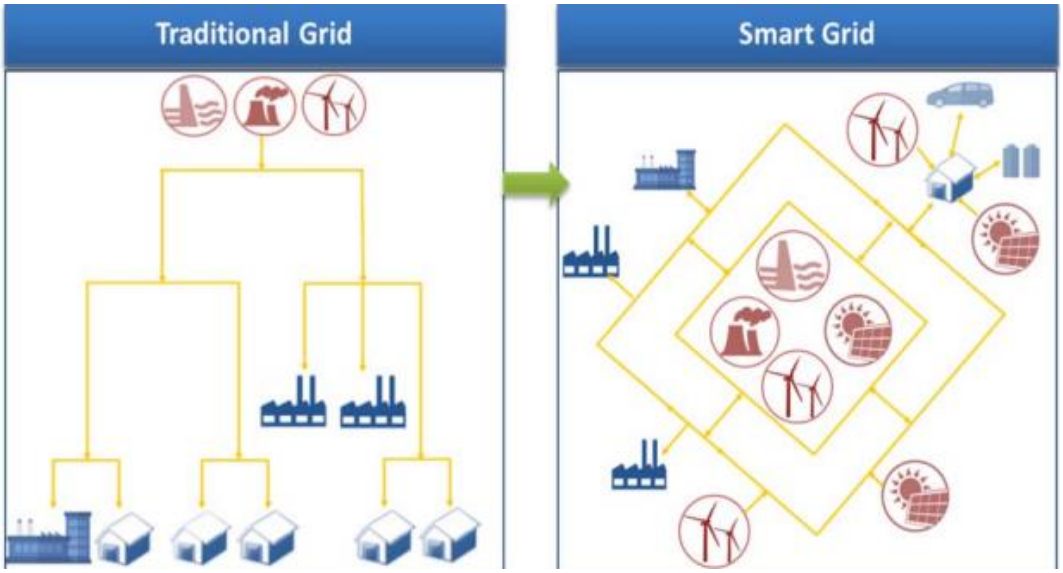


Figure 1 Difference between tradition grid and smart grid

Transition of traditional grid into smart grid shifted the network topology, especially in the distribution system. At the distribution side, the grid is no longer unidirectional radial system with one-way communication method, rather the system has become a multi-dimensional one with two-way communication method, incorporating active participation from the end-users [2].

The current trend of transforming a traditional grid into a smart one brings about several benefits in the grid-operation. The advantages of smart grid can be summarized as follows [3]:

- It provides better situational awareness.
- Maximize utilization of assets.
- Higher efficiency
- Higher reliability with the help of autonomous control
- Higher resiliency against cyber-attack
- Utilizing renewable energy resources is a step toward global energy sustainability
- Increased market efficiency by using bundled products of energy and performing ancillary services.
- Lower power quality issues such as voltage sag, swell, oscillation events, etc.

In a smart grid, Information and Communication Technology (ICT) devices and systems are deployed for the monitoring and maintenance of the physical grid. The ICT devices and systems

are referred as the cyber system and the physical power grid is referred as the physical system. Combining the cyber system and physical grid begets a complex infrastructure, which is known as Cyber-Physical System (CPS) [4].

1.2 Cyber Physical System (CPS)

An important component of a CPS is the Supervisory Control and Data Acquisition (SCADA) system, which is a computer-based system dedicated for gathering and analyzing real-time data. SCADA is used to monitor and control time-sensitive critical infrastructure data dynamically [5]. CPS of smart-power grid is like other systems that uses SCADA, such as natural gas pipelines, highway transportation system, water management system etc. In all these cases the cyber system is similar, however, the physical system is different. In a CPS, co-simulation of continuous time and discrete event is necessary to ensure the dynamic operation. In [6], a co-simulation environment with three layered architectures has been proposed to study the real-time supervision and control of power grid by Transmission Operator (TO) as well as to assess the SCADA and communication security performance. The model is also a test-bed to investigate the cyber-security of CPS.

As mentioned before, physical side of CPS is the power system, where both the static and dynamic models are used. The power grid elements are divided into substations. The cyber side of CPS includes SCADA functionalities. SCADA is used for real time communication between the physical grid and Transmission Operator (TO). SCADA provides the ICT functions which are necessary for the TO. Primary functions of SCADA are gathering measurement data and breaker status, establishing communication over the Wide Area Network (WAN) and control of Remote Terminal Units (RTUs) and Intelligent Electronic Devices (IEDs). The TO receives power system data through SCADA and take actions to maintain stability and safety of the grid.

Fig. 2., demonstrates a sample CPS model, where the three layers are arranged in hierarchical structure. The physical layer is divided into substations and few substations form a region. Each region is controlled by a control center. The cyber layer has two labels: control center label and substation label. In the substation label of cyber layer, the ICT devices are connected with the power devices through Local Operating Network (LON). The power devices are IEDs, RTUs etc.

The Local Area Network (LAN) of the control center has Human Machine Interfaces (HMIs), remote access stations, firewalls, routers and servers.

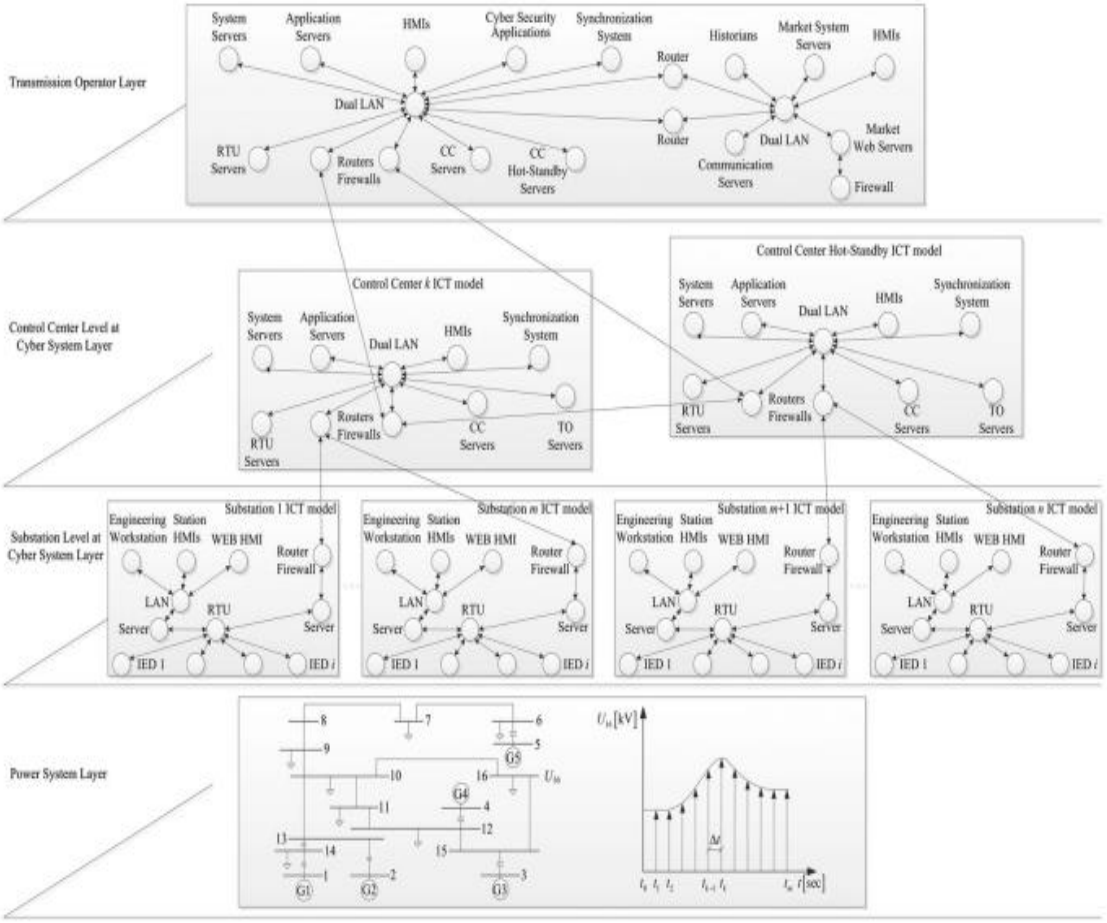


Figure 2 Cyber Physical System [6]

The TO and the control center of cyber layer use several ICT devices such as workstations servers, redundant application servers, system servers, hot standby connections, firewalls and routers. Routers are used to established two-way communication. TO receives real time data from the substation, where LAN is used to establish interconnection between the ICT components. TO uses the real time data through the ICT network and perform state estimation, unit commitment, economic dispatch and stability and safety checks. Analyzing the result, TO sends required commands to the substation with the help of two-way communication method [7].

TO has two separate parts: Energy Management System (EMS) and Market system. Data from substation that are fed into the control center are sent to the TO's EMS using SCADA. SCADA sends data to EMS at a rate of 4 or 6 frame per second. Though this rate is fast enough for small area network, it is inadequate for Wide Area Monitoring (WAM). Since the communication between ICT devices are incorporated with First-In-First-Out (FIFO) queuing method [6], the WAM may cause higher delay in communication network. The transmission rate in SCADA system may face loss in data packet due to this delay [8]. Therefore, a faster frame rate is required to ensure that data-loss is minimized. Moreover, this issue may be resolved by using synchronous or centrally timestamped data. Having time-stamped data has several advantages, such as the ability of tracking voltage and current phase angle and magnitude values, as well as post-analysis of electrical events. Power network stress is marked by separation in phase angle that helps in real-time monitoring and post analysis especially in case of contingencies and blackout. Therefore, Phasor Measurement Unit (PMU) has become an integral part of CPS which uses GPS-signal to create accurate time-stamped data [9].

1.3 Phasor Measurement Unit (PMU)

Synchrophasor data refers to the time synchronized phasor data, i.e., phase angle and magnitude of the sinusoidal voltage and current signals. Fig. 3. Depicts the synchrophasor form of the signal x(t).

The sinusoidal signal x(t) can be expressed as:

$$x(t) = X_m \cos(\omega t + \phi) = \left(\frac{X_m}{\sqrt{2}}\right) e^{j\phi} \quad (1)$$

Magnitude of the signal is the $\left(\frac{X_m}{\sqrt{2}}\right)$ from eqn (1) and the phase angle is the ϕ. The value of ϕ depends on the time scale. It refers to the temporal position of instantaneous signal from the reference time (t = 0) at a nominal frequency synchronized with Universal Time Coordinate (UTC). ϕ is 0 degrees when UTC second rollover (1 Pulse/ Second or PPS) occurs for the first time. At nominal frequency f_0 (50 Hz or 60 Hz), the sinusoidal signal can be expressed as phasor form as in eqn (2).

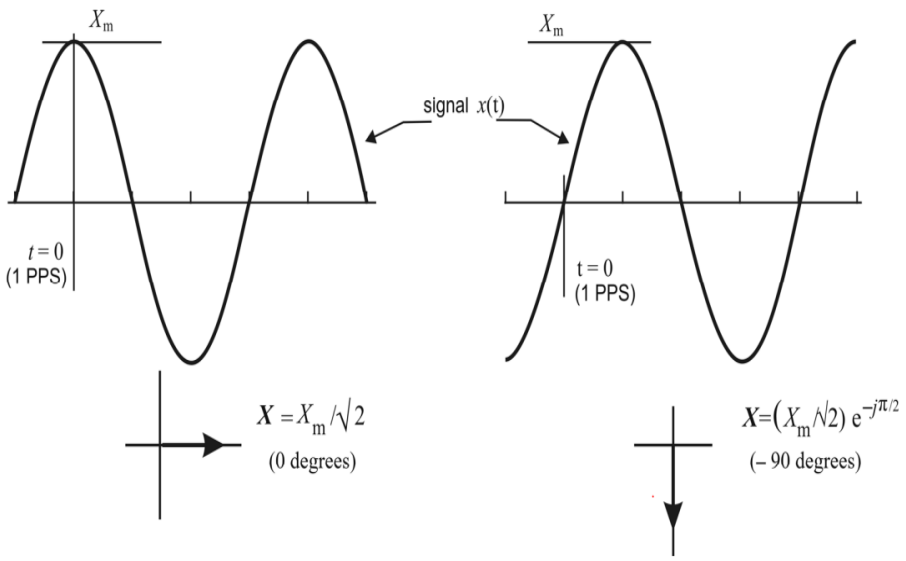


Figure 3 Synchrophasor representation [10]

$$x(t) = X_m \cos(2\,\pi\, f_0 t + \phi) \qquad (2)$$

In practical scenario, both magnitude X_m and frequency f are functions of time. When the time varying frequency f(t) is different from the nominal frequency f_0, it is called off-nominal frequency. g = f(t) – f = Δf is the offset between off-nominal and nominal frequencies. The phasor representation for off-nominal frequency becomes:

$$x(t) = \frac{X_m}{\sqrt{2}}\, e^{j(2\pi \int g dt + \phi)} \qquad (3)$$

If the frequency offset from nominal frequency is constant Δf remains constant, eqn (3) becomes [10]:

$$x(t) = \frac{X_m}{\sqrt{2}}\, e^{j(2\,\pi\,\Delta f\; t + \phi)} \qquad (4)$$

For a constant offset, the magnitude of sinusoidal signal is also constant, however the phase angle will change at a rate of $2\pi(f - f_0)T_0$. The phase angle values will increase up to 180^o, and then wraps around to -180^o (fig 4).

One of the key PMU features is high reporting rate. Generally, the reporting rate of 30 frame/second or 60 frame/second. For 30 frame/sec, the PMU data transmission rate is

approximately 120 times faster than SCADA. Data transmission rate depends on the nominal system frequency as described in table 1.

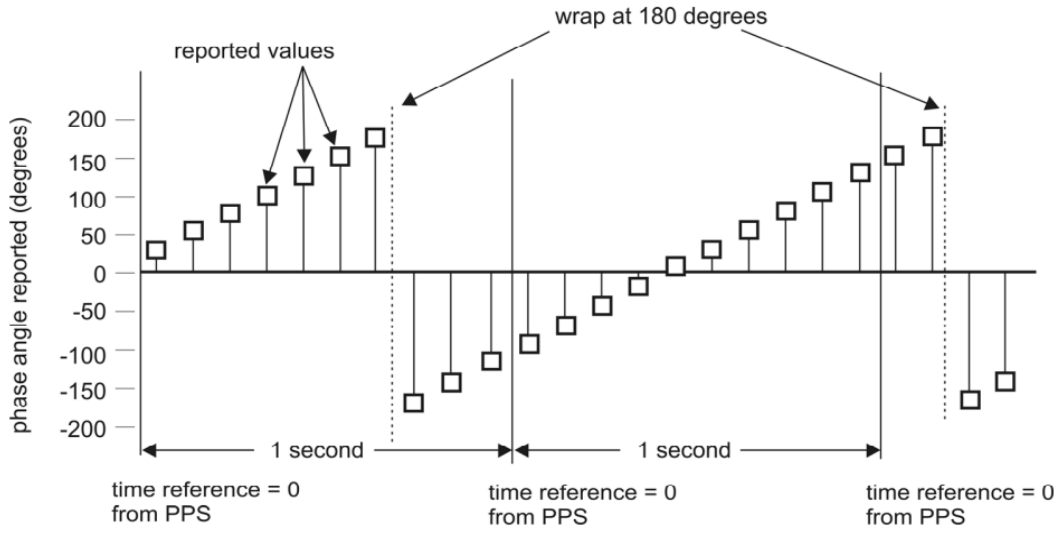


Figure 4 Phase angle variation for constant off nominal frequency [10]

Table 1 PMU reporting rates [10]

System frequency	50 Hz			60 Hz					
Reporting rate F_s (frame/second)	10	25	50	10	12	15	20	30	60

1.3.1 PMU Functional Block

A PMU generally uses signal processing block and Analog to Digital Converters (ADC) for processing synchrophasor data. The analog signal is passed through a low pass filter which is then converted into digital signal. UTC time stamp is received from GPS signal. The time stamp is assigned to the digital signal with oscillator. The output is put into a phasor microprocessor where the magnitude and phase angle are calculated using Discrete Fourier Transform (DFT) (fig. 5). For a set of singe phase signal $\{x_i\}$, the synchrophasor estimate for i^{th} sample time can be expressed as:

$$X(i) = \sqrt{2} \times \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} x_{i+k} \times W_k \times \mathrm{e}^{(-j(i+k)\Delta t\, \omega_0)} \quad (5)$$

where,

$\omega_0 = 2\,\pi\, f_0$ (f_0 is nominal frequency)

N = FIR filter order

Δt = 1/ sampling frequency

$x_i = i^{th}$ sample

W_k = low-pass filter coefficient

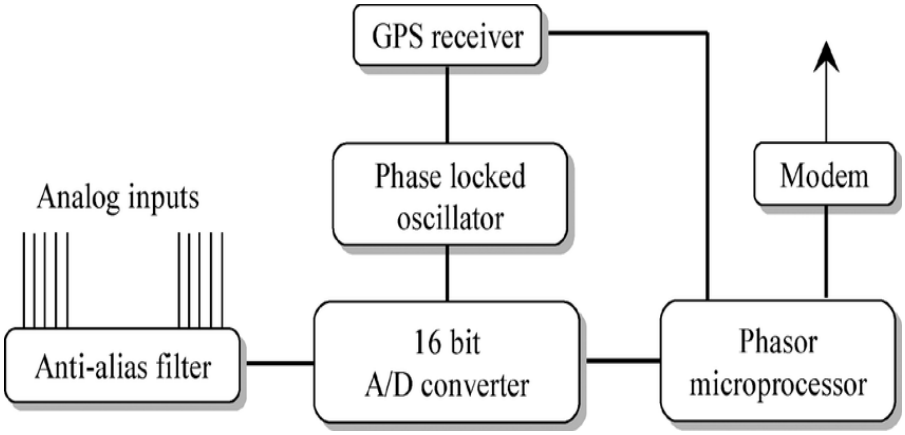


Figure 5 Functional block diagram of PMU [9]

1.3.2 Time-tagging

The time-tagging of synchrophasor measurement with UTC follow IEEE 37.118.2 standard protocol [12]. According to this protocol the time information has three parts:

- second-of-century (SOC) count,
- fraction-of-second (FRACSEC) count, and
- message time quality flag.

SOC refers to the number of seconds passed from a reference time instance. SOC is a 4-byte binary number, where the reference point in time is the midnight of January 1, 19710 (00:00:00 UTC time). With 32-bit SOC value, the current format can provide timestamp for 136 years. To address the issue of leap second, a second is added or deleted to make the timestamp consistent with UTC. When leap second is added, two consecutive second count have same SOC value, whereas a second

is removed when leap second is deleted. The SOC will roll over to 0 in 2106, same as the UNIX, LINUX, and DOS system.

Since PMU data transmission rate is generally 30 or 60 fps, the timestamp needs to configure to describe the fractional second count, which is represented by the FRACSEC count. FRACSEC is a 32-bit binary number, where the first 8 bits represents the message time quality and the last 24 represents the fraction of second count. The fraction of second count is divided by the TIME_BASE to get the fractional timestamp. The TIME_BASE information is provided by the configuration frame of the data transmitted by PMU.

$$Time = SOC + \frac{FRACSEC}{TIME_{BASE}}$$

The 8-bit message time quality has two parts. Bit 6 to 4 describe the leap-second information, and bit 3-0 describe the range for maximum time-error from UTC.

1.3.3 Synchrophasor Connections

PMU provides synchrophasor data as well as frequency and Rate of Change of Frequency (ROCOF) estimates. On demand, PMU transmits megawatts (MW) and megavars (MVAR) in Boolean status word. PMU can either be a single device with synchrophasor ability or it can be combination of devices such as protective relay, DFR, meter etc. PMU's data is sent to Phasor Data Concentrators (PDC) through a communication network. For the case when there are multiple Intelligent Electronic Devices (IEDs) with synchrophasor measurement capabilities in a single substation, multiple data streams are passed to a hardware PDC in a real time manner. PMUs and PDCs can also store data locally to be used in off-line calculations. The main task of a PDC is to aggregate the PMU data, however there are additional software developments that allow PDCs additional functionalities such as visualization of information, providing limited situational awareness, providing partial analytical and control functionality.

From structural point of view, PDC can be regarded as node in the network. For simpler network and low number of PMUs, one single PDC is used to aggregate data from all the PMU. However, in practical scenario, there is large number of synchrophasor data-streams. Local PDCs are connected to clusters of PMUs, and the output stream of the several local PDCs are fed into another PDC, creating a PDC-PDC communication channel and using the same protocol as IEEE 37.118.2.

The second layer PDCs may again fed the synchrophasor data-stream into either next layer of PDC or to the control center, depending on the grid structure. PMU or the lower level PDC sends synchrophasor voltage and current magnitudes and angle data, frequency estimates, ROCOF values along with the time stamp information to the next layer PDC. PDC then align all synchrophasor measurements with the timestamp it receives using SOC and FRACSEC information.

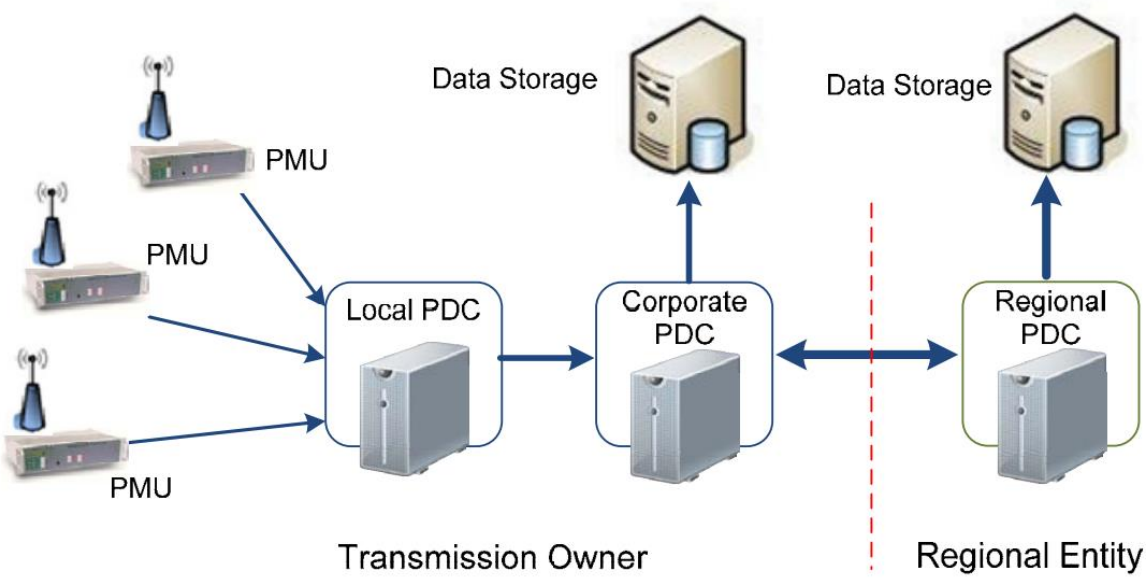


Figure 6 PMU and PDC Network [10]

System operators can make PDC perform some additional tasks such as:

- Quality check and insert necessary flag.
- Checking disturbance flag
- Record data for off-line analysis
- Displaying the result
- Working as a interface to SCADA system

The first layer local PDCs do the aggregating and time-alignment of synchrophasor data. The mid-layer PDCs receive data and perform quality checks and aggregate the data into next layer. The highest layer PDC, which is also referred as the SuperPDC archive data for off-line use. Similar to PMU, PDC can be either a stand-alone device or can be a collection of devices. Also, PDC can be a system of hardware and software and different parts of the hardware/ software operates

individually. PDC can be a useful tool for cyber-security analysis of the power grid, as well as for providing additional protection for PMUs.

1.4 Cyber-Attacks on PMU

Despite PMU is reliable in terms of time-tagged data and accuracy, it is still susceptible to various type of cyber-attacks [13] [14] [15]. The most common type of cyber-attack against PMU is the malicious data injection attack such as false data injection attack (FDIA). Other types of malicious data injection attacks include data modification and replay attacks [16]. In FDIA, the attacker injects bad data into the system and thereby modifying the power system data fed into the control center. If the state estimation is performed with falsified data, the control center may take wrong decision such as taking restorative action or emergency shut down even though no operating condition is not violated. Another frequent type of attack that can hamper PMU integrated grid performance is Denial of Service (DoS) attack. In this case attack inject large volume of data through the communication channel, therefore the data transfer to destination gets blocked due to the large volume of data traffic [17] [18]. Man In the Middle (MITM) and Side Channel attacks are also critical for PMU [19] [20].

Since PMU relies on GPS signal for time synchronization, it opens the possibility of a new type of attack: GPS-spoofing attack. PMU generally uses public GPS which is vulnerable to cyber-attacks. GPS-spoofing attack poses very serious concern over the cyber security of the CPS [21], such as misoperation of transmission line differential relay [18]. In this type of attack, the attackers fake the GPS signal and become able to modify the timestamp of PMU data [22] and the PMU signal after attack looks similar in shape to the signal with no attack [23] [24].

From the defender's perspective, protection of PMU integrated smart grid against cyber-attacks is crucial for smooth operation of the grid. Most works are based on detecting FDIA. Detection of FDIA is like conventional Bad Data Detection (BDD) method [25]. Conventional BDD algorithms observe the residues of the measured and expected variables and do statistical test to find the outliers. GPS-spoofing attack can also be considered as a type of BDD, since modification in time-stamps lead to the shift in phase angles. In GPS-spoofing attack, the voltage and current magnitude remain unchanged. As a result, it is similar to a FDIA with only the phase angle data modified. However, many researchers focused on creating attack that cannot be detected by BDD [26] [27].

This statement is also true for GPS-spoofing attack if the attack is considered as a variation of FDIA with corrupted phase angle data [28]. These types of undetectable and stealthy attacks can still be prevented by placing PMUs into optimal places of the grid [29].

Even though FDIA and GPS-spoofing attacks both can damage the grid operation by feeding falsified measurements into the control center, distinguishing these two types of attacks is necessary. Both types of attacks can be detected using BDD methods. However, the restorative actions are different for GPS-spoofing attack. Therefore, algorithm needs to be developed to differentiate GPS-spoofing attack from FDIA

1.5 Outline

This thesis work aims at distinguishing FDIA from GPS-spoofing attack. For this purpose, low-rank Hankel Matrix property and unwrapping of phase angle data can be two useful tools. In section 2, literature review on cyber-attack on PMU is discussed. In section 3, the proposed methodology using low-rank Hankel Matrix property and unwrapped phase angle data is described.

In section 4, the simulation setup is explained. The UTC timestamp is created with MATLAB *daytime* function. We have used synchrophasor data from an IEEE 13 node test system in SIMULINK. The GPS-spoofing attack is modelled as injecting a time-delay to the UTC timestamp.

In section 5, the simulation result is discussed. It has been observed that the proposed algorithm can distinguish GPS-spoofing attack and FDIA using low-rank Hankel Matrix approximation. The GPS-spoofing attack demonstrates higher low-rank approximation error than that of FDIA, thereby the higher error for low-rank approximation indicates a GPS-spoofing attack.

Section 6 describes the concluding remarks.

Chapter 2

Previous Works on Cyber Attack on PMU

Cyber-attack on a CPS refers to the security breach of the cyber system, which will impact the stability and safety of grid operation. One of the notable incidents of cyber-attack on power system is the Ukraine power grid hack in 2015 [30]. It is the first confirmed successful cyber-attack on a power grid and the perpetrator was Russian hacker group "Sandworm". Since then, works on cyber-attack boosted up. Synchrophasor communications and PMU devices are also vulnerable attacks. Therefore, special focus has been given for the research on cyber-attack against PMUs. Cyber-attack on PMU integrated CPS can be investigated from two perspectives: 1) Attackers' and 2) Defenders'. From attackers' perspective, the goal is to exploit the vulnerabilities of PMU. Most attacks covered in previous literature are on FDIA. Timing attack has also been explored by several researchers.

There are three aspects of from the defenders' perspective: 1) preventive, 2) detection and 3) responsive. Detection of FDIA has been the focus of most of the research works, whereas some research works have been done on pre-deployment of PMUs to the optimal positions to prevent FDIA. The responsive part is based on the accurate detection of attack, and it is usually done by taking countermeasure to mitigate the deviation of measured values from nominal points (fig 7).

2.1 Creating Undetectable Attacks

For an n bus system, the state variable can be represented as the vector $x = [x_1, x_2, \ldots . x_n]$ and the measurement variables can be represented with the vector $z = [z_1, z_2, \ldots . z_m]$. Here m is the number of meters (m<<n). The relation between z and x can be expressed as

$$\boldsymbol{z = H x + e} \qquad \textbf{(6)}$$

H is the Jacobean matrix that represents the non-linear relation between the state variable and the measurement variable and $e = [e_1, e_2, \ldots . e_m]$ is the measurement error vector. Any kind of attack will go undetected if $|| z - H \hat{x} ||_2$ is less than threshold τ, where x is the estimated state variable which can be calculated from the relation $\hat{x} = (H^T H)^{-1} H^T z$. If the attack **a** is applied, the measurement vector under attack will be $z_a = z + a$ and the estimated state variables under attack will be $\widehat{x_a} = \hat{x} + a$. The estimated attack vector is $c = (H^T H)^{\{-1\}} H^T$a. After attack, $|| z\_a - H \widehat{x_a} ||_2 = || z + a - H \hat{x} - H c ||_2 + || a - H c ||_2$. If a = Hc, then $|| z_a - H \widehat{x_a} ||_2 \leq \tau$. Therefore, the attack goes undetected [1].

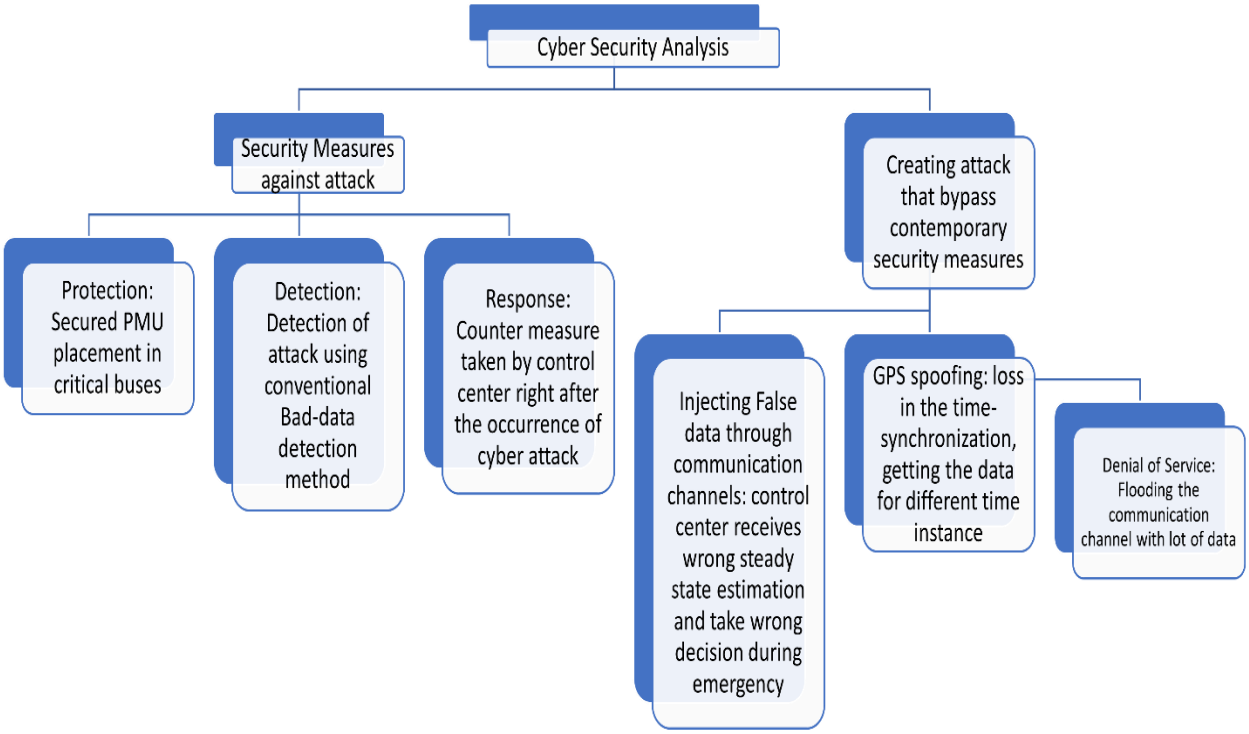


Figure 7 Research Trend in Cyber Security Analysis for PMU

This type of undetectable attack relies on the threshold τ of the control center. Also, the attack depends on the H matrix, which corresponds to the PMU locations in the bus. To defend against this attack, the defender can arrange the PMUs in specific locations so that the defender cannot exploit the H matrix to find an undetectable attack, thereby making the attack infeasible. If S refers

to the set of protected measurement and $\bar{S}$ is the set of unprotected measurements, then the matrix H^S represents the protected portion of H matrix. $H^{\bar{S}} = H - H^S$. The attacker needs to create an undetectable attack using the unprotected portion of H matrix. For this purpose, they need to satisfy following:

$$H^S c = 0 \tag{7}$$

When the number of protected measurements is very large, close to the number of buses n, the H^S becomes a full rank matrix. Thus $rank(H^S) = n$, which implies the solution of the eqn. (7) is zero, or c = 0. The attacker can only create a zero vector for attack, making the attack infeasible. However, for the case of $rank(H^S) \ll n$, there exists infinite number of non-zero solutions c for eqn (7). The attackers try to damage the grid operation and make a significant impact. As a result, c must be greater than a significance threshold γ [31]. The attacker needs to solve the following optimization criterion to create the optimum c vector:

$$\begin{aligned} &\min_c \left|\left|H^{\bar{S}}\right|\right|_0 \\ &s.t. H^S = 0 \\ &\left|\left|c\right|\right|_\infty > \gamma \end{aligned} \tag{8}$$

The solution for the optimization criterion (8) must be sparse to make the attack feasible. To find a sparse c vector is a NP-hard one, since the constraints in (8) are non-convex functions. One of the possible solutions is to make a random attack vector c, which will be perceived as a white noise by the control center.

l_1 relaxation of the second constraint of (8) can be another possible solution to get the sparse optimum solution faster.

J. Hao et. al. [32] proposed a probabilistic shrinkage function-based estimation of attack vector for the attacker. Attack vector can be created by shrinking a column vector of the basis matrix U from the null space of $F = (H(H^T H)^{-1} H^T - I)a$. u vector is the column with largest variance from U. u vector is adjusted with scaling factor ϵ_0 which is:

$$\epsilon_0 = \frac{C}{\max(u)}$$

The shrinkage function S(x) can be expressed as:

$$S(x) = \frac{x}{|x|-t}\max(|x| - t, 0)\,; \quad \text{for } |x| \neq t \qquad (9)$$

$$= 0; \text{ for } x = t$$

Therefore, the attack vector = $S(\epsilon_0 u)$, which must be random variable by keeping threshold t within the range $[-3\,\sigma,\ 3\,\sigma]$. The key idea of this model is that the $null(F)$ has most elements are small values, with variance less than the variance of white noise, thus will be considered as random white noise by the control center. However, there are small number of large values, which can be regarded as random noise by shrinking the attack vector with eqn. (9).

Most previous works on attacker's perspective assume that the attacker has the complete knowledge of grid topology. If the attacker has knowledge of the complete topology of the grid, they can use the H matrix information to create the attack vector c described previously. However, the algorithms discussed previously cannot be used if the whole H matrix can't be determined. This scenario occurs when the attacker doesn't know the complete topology of the grid. To tackle this issue, Y. Li et. al. [33] developed a Kernel Independent Component Analysis (ICA) based attack model. According to this model, a partial knowledge of grid topology and H matrix is enough to create undetectable attack. Attacker may know H_p, a part of the H matrix, each row p refers to each known bus of the grid. The method takes advantage of the Independent Component Analysis (ICA), where consecutive measurements over a window is required to formulate the attack. ICA of the estimated measurement is $Hx = H\ A\ \nu$. A is the impurity matrix and ν is the independent component. The HA can be written as the Jacobian matrix H_p, i.e. known portion of H matrix. Attacker tries to reduce the probability of detection, which can be expressed as:

$$V = 1 + \exp\left[\left(\frac{a}{L}\right)^T D\ \left(\frac{a}{L}\right)\right] \qquad (10)$$

Reducing V means increasing the inverse of V, which is the probability of attack being undetectable. Moreover, the attackers need to ensure the formulated attack is profitable, which is analytically expressed as $n^T a$, n is the cost. The overall undetectability of the attack is:

$$U = p\, n^T a + q\,\frac{2}{V} \tag{11}$$

Since the U must be maximized, the attackers will create the attack vector a by solving the following optimization formula:

$$\begin{aligned} &\max_{a} U \\ s.t. &\left|\left|a - H_p c\right|\right| < \tau \\ &N(a + L) \leqslant 0, \\ &a^T\, \bar{N}\, a = 0, \end{aligned} \tag{12}$$

L is the measurement vector z; c corresponds to estimated attack vector $c = (H^T H)^{-1} H^T a$. N = $diag(m)$ where m is the number of measurements.

$$\begin{aligned} N(i,i) &= 1; \text{ if the } i^{th} \text{ measurement is attacked} \\ &= 0; \text{ otherwise} \end{aligned}$$

2.2 Creating GPS-Spoofing Attack

Another important type of attack is GPS-spoofing attack, where the attacker interferes with the GPS signal and provide falsified time-stamp to the PMUs. GPS-spoofing attack will lead to an equivalent time-delay into the angle calculation. Therefore GPS-spoofing attack can also be referred as timing attack.

During timing attack, the phasor data are shifted in time axis. Timing-attack can be regarded as a special case of FDIA, where only the phase angle measurements are modified. The voltage, current and frequency magnitudes remain same. Let α_i be the phase shift due to timing attack. It is the phase angle difference between the actual values and the measured values for i_{th} sample. The change in the phasor measurement z can be written as:

$$\Delta z = z_m(\cos \alpha_i + j \sin \alpha_i - 1\,) \tag{13}$$

The attack will go undetected if and only if $\left|\left|(H(H^TH)^{-1}\,H^T - I)\Delta z\right|\right|_2 = \left|\left|F\,\Delta z\right|\right| = 0$ [31]. Assuming M is the set of the measurements and **a** is the attack vector with $a = [a_1, a_2, \ldots a_m]$, A is the set of the elements of attack vector, p is the cardinality of A. We can consider an attack measurement indicator matrix $\psi\ (m \times 1)$ which is:

$$\psi_j = 1; for\ j \in A$$
$$=0;\ \text{for}\ j \notin A$$

The attack vector can be formulated as $W = \psi^T\ diag(z)^*\ F^*\ F\ diag(z)\psi$ and the condition for undetectability in [31] becomes:

$$W(\bar{u} - 1) = 0$$

Here $(\cdot)^*$ is the conjugate function $\bar{u} = \cos\alpha_i + j\sin\alpha_i$. Using the attack vector W, the attacker will not be able to create a successful attack by shifting one phase angle only. In this case, only possible shift in phase angle is 0 [28]. However, it is possible to construct a successful attack by modifying two phase angle data. For this case, the undetectable attack vectors are:

$$\alpha_1 = 2\arg\left(W_{1,1} + W_{1,2}\right) mod(2\pi)$$
$$\alpha_2 = -2\arg\left(W_{1,1}\right) + 2\arg\left(W_{1,1} + W_{1,2}\right) mod(2\,\pi) \quad (14)$$

2.3 Detection of Attack and Countermeasure for FDIA

Several works have been done from the defender's perspective. The defenders' goal is to maximize the attack detection probability and make the communication channel secure. Most works against cyber-attack are focused on the conventional χ^2 based bad data detection scheme.

A Mixed-Integer Linear Programming (MILP) based attack prevention method was proposed in [34]. In the proposed method, the threat level is assessed initially. The maximum threat level of the PMUs in the system is minimized using optimization technique. A PMU is disconnected if the threat level associated with it is higher after first minimization step.

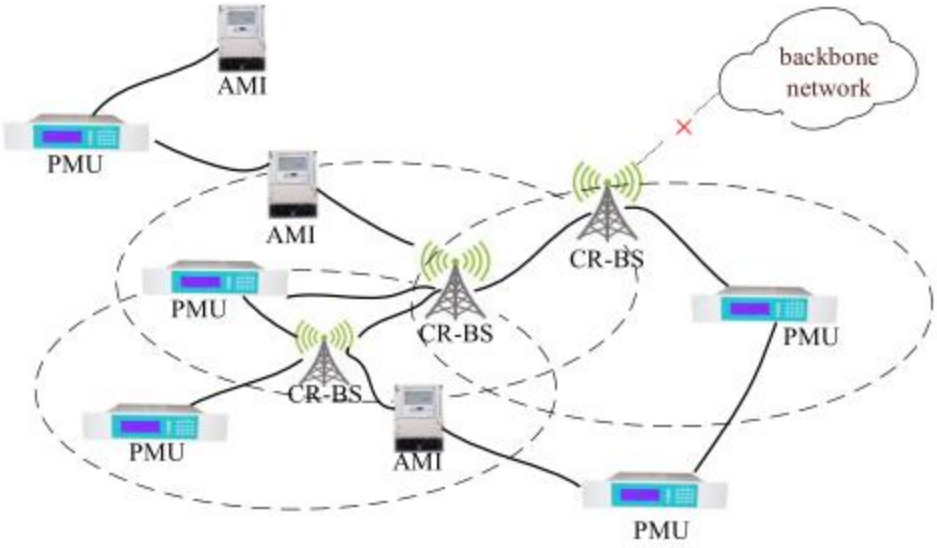


Figure 8 Cognitive Radio based PMU Data Transmission [35]

Vulnerability due to cyber-attack was analyzed for conventional the measurements and data-format of PMUs [35]. Independent Component Analysis (ICA) based signal separation model is utilized to create attack on PMU measurements. The authors also proposed a Cognitive Radio (CR) based network topology to ensure secure PMU data transmission. The CR based topology makes PMUs and AMIs access the Cognitive Radio-Base Station using multi-hop relay nodes with other PMU/AMIs. PMUs and AMIs can access the CR base station directly too, depending on the grid complexity. The data transmission spectrum is not constant. CR base stations has the ability to change the transmission spectrum dynamically. This structure makes it harder for the attacker to cause an intrusion.

Another aspect of the proposed topology in [35] is the network being not directly connected to the backbone network (Fig 7). The CR-BSs having high spectrum hole and enough wireless-link capacity ensures the network security against cyber-attack. From the result, it can be observed that lower Package Loss Rate (PLR) and higher Data Authentication Rate (DAR) are possible for proposed Cognitive radio-based structure [35].

Attackers try to maximize the impact of attack by inducing minimum effort. Therefore, a common target for the attackers is to construct the attack in some critical nodes in such a way that the interconnection nodes of the grid also get tripped. This phenomenon is an example of cascading failure. A Recovery-based Model Predictive Control scheme (RMPC) has proposed in [36] which addressed the cascading failure due to cyber-attack. The post-attack operating conditions and

measurements are used to formulate an optimization algorithm over finite time horizon. The control center decides the necessary action against this attack by solving the optimization algorithm. During attack, the control center calculates the state-k variable with the help of previous state variable information. The length of previous state variable dataset depends on the length of compromised dataset. The table 2 describes the dataset length for RMPC algorithm.

Table 2 Dataset length for RMPC algorithm

State information	Dataset length
Current state	K
Number of consecutive compromised dataset	N
Historical dataset to calculate k-state	K-N

RMPC algorithm works mainly for data recovery side, rather than attack detection. The compromised data are replaced by estimated dataset using the optimization technique.

For the case of successive attacks, a detection method was proposed in [37] where the matrix decomposition is utilized. The measurement matrix comprising the attack can be decomposed into a low-rank matrix plus a transformed column-sparse matrix [38]. The measurement matrix after the attack can be represented as;

$$L = \bar{Z} + \bar{A} + \epsilon$$

Z is the measurement matrix without attack, A is the attack matrix, and ϵ is a white noise in the measurement. For the Jacobian matrix H which expresses the non-linear relation between the state variables X and measurement variables Z (Z = HX). From the previous discussion it is evident that the attack vector A is equal to the $\bar{c}H$. If the matrix L can be decomposed into the low-rank approximation of Z and the column sparse matrix $\bar{c}$, we can separate the attack estimation vector $\bar{c}$ [37].

The countermeasures against cyber-attack mentioned in this subsection are detection-based. So, the control center waits for the attack to happen, after a while the attack gets detected using state-estimation. Much of the research works have been concentrated on making the state estimation faster and more robust. However, there is an underlying problem in this approach, which is the defender can't prevent the attack from happening. However, there are several works that considered the prevention against cyber-attacks by different techniques.

The most common prevention technique is the optimal placement of PMUs in the critical grid locations to make the whole grid observable. Due to cost and physical constraints, it is not feasible to install PMU at every node and branch of the grid. Therefore, the PMUs are installed at the locations where the observability and security of the system is maximized. In [31], a greedy-based optimal placement algorithm was proposed. The greedy algorithm place PMU at all the locations one after another at each step and measure the system security after each step. The PMU will be installed at the location which provides the highest security. Assuming H^S is the part of the H matrix, representing the secured measurements. If a PMU is added, a new row will be added to H^S matrix. The new matrix must follow the $Hc = 0$ criterion to avoid detection. The modified criterion after adding a PMU can be expressed as eqn (15).

$$\begin{bmatrix} H^S \\ H^{PMU} \end{bmatrix} c \quad = 0 \tag{15}$$

For $H^{PMU}c = 0$, the attack will remain stealthy. So, adding a PMU will not help detecting the attack. In contrast, for $H^{PMU}c \neq 0$, the attack will no longer be stealthy and adding a PMU will make the control center able to detect the attack. It forces the attacker to find alternate attack vector since the existing attack vector is no longer feasible.

A least effort Reduced Row Echelon (RRE) model was proposed in [39] to create an attack by that is not detected by aforementioned $H^{PMU}c \neq 0$ method. The defender or the control center can utilize the same technique to place PMUs at the locations provided by RRE form. Another methodology that can be used for PMU placement is the pre-deployment based greedy-algorithm. This algorithm was proposed in [29], where the edge buses and the adjacent to edge buses are first considered for PMU placement. After the edge buses are covered, greedy-algorithm mentioned in [31] can be used. This hybrid placement algorithm reduces the numerical complexity of greedy-algorithm alone.

2.4 Countermeasures for GPS-Spoofing Attack

GPS-spoofing attack can be regarded as a form of timing attack, where a time shift is introduced to the data. Recently several researchers worked on GPS-spoofing attack detection and mitigation. Some works on the time-delay attack on conventional power grid can also be utilized for PMUs. For example, X. Lou et. Al. [40] proposed a risk assessment and mitigation algorithm for time-

delay attack against Automatic Generation Control (AGC). The stability-safety check criterion is exploited in this algorithm. At first, the proposed method assesses the stability of the system. If the system fails in the stability assessment, it can be assumed that the system is unsafe. As long as the stability check is failed, mitigation scheme is initiated. If the system is marked as stable after the stability assessment, the safety assessment is performed. If the system shows to be unsafe, the mitigation scheme is initiated again.

Classification of stability is an important factor for stability-safety assessment. A boundary line-based stability safety method can be considered. System is stable when the operating condition is below the boundary line. The safety boundary-line is more complex for large number of load buses, since the safety depends on the load factors. To reduce complexity, Monte-Carlo-based method has shown to be useful to sample the operating points. Moreover, for safety classifications, a set of ground created. The ground truth dataset is divided into test and training dataset to apply ELM and the AGC's safety is marked with corresponding class for specific input operating points. For the mitigation scheme during unsafe condition, a two-tier technique used where the gain k is tuned with a PID controller and load-shedding is applied (Fig 9).

Since most PMUs use civilian GPS timing signals, spoofing of GPS signal is possible by attackers. Attackers manipulate GPS signal to create GPS-spoofing attack, causing loss of synchronization. A GPS-spoofing attack produces an offset in the phase angle data, similar to the one proposed by [10]. Detection of GPS-spoofing attack using cross-layer detection mechanism was proposed in [11]. In this work, a GPS carrier-to-noise ratio (CNR) is utilized for the detection of GPS spoofing attack. Two patch-monopole hybrid antennas are set at two GPS receivers, then the standard deviation of the power ratio is computed. The power ratio is defined by:

$$R_i = \left(\frac{C}{No}\right)_{i,1}(dB) - \left(\frac{C}{No}\right)_{i,2}(dB)$$

For normal condition, GPS-signals are being received from different satellites for different channels randomly, causing a difference between power ratio. However, for GPS-spoofing, signals for all the channels coming from same source, making the power ratio close to zero. In this case, the standard deviation of power ratio is less than a predetermined threshold [11].

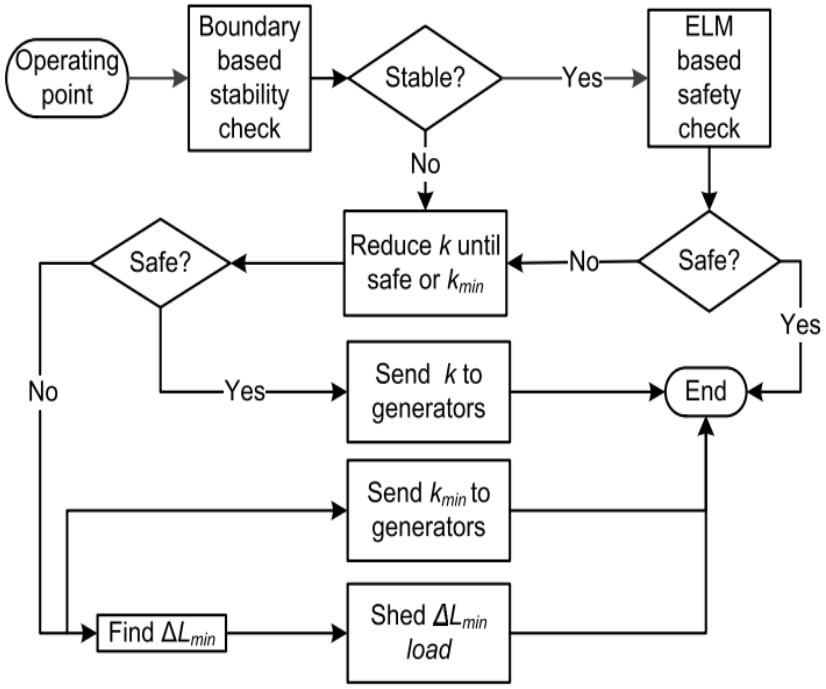


Figure 9 Two Tier Mitigation Technique for Timing Attack

A probing technique is implemented to identify the spoofed PMUs' locations [12]. Moreover, the amount of phase-shift due to the GPS-spoofing is also estimated by narrowing down the searching range of phase angle shift. The spoofed PMU data can be corrected by minimizing the following Weighted Measurement Residuals (WMR):

$$J_{corr} = r_{corr}^{T}\, W\, r_{corr}$$

Where,

$$r_{corr} = z_{corr} - H\, x_{corr}$$

$$z_{corr} = A^{-1}\, z_{spf}$$

$$A = Attack\ matrix$$

$$W = [diag(\sigma_1^2, \sigma_2^2, \ldots . \sigma_m^2)]^{-1}$$

A model for detection of time varying GPS-spoofing attack is proposed in [13]. The proposed method uses data from PMUs and SCADA. Dynamic fusion estimator is utilized to estimate power system states. Estimated rotor angles are compared with the measurements and the phase shift due to GPS-spoofing is calculated.

A light-weight threshold-based detection of GPS-spoofing attack is proposed by X. Wei et. al. [14]. The attacker will replay a previously captured signal to destroy the synchronization of PMU

time-stamp, which results in the existence of autocorrelation to some extent in the PMU signal. The correlation between signal is computed within a window. At time instance t, the window of length T can be expressed as $W_t = [P_0, P_1, \dots P_{T-1}]^T$. The correlation between the window W_t and W_{t-1} is calculated. The GPS-spoofed signal is detected by accepting or rejecting the following hypothesis test:

$$H_0 : Elements\ of\ W\ are\ not\ correlated$$

$$H_1 : Elements\ of\ W\ are\ correlated$$

In [15], a recovery of PMU data after GPS-spoofing attack is proposed where sparse optimization technique is utilized. An iterative greedy algorithm is used which exploits the network topology information. The attack vector is estimated by solving the optimization problem first. With the estimated attack vector, the falsified data is recovered as in Fig 2.

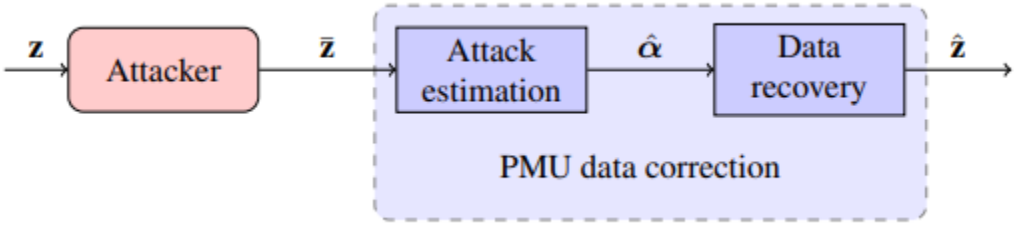


Figure 10 Data Correction for GPS-spoofing attack

A mathematical model for online pattern recognition and data correction of GPS-spoofing attack has been proposed in [16]. In this method the phase angle deviation between the both ends of the transmission line is estimated using moving window method. An advantage of this method is that it doesn't need transmission line parameters to estimate phase angle deviation.

2.5 Contribution of this work

From the previous discussion, it can be concluded that cyber-attack on synchrophasor data has become a widely explored area. However, GPS-spoofing and FDIA attacks require different responsive actions. State-estimation based attack detection method works similar way for both types of attacks. It implies that using state-estimation based detection alone doesn't give the control

center enough information about the attack type. This scenario is specifically more critical for those attack detection methods which consider GPS-spoofing attack as another FDIA with falsified phase angle data.

Since identifying correct attack type is paramount, we have attempted to develop an algorithm to distinguish these two attacks. Previous researchers exploited low-rank approximation of Hankel Matrix to differentiate between FDIA and physical events. We have demonstrated that, together with angle unwrapping algorithm, low-rank approximation of Hankel Matrix can help us separating GPS-spoofing attack with FDIA.

The next chapter discusses our proposed algorithm on GPS-spoofing attack detection and its ability to distinguish this type of attack from conventional FDIA.

Chapter 3

GPS-Spoofing Detection: Proposed Method

The goal of this work to propose an algorithm to detect GPS-spoofing attack and will make us able to distinguish GPS-spoofing attack from FDIA. Previously, low-rank approximation of Hankel matrix has been proved to be useful against anomaly detection of power grid data. In this work the combination of low-rank approximation of Hankel matrix and unwrapped phase angle data are considered to detect GPS-spoofing attack. In the following subsection, our proposed model and the corresponding background is discussed.

3.1 Hankel Matrix and Low Rank Approximation

Interconnected PMUs in the power grid topology shows similar behavior as their neighboring PMUs. Since PMUs provide time-synchronized measurements, a large number of synchrophasor dataset from neighboring PMUs show coherent characteristics [41], [42], [43]. Each row of the measurement matrix of multiple PMU channels contains time-series data of single PMU channel, where each column represents PMU data from all the channels for each time instant. As neighboring PMUs display a correlation in time-series, matrix contains a low number of significant singular values. Thus, PMU dataset can be considered as approximately low rank. The low-rank property of the PMU data is useful for several applications discussed in previous literatures such as missing data recovery [44], event and cyber-attack detection [37] [45].
The PMU dataset can be expressed as the matrix Y, which is as follows:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}$$

Y is a $m \times n$ matrix, m is the number of PMU channels and n is the number of time instances over a single time-window. The first step of taking low rank approximation of a matrix is decompose the matrix using Singular Value Decomposition (SVD). SVD of Y can be expressed as $Y = U \sum V^*$. SVD of a matrix is useful for approximating the matrix Y as a rank r where $r < rank(Y)$. This approximation can be done by taking the largest r singular values from Σ. With these r number of singular values, another diagonal matrix Σ^r can be formed. $U\ \Sigma^r V$ represents the low-rank form of original matrix Y. The low-rank approximation error can be formulated as:

$$e^r = \frac{\left|\left|U\ \Sigma^r\ V^* - Y\right|\right|_F}{\left|\left|Y\right|\right|_F} \times 100\ \% \qquad (3.1)$$

Electrical events such as line outages, oscillation events, transformer events etc disrupt the power grid normal operating conditions. During these events, PMU data exhibit anomalous behavior. Conventional Bad Data Detection (BDD) based anomaly detection is generally used for electrical event detection. Both electrical events and cyber-attack cause erroneous measurements, as a result conventional BDD scheme fails to distinguish between electrical event and cyber-attack.

An important difference between the erroneous measurements due to cyber-attack vs the erroneous measurements due to event is the time-series correlation. An electrical event is a physical incident and it affects the physical interconnection in the grid. So, it is expected that during an electrical event the neighboring bus-measurements should also be affected. This scenario implies that the neighboring PMUs during electrical event must exhibit a temporal relation in their anomalous measurements [46].

On the other hand, an FDIA is a targeted intrusion by the hackers. The attacker injects falsified measurements by interfering the communication channel between different layers. Instigating FDIA into one node of the grid doesn't impact the neighboring nodes. PMU dataset doesn't show any temporal correlation during attack on specific PMU data. This opposite trait can be utilized to differentiate FDIA and event.

Ref [37] uses the low-rank approximation of the Hankel matrix distinguish FDIA and electrical event, exploiting the difference between the temporal behavior of Y matrix during these two incidents. A Hankel matrix is a square matrix constructed by making each ascending skew-diagonal elements constant. This skew-diagonal ascent goes from left to right. Each row is formed by right-shifting previous row. Hankel matrix is particularly useful for time-series analysis as well as for state-space representations of data. Considering a dataset of window length k, and the dataset is B = $\{b_0, b, \ldots\ldots b_k\}$, then the Hankel matrix can be expressed as the following:

$$H_m = \begin{bmatrix} b_0 & b_1 & \cdots & b_{\frac{k}{2}+1} \\ b_1 & b_2 & \ldots & b_{\frac{k}{2}+2} \\ \vdots & \ldots & \ddots & \vdots \\ b_{\frac{k}{2}+1} & b_{\frac{k}{2}+2} & \cdots & b_k \end{bmatrix} \tag{3.2}$$

H_m is a $\left(\frac{k}{2}+2\right) \times \left(\frac{k}{2}+2\right)$ square matrix. Columns of the matrix in eqn (3.2) together represent the temporal variation and rows together represent the spatial variation of dataset. In [47], SVD of H_m has been used to decompose the signal into temporal and spatial variation. Since low rank approximation of a matrix is deducted directly from SVD, the low rank approximation of Hankel Matrix can be used to analyze time-series PMU data.

As discussed above, during an electrical event, a temporal correlation can be found among the data of all the channels. This temporal correlation will no longer exist if a random column permutation is performed. Without temporal correlation, the number of significant singular values is higher, giving larger low rank approximation error [48].

For FDIA, PMU data of only the affected node will be modified. The neighboring PMUs don't show any change in dataset. While there is no temporal relation among the PMU channels, a random permutation will not have any effect on the low-rank approximation of Hankel Matrix [48]. The complete model can be visualized in algorithm 1.

Algorithm 1: Distinguishing FDIA from electrical events in PMU

Initialization Receive time-series measurements from PMU as the $\boldsymbol{m} \times \boldsymbol{n}$ matrix Y; m = channel numbers, n = data window length

Step 1: Create a $\boldsymbol{m}\left(\frac{\boldsymbol{n}}{\boldsymbol{2}}+\boldsymbol{2}\right) \times \left(\frac{\boldsymbol{n}}{\boldsymbol{2}}+\boldsymbol{2}\right)$ Hankel matrix $\boldsymbol{H_m}$;

Step 2: Calculate low rank approximation error $\boldsymbol{e^r}$ with varying e ($\boldsymbol{r \leq rank(H)}$);

Step 3: Perform random column permutation on the Hankel matrix $\boldsymbol{H_m}$ and create new matrix $\overline{\boldsymbol{H_m}}$;

Step 4: Calculate low rank approximation error $\boldsymbol{e^{rr}}$ with varying rank r $\left(\boldsymbol{r \leq rank(\overline{H_m})}\right)$;

Step 5: If $\boldsymbol{e^{rr} > e^r}$, it is an electrical event;

Step 6: else, it is a FDIA;

3.2 Angle Unwrapping Technique

Generally, power grid frequency is kept nominal at either 50 Hz (Europe) or 60 Hz (USA). In practical scenario, the frequency often fluctuates from this nominal values. These fluctuations result in phase angle data being deviate suddenly. According to IEEE C37.118.1-2011 synchrophasor standard [10], the phase angle must be kept between +π to −π (radian). To meet this requirement, the phase angle data are wrapped around by 2π radian, causing a transition between $+/-\pi$ to $-/+\pi$.

While performing data analysis, discontinuities of phase angle data above π and below π causes complexity, especially when the target is the recovery of missing data. To address this complexity, unwrapping of phase angle data can be an important tool. Previous phase angle unwrapping techniques available in MATLAB and Python were off-line, with a goal to use unwrapped data for post-processing [49]. To unwrap the phase angle in real-time for continuous processing, an efficient algorithm was proposed in [49]. Despite its efficiency, unwrapped phase angle with this algorithm may grow very large over time [50]. To tackle the problem of unwrapped angle growing too large, a Roll-Over Counter (ROC) can be considered for unwrapping phase angle [50]. An ROC represents the number of times the phase angle shifts from $+/-\pi$ to $-/+\pi$. This counter then gives the unwrapped angle with the following formula:

$$\min_{N} |\theta_{i+1} - \theta_i + 360\, N| \quad (3.3)$$

$$ROC(i+1) = ROC(i) + N \quad (3.4)$$

In eqn (3.3), θ_{i+1} and θ_i represents two consecutive phase angle data. The solution of the minimizer in eqn (3.3) is the integer that is added to current ROC (eqn 3.4) value whenever there is a transition between +/- π to -/+ π. The minimizer (3.3) indicates that N = 1 when the phase angle changes from +π to −π and N = -1 when phase angle changes from −π to +π. Fig. 11 depicts the raw phase angle data for a sample simulated PMU in MATLAB and fig. 12 represents the corresponding unwrapped phase angle data using equation (3.3).

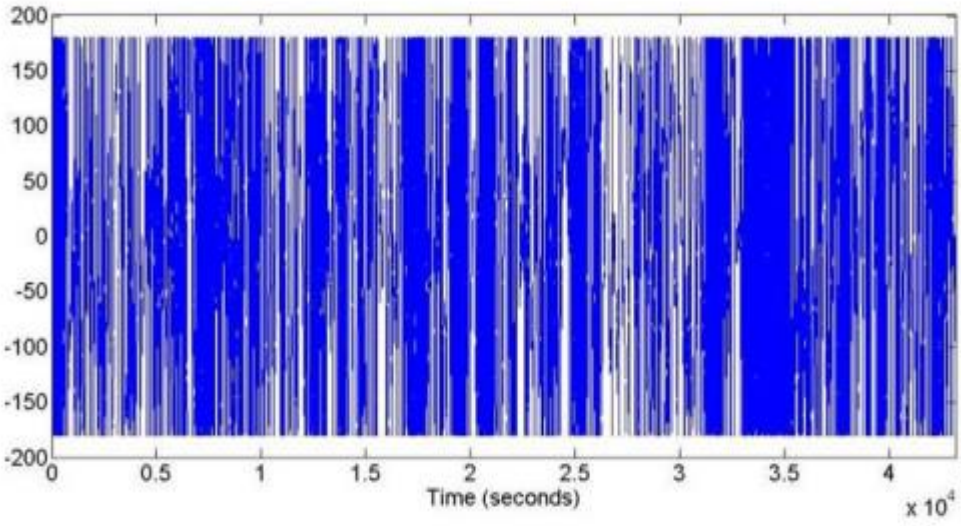


Figure 11 Raw Phase angle data from a PMU

The Unwrapped phase angle data demonstrate some unique behaviors and are different from raw phase angle data. This technique is especially useful for islanding detection and phase angle difference calculation.

The ROC the gain in cycles of the signal with respect to the UTC in the GPS signal. ROC value indicates that a clock running on the signal at system frequency F has a gain ROC/F seconds compared to the UTC clock since the beginning the calculation. The ROC may can grow to large values for the frequency being off nominal for a long period.

At the later part of this chapter, we show that the unwrapped phase angle values can be useful for analyzing cyber-attack into PMUs, particularly for GPS-Spoofing attack detection.

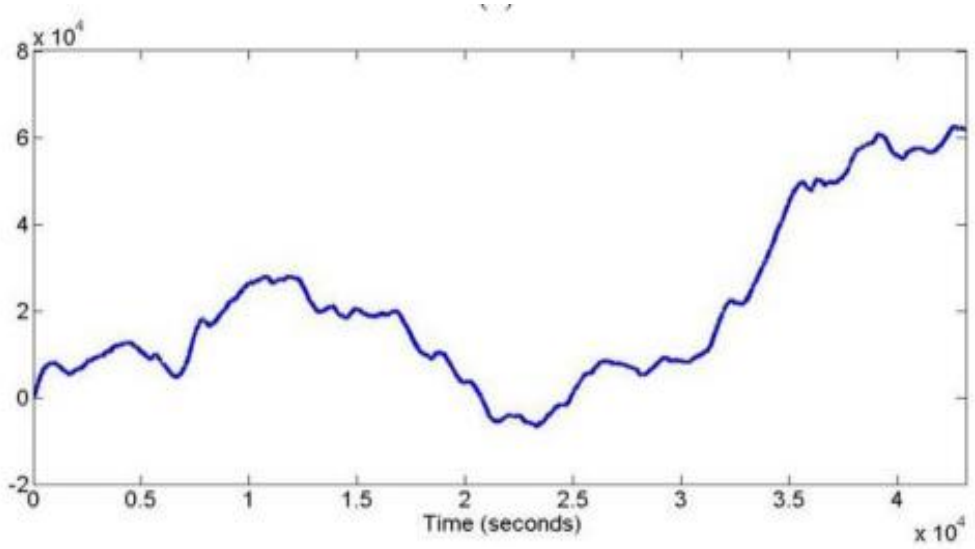


Figure 12 Unwrapped phase angle data with eqn (3.3)

3.3 Timing Attack Detection with Unwrapped Angle

When phase-angle data gets wrapped-up from +/- π -/+ π, unwrapping technique helps us avoiding this transition by 2π. According to the chapter 2, attackers may inject falsified data into the communication channel and thereby make the PDC or the control center take decision based on the wrong measurements. Control center receives the data, perform state estimation and error detection processes. If the data that the control center uses are corrupt, the unit commitment and economic dispatch calculation made by the control center may become wrong and the stable and safe operation of the grid may be hampered. For the case of False Data Injection Attack (FDIA), the attacker modifies the measurement value directly thought the communication channel. The PMU provides phasor data, and the dataset contains the phase angle information of each phase and the positive sequence components of both voltage and current. For FDIA, attacker can target the phase angle values only, which is sent as 4- or 8-byte binary data of the *PHASOR* field in the data frame configuration of IEEE C37.118.2-2011 protocol [12]. For the GPS-spoofing attack (GSA), the attacker mimics the real-GPS signal with a fake one, the timestamp that the PDC or control

center receives are shifted in time-axis. In this case, The *SOC* and/or the *FRACSEC* information in the data frame is wrong.

In both attacks, the phase angle data is incorrect, therefore conventional state-estimation based BDD methods can't distinguish these two types of attack. Nevertheless, it is imperative for the control center to correctly identify the attack type to take accurate restoration technique. Thought, just looking at the phase angle data will not provide enough information to distinguish FDIA and GSA, unwrapped phase angle data can provide some insight into it. During FDIA, for the time instance t, the attacker changes the phase angle in eqn (3.3), changing θ(t) to $\theta'(t)$ by adding an attack value a(t). During the next time instance t+1, attacker modifies the angle data by adding attack value a(t+1). Assuming original θ(t) is ∼ +180 and θ (t+ 1) is ∼ −180, the modified phase angle data received by control center will be:

$$\theta'(t) = \theta(t) + a(t)$$
$$\theta'(t+1) = \theta(t+1) + a(t+1) \tag{3.5}$$

From the eqn (3.5), it is evident that adding a falsified attack value changes the phase angle measurements. However, adding an attack value doesn't change the instance when the angular data transits from +/- π to $-/+\pi$. Also, the goal of the attacker is to create a stealthy attack so that it goes undetected by conventional χ^2 based BDD algorithms. This condition makes attacker to construct the $a(t)$ value very small to avoid detection. Considering the small $a(t)$, it can be safely assumed that no additional transition between $+/-\pi$ to $-/+\pi$ will occur due to FDIA. From eqn (3.3), since the transition instance between +/- π to $-/+\pi$ is not changed, the value of N will not change either. The ROC from eqn (3.4) will remain same. As the unwrapped phase angle value depends on N value and the ROC, the attack from eqn (3.5) will not disrupt the shape of unwrapped phase angle graph. FDIA will cause a small shift in the magnitude, keeping the shape of the curve like unwrapped phase angle graph under no-attack condition.

During GSA, the timestamp will be shifted, which implies that the phase angle data will be shifted toward the horizontal time axis. For GSA, the transition instance between +/- π to $-/+\pi$ will be shifted in the horizontal axis. If the attacker introduces a time delay T to the measurements, the new phase angle measurement will become as follows,

$$\theta''(t) = \theta(t + T) \tag{3.6}$$

Control center receives the phase angle value from PMU at t^{th} time instance, which is the actually data from $(t + T)^{th}$ time instance. Contrary to the FDIA, GSA causes change the transition point between +/- π to $-/+\pi$. ROC from eqn (3.4) changes when there is a transition between +/- π to $-/+\pi$ and the N value will be either +1 or -1, depending on the direction of the transition. As the ROC value will be changed at the moment of GSA, the unwrapped phase angle graph should also change at that moment. This incident will create a certain unexpected spike (positive or negative direction) in the unwrapped phase angle graph.

Looking at the raw phase angle data don't give us enough information about the attack type. However, looking at the unwrapped phase angle graph can give us important information about the attack type. FDIA doesn't create any significant distortion in the unwrapped phase angle curve. For GSA, the unwrapped phase angle curve gets distorted with a spike now of the attack. A sudden spike in the unwrapped phase angle data indicates a delay in the time-axis, suggesting a shift in the timestamp. In short, we can differentiate the GSA from FDIA by exploiting the unwrapped phase angle behavior.

3.4 Detecting Unwrapped Phase Angle Graph Distortion with Hankel Matrix

As per the discussion in section 3.1, a bad data in the measurement stream can either due to cyber-attack or due to electrical event. There exists a temporal relation among the PMU channels during event and there is no temporal relation among the channels during cyber-attack. The existence of temporal relation is reflected by the random permutation of column matrix of low-rank approximation of Hankel matrix.

This method uses magnitudes or raw phase angle data from PMU channels. Raw phase angle data alone is not enough to identify the GSA or timing attack. Section 3.3 suggests that unwrapping of phase angle data can be a useful tool in identifying GSA. To identify the unwrapped angle curve distortion, the low rank approximation of Hankel matrix can again be utilized. In this case, the temporal relation among the PMU channels is not a factor. The low rank approximation of Hankel matrix of individual channel needs to be analyzed separately to detect the attack type. The key idea

is that since the FDIA doesn't change the shape of unwrapped phase angle curve, the low rank approximation error of Hankel matrix over consecutive time-window shouldn't change significantly from that of normal condition. Conversely, during GSA, the affected PMU's unwrapped phase angle curve gets distorted. As a result, the low rank approximation error of Hankel matrix changes significantly at the instance of GSA occurrence. For FDIA, the phase angle data is added by the attack value a, whereas for GSA, the corresponding data is the phase angle data from T sec. later plus the additional ROC due to the transition between +/- π to $-/+\pi$. The change in Hankel matrix due to FDIA and GSA are depicted in fig 13a, 13b and 13c.

$$Y = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_{\frac{n}{2}+1} \\ \theta_1 & \theta_2 & \dots & \theta_{\frac{n}{2}+2} \\ \vdots & \dots & \ddots & \vdots \\ \theta_{\frac{n}{2}+1} & \theta_{\frac{n}{2}+2} & \cdots & y_{n-1} \end{bmatrix}$$

(a)

FDIA at time t

$$Y = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_{\frac{n}{2}+1} \\ \theta_1 & \theta_2 & \dots & \theta_{\frac{n}{2}+2} \\ . & \theta_t + a & \dots & \theta_{\frac{n}{2}+t} + a \\ \vdots & \dots & \ddots & \vdots \\ \theta_{\frac{n}{2}+1} + a & \theta_{\frac{n}{2}+2} + a & \cdots & y_{n-1} + a \end{bmatrix}$$

(b)

GSA at time t

$$Y = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_{\frac{n}{2}+1} \\ \theta_1 & \theta_2 & \dots & \theta_{\frac{n}{2}+2} \\ . & \theta_{t+T} + ROC & \dots & \theta_{\frac{n}{2}+t+T} + ROC \\ \vdots & \dots & \ddots & \vdots \\ \theta_{\frac{n}{2}+1+T} + ROC & \theta_{\frac{n}{2}+2+T} + ROC & \cdots & y_{n-1+T} + ROC \end{bmatrix}$$

(c)

Figure 13 Hankel matrix for a) normal condition b) FDIA and c) GSA

The low rank approximation error is computed using eqn (3.1). The attack detection process is performed over moving time window. If the attack is detected using algorithm 1 for a specific time window, and the corresponding error sharply increases with unwrapped phase angle-based

Hankel matrix, then it can be concluded that the attack is GSA. The algorithm 2 describes the procedure to detect GSA and to distinguish GSA from FDIA.

Algorithm 2: Detection of GSA and distinguishing from FDIA

Initialization Receive time-series angular (voltage or current) measurements from PMU as the matrix Y of time instances $\boldsymbol{t_0, t_1, .., t_{n-1}}$, m is the number of channels; n is the data length of each window.

Step 1: Create a m (n/2 + 2) × (n/2 + 2) Hankel matrix H;

Step 2: Calculate the low rank approximation error e r with varying rank r (r ≤ rank(H));

Step 3: Do a random column permutation on the Hankel matrix H and create a new matrix $\bar{\boldsymbol{H}}$;

Step 4: Calculate the low rank approximation error $\boldsymbol{e^{rr}}$ with varying rank r (r ≤ rank($\bar{\boldsymbol{H}}$));

Step 5: If $\boldsymbol{e^{rr} > e^{r}}$, it is an electrical event. Go to step 1 with the next n data sample;

Step 6: else, it is an attack. Continue to Step 7;

Step 7: Unwrap the angular data and create Hankel matrix $\boldsymbol{H_u}$ with data from each individual channel, matrix dimension: (n/2 + 2) × (n/2 + 2);

Step 8: Calculate the low rank approximation error $\boldsymbol{e^{ru}}$ with varying rank r (r ≤ rank($\boldsymbol{H_u}$));

Step 9: if the low rank approximation error $\boldsymbol{e^{ru} > e^{r}}$, then it is a timing attack;

Step 10: Else, it is FDIA

Step 11: Proceed with next time window

Chapter 4

Results

In this chapter, the numerical simulation setup and the results are summarized. At the end of this chapter, the overall contribution of this thesis is discussed.

4.1 Simulation Setup

We have verified the proposed methodology in IEEE 13 bus system. To simulate the timing attack, we have created an experimental setup, where the GPS timestamp is simulated using MATLAB datetime function. It provides a series of universal time reference (UTC) timestamp beginning from a specified point of date and time. In this work the timestamp starts at the midnight (00:00:00) of June 1, 2021 with a sampling rate of 1/60 sec. The first data is at 00:00:00, the second data is at 00:00:0.0167, and so on. The last data of a single day is at 23:59:59. The simulation of IEEE 13 node test system has been done with MATLAB SIMULINK. The PMUs have been added at buses 632, 633, 634, 671, 672 and 692 (fig 2). The PMUs provide positive sequence voltage and current magnitude and angle as well as frequency data.

At the first part we have created the experimental setup with numerical simulation. The SIMULINK model of PMU is used to receive synchrophasor data. This data is synchronized with the timestamp from MATLAB *daytime* function. The data is fed into python to implement algorithm 2. The python code calculates the low rank approximation error for the rank 1 to rank 10 and returns the average approximation error for each time window.

The simulated GPS timestamp produced with the functional block of fig. 15 is synchronized with the positive sequence voltage angle data. To create a GSA, we have shifted the GPS timestamp

by a time T. In this way the voltage angle values loss synchronization and the control center will receive with corrupted and shifted timestamp. In this way we can produce a resemblance of the GSA on PMU.

The total simulation is performed over a time span of 100 sec. For the FDIA, the attack value a is injected as a random variable, having value within the range of -2^o to 2^o. Both FDIA and the GSA is initiated at 58.333 second (timestamp 00:00:58.333).

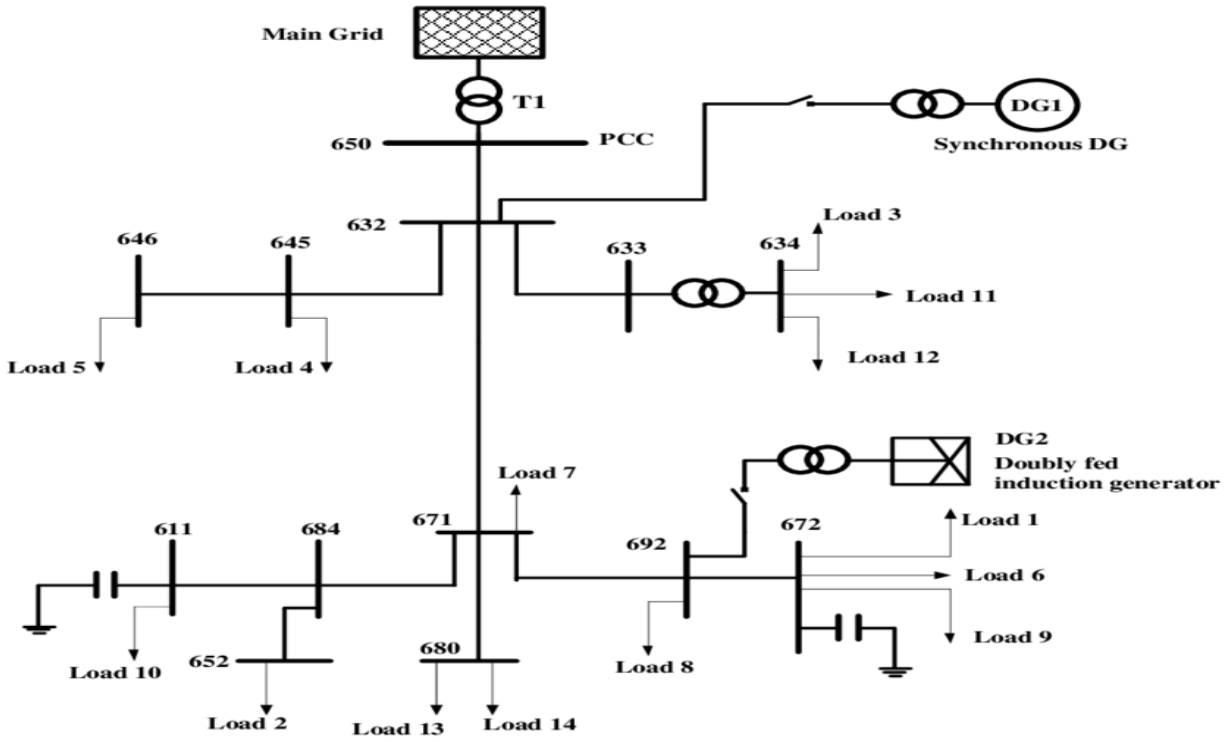


Figure 14 IEEE 13 Node Test System

Fig 16a depicts the FDIA and fig 16b depicts the GSA or timing attack. The unwrapping technique described in section 3 is applied for all the three cases: normal condition, FDIA and timing attack (GSA). From fig 17, it can be observed that at the moment of attack, the unwrapped angle data starts deviated for GSA. However, for FDIA, the graph is not distorted much.

To differentiate FDIA from GSA, algorithm 2 is implemented. The Hankel matrix is created with the datapoints of each time window. Time window consists of 3 sec of data (180 samples). Therefore, the length n from algorithm 2 is 180. The Hankel matrix is a 92×92 matrix. The first time window is 00:00:00 to 00:00:03, the second time window is 00:00:00.0167 to 00:00:03.0167, and so on. The last time window is 00:01:37 to 00:01:40.

4.2 Result at the Moment of Attack

During FDIA and GSA, the low rank approximation error from eqn (3.1) is calculated over the rank 1 to 10. Fig 18 displays the low rank approximation error of FDIA and GSA. From the figure, it is evident that the error is higher for GSA. Rank 1 approximation error increases by 700% for a GSA which creates 3 sec time shift.

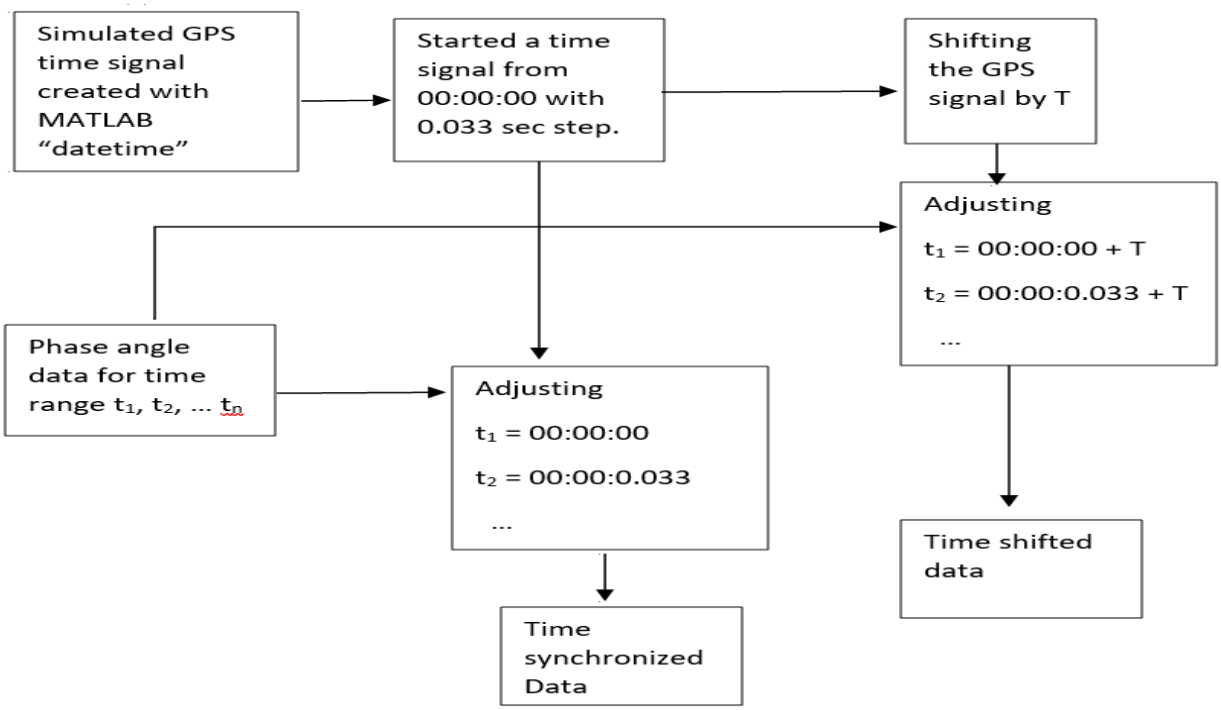


Figure 15 Block Diagram of Simulation setup to analyze timing attack

Fig 19 depicts the low rank approximation error for different GSA. For 3 sec time shift, the error is highest, especially for rank r = 1. For rank = 5, the error is not significant. The higher the time-shift due to GSA, the higher the error is. For time shift of 1 sec, the low rank approximation error is insignificant.

4.3 Result for moving time window

The average low rank approximation error is observed over the moving time window. From fig 20, it can be concluded that for GSA, the average low rank approximation error is higher than

that of normal condition. However, for FDIA, the low rank approximation error doesn't demonstrate any significant change.

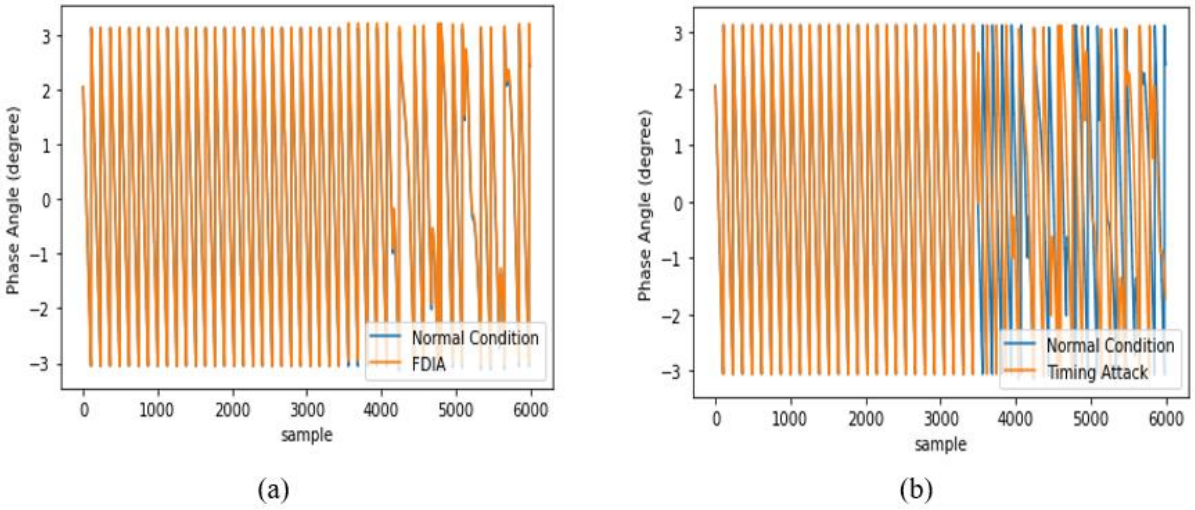


Figure 16 a) FDIA at 58.33 sec (samle no 3500) b) Timing attack/ GSA at the same time

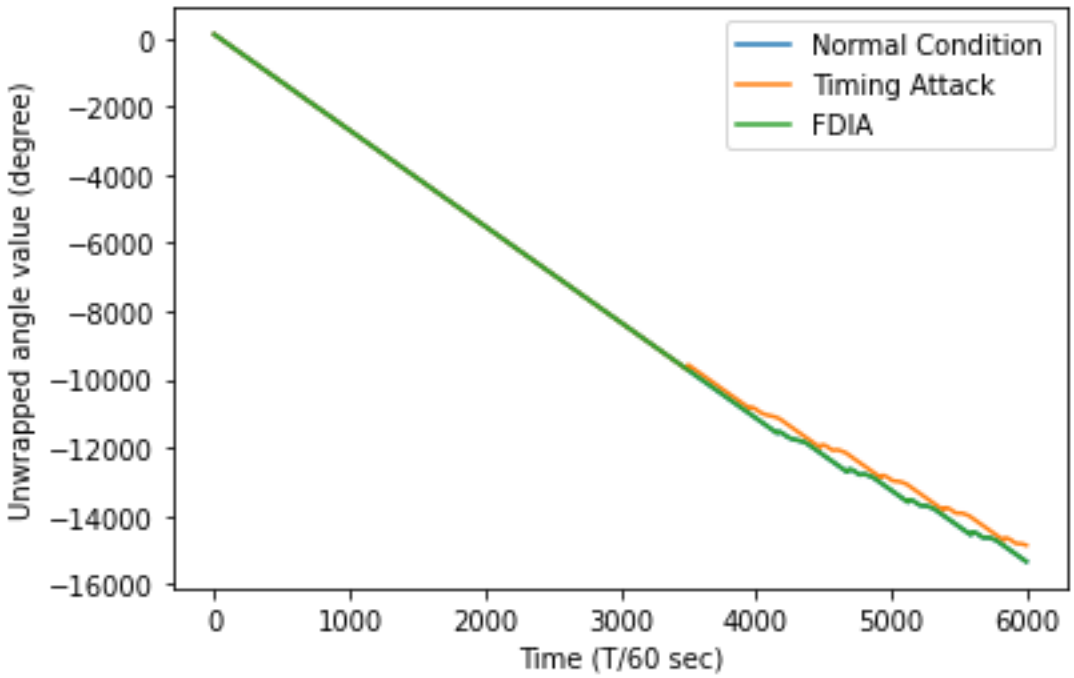


Figure 17 Unwrapped phase angle data

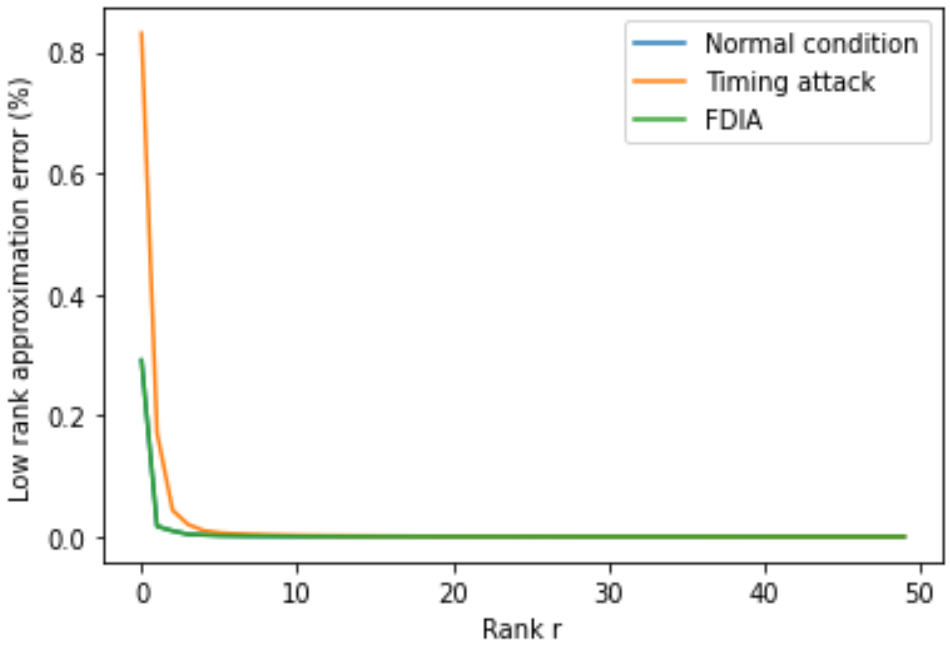


Figure 18 Low rank approximation error under normal condition, FDIA and GSA

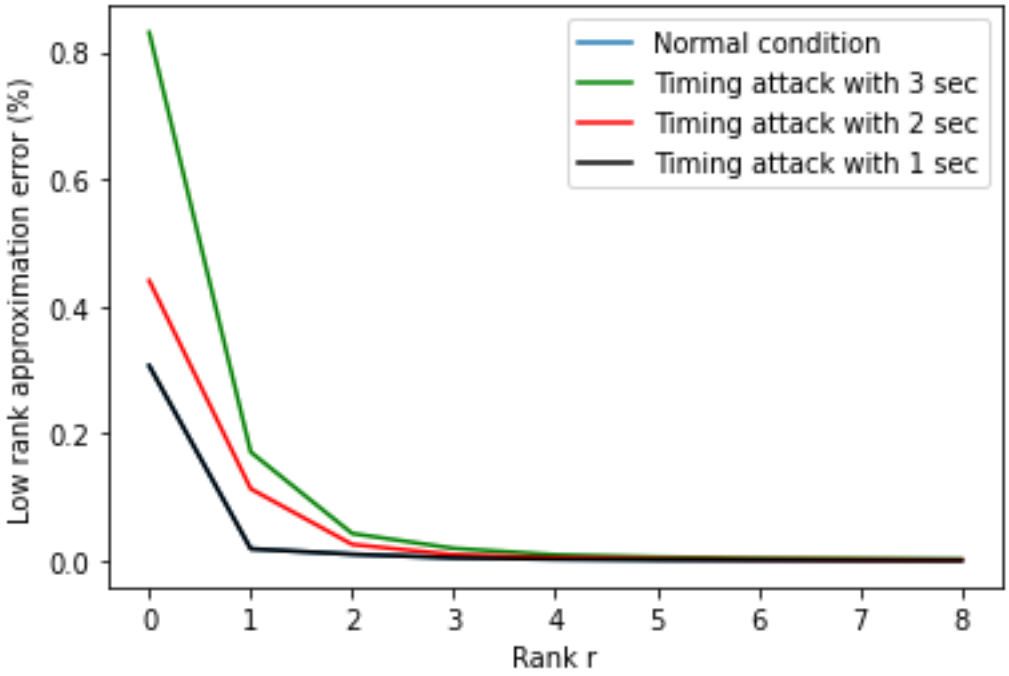


Figure 19 Low rank approximation error under GSA with varying time-shift

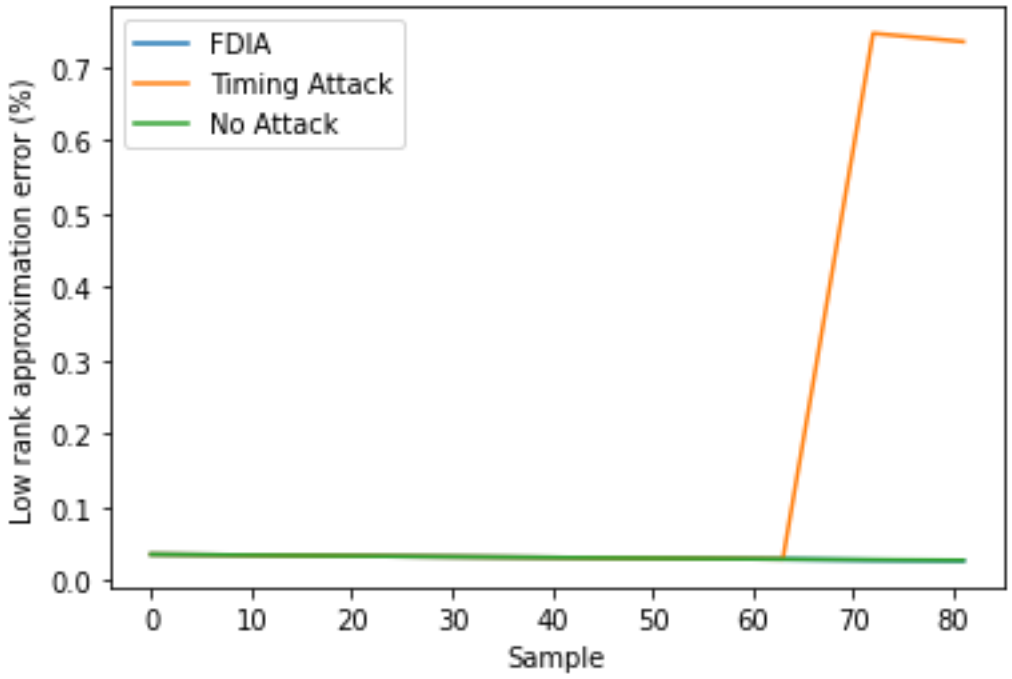


Figure 20 Low rank approximation error over moving time window

4.4 Contribution

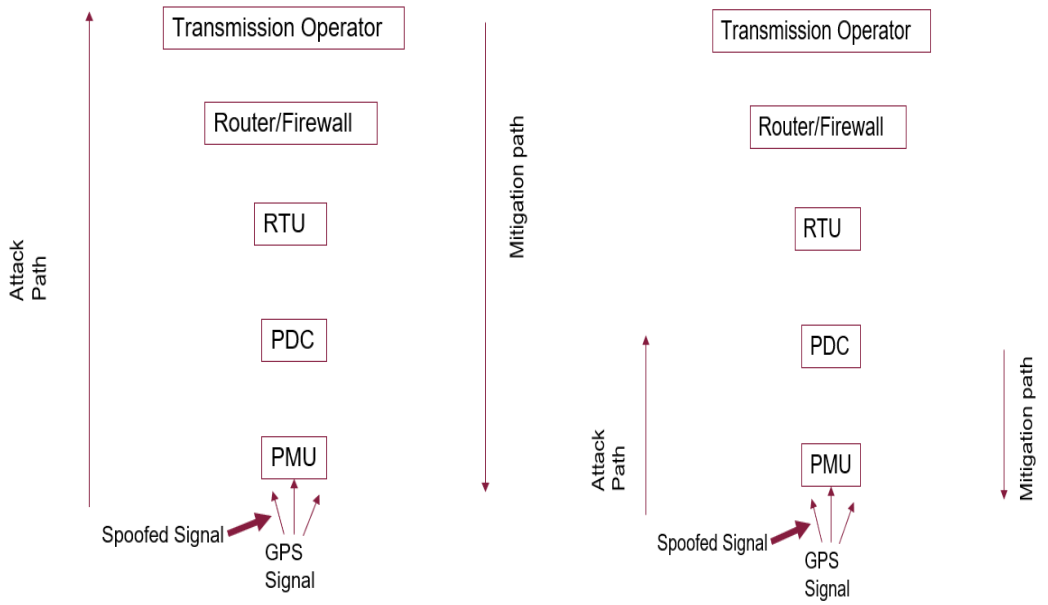
Contribution of this work can be summarized as follows:

- Provided a thorough literature review on the cyber-attack on PMU
- Discussed the Hankel matrix-based attack and electrical event identification technique
- Proposed a method for differentiating False Data Injection Attack from GPS-spoofing attack using unwrapped phase angle data and low rank approximation of Hankel matrix
- Developed a real-time model for successfully identifying electrical event, FDIA and GPS-spoofing attack.
- Verified proposed method with numerical simulation

4.5 Limitations

One limitation of this work is that the proposed method, like other state-estimation based detection method, relies on the data received to the control center. Therefore, in this scenario, the attack is already happened, and the control center mitigates the impact of attack after the occurrence of the attack.

One possible way to utilize the proposed method in preventing attack from propagating to control center, is applying this algorithm in the PDC level, since PDC has the options to incorporate error detection algorithm. However, this will require the higher numerical power for PDCs. Fig 42 depicts the attack and mitigation path for PDC incorporated with attack detection algorithms. This scheme also reduces the communication delay while taking mitigation action



(a) (b)

Figure 21 Attack and mitigation path for a) conventional structure and b) PDC incorporated with detection algorithm

Chapter 5

Conclusion and Future Research Scope

5.1 Conclusion

Cyber-attack on synchrophasor data has become a widely explored area. However, GPS-spoofing and FDIA attacks require different responsive actions. State-estimation based attack detection method works similar way for both types of attacks. It implies that using state-estimation based detection alone doesn't give the control center enough information about the attack type. This scenario is specifically more critical for those attack detection methods which consider GPS-spoofing attack as another FDIA with falsified phase angle data. Since identifying correct attack type is paramount, we have attempted to develop an algorithm to distinguish these two attacks. Previous researchers exploited low-rank approximation of Hankel Matrix to differentiate between FDIA and physical events. We have demonstrated that, together with angle unwrapping algorithm, low-rank approximation of Hankel Matrix can help us separating GPS-spoofing attack with FDIA.

The simulation result verifies our proposed algorithm on GPS-spoofing attack detection and its ability to distinguish this type of attack from conventional FDIA. It has been demonstrated that the GSA with 3 second time-shift creates an low-rank approximation error 700% higher than that of normal condition, whereas FDIA doesn't produce any significant change in low-rank approximation error from that of normal condition. Finally, we have proposed a real-time method for successful identification of event, FDIA and GSA.

5.2 Future Research Scope

There is lot of scope for future research such as:

- The proposed real time attack identification can be verified with experimental setup
- Undetectable GSA algorithm can be developed to avoid detection by the control center, we are currently working on it.
- Role of Phasor Data Concentrator (PDC) from both the attackers' and defenders' perspective need to be analyzed
- Vulnerabilities of TCP/UDP based communication using IEEE C37.118.2-2011 protocol need to be investigated

References:

[1] https://www.govinfo.gov/content/pkg/PLAW-110publ140/html/PLAW-110publ140.htm

[2] N. Anku, J. Abayatcye and S. Oguah, "Smart grid: An assessment of opportunities and challenges in its deployment in the ghana power system," *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, 2013, pp. 1-5, doi: 10.1109/ISGT.2013.6497800.

[3] K. Moslehi and R. Kumar, "A Reliability Perspective of the Smart Grid," in IEEE Transactions on Smart Grid, vol. 1, no. 1, pp. 57-64, June 2010, doi: 10.1109/TSG.2010.2046346.

[4] Liu C-C, Stefanov A, Hong J, Panciatici P. Intruders in the grid. IEEE Power Energy Magazine 2012; 10(1):58–66

[5] Queiroz C, Mahmood A, Tari Z. SCADASim – a framework for building SCADA simulations. IEEE Transactions on Smart Grid 2011; 2(4):589–597.

[6] Stefanov, A., Liu, C., Govindarasu, M. and Wu, S. (2015), SCADA modeling for performance and vulnerability assessment of integrated cyber–physical systems. *Int. Trans. Electr. Energ. Syst.*, 25: 498– 519. doi: 10.1002/etep.1862.

[7] Savulescu SC. Real-Time Stability Assessment in Modern Power System Control Centers, Wiley, IEEE Press on Power Engineering, 2009

[8] Mohanty, Monalisa & Kant, Ravi & Kumar, Asit & Sahu, Debasis & Choudhury, Subhashree. (2020). A Brief Review on Synchro Phasor Technology and Phasor Measurement Unit. 10.1007/978-981-15-5262-5_53.

[9] Phadke, A.G., Thorp, J.S.: Synchronized phasor measurements and their applications, vol. 1. Springer, New York (2008)

[10] "IEEE Standard for Synchrophasor Measurements for Power Systems," in IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005) , vol., no., pp.1-61, 28 Dec. 2011, doi: 10.1109/IEEESTD.2011.6111219.

[11] Aminifar, Farrokh & Shouraki, Saeed & Fotuhi-Firuzabad, Mahmoud & Shahidehpour, M.. (2010). Reliability Modeling of PMUs Using Fuzzy Sets. Power Delivery, IEEE Transactions on. 25. 2384 - 2391. 10.1109/TPWRD.2010.2051821.

[12] "IEEE Standard for Synchrophasor Data Transfer for Power Systems," in IEEE Std C37.118.2-2011 (Revision of IEEE Std C37.118-2005) , vol., no., pp.1-53, 28 Dec. 2011, doi: 10.1109/IEEESTD.2011.6111222.

[13] C. Tu, X. He, Z. Shuai, and F. Jiang, “Big data issues in smart grid – A review,” Renewable and Sustainable Energy Reviews, vol. 79, pp. 1099–1107, 11 2017.

[14] W. Li, L. Vanfretti, and J. H. Chow, “Pseudo-Dynamic Network Modeling for PMU-Based State Estimation of Hybrid AC/DC Grids,” IEEE Access, vol. 6, pp. 4006–4016, 11 2017.

[15] Z. Shuai, W. Huang, C. Shen, J. Ge, and Z. J. Shen, “Characteristics and Restraining Method of Fast Transient Inrush Fault Currents in Synchronverters,” IEEE Transactions on Industrial Electronics, vol. 64, pp. 7487–7497, 9 2017.

[16] M. N. Aman, K. Javed, B. Sikdar, and K. C. Chua, “Detecting data tampering attacks in synchrophasor networks using time hopping,” IEEE PES Innovative Smart Grid Technologies Conference Europe, 7 2016.

[17] A. Chawla, A. Singh, P. Agrawal, B. K. Panigrahi, B. R. Bhalja, and K. Paul, “Denial-of-Service Attacks Pre-Emptive and Detection Framework for Synchrophasor Based Wide Area Protection Applications,” IEEE Systems Journal, pp. 1–12, 2021.

[18] J. Xie, A. P. Meliopoulos, and G. Cokkinides, “PMU-based line differential protection under GPS spoofing attack,” Proceedings of the Annual Hawaii International Conference on System Sciences, vol. 2020-Janua, pp. 3350–3358, 2021.

[19] C. Beasley, X. Zhong, J. Deng, R. Brooks, and G. K. Venayagamoorthy, “A survey of electric power synchrophasor network cyber security,” IEEE PES Innovative Smart Grid Technologies Conference Europe, vol. 2015-Janua, 1 2015.
[20] X. Zhong, P. Arunagirinathan, A. Ahmadi, R. Brooks, and G. K. Venayagamoorthy, “Side-channels in electric power synchrophasor network data traffic,” ACM International Conference Proceeding Series, vol. 06- 08-Apri, 4 2015.
[21] Z. Zhang, S. Gong, A. D. Dimitrovski, and H. Li, “Time synchronization attack in smart grid: Impact and analysis,” IEEE Transactions on Smart Grid, vol. 4, no. 1, pp. 87–98, 2013.
[22] X. Fan, L. Du, and D. Duan, “Synchrophasor data correction under GPS spoofing attack: A state estimation-based approach,” IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 4538–4546, 2018.
[23] X. Jiang, J. Zhang, B. J. Harding, J. J. Makela, and A. D. Dom´ınguezGarc´ıa, “Spoofing GPS receiver clock offset of phasor measurement units,” IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 3253– 3262, 2013.
[24] I. Akkaya, E. A. Lee, and P. Derler, “Model-based evaluation of GPS spoofing attacks on power grid sensors,” 2013 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, MSCPES 2013, 2013.
[25] O. Vukovic and G. Dan, “Security of fully distributed power system state estimation: Detection and mitigation of data integrity attacks,” IEEE Journal on Selected Areas in Communications, vol. 32, no. 7, pp. 1500– 1508, 2014.
[26] LiuYao, NingPeng, and R. K., “False data injection attacks against state estimation in electric power grids,” ACM Transactions on Information and System Security (TISSEC), vol. 14, p. 33, 6 2011.
[27] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, “False data injection on state estimation in power systems-attacks, impacts, and defense: A survey,” IEEE Transactions on Industrial Informatics, vol. 13, pp. 411–423, 4 2017.
[28] S. Barreto, M. Pignati, G. Dan, J. Y. Le Boudec, and M. Paolone, “Un- ´ detectable Timing-Attack on Linear State-Estimation by Using Rank-1 Approximation,” IEEE Transactions on Smart Grid, vol. 9, pp. 3530– 3542, 7 2018.
[29] G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: Theory and applications,” Neurocomputing, vol. 70, pp. 489–501, 12 2006.
[30] A. Shehod, “Ukraine power grid cyberattack and us susceptibility: Cybersecurity implications of smart grid advancements in the us,” Cybersecurity Interdisciplinary Systems Laboratory (CISL), Sloan School of Management, Room E62-422, Massachusetts Institute of Technology, Cambridge, MA, vol. 2016, Dec 2016.
[31] T. T. Kim and H. V. Poor, “Attacks on Power Grids,” IEEE Transactions on Smart Grid, vol. 2, no. 2, pp. 326–333, 2011.
[32] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, “Sparse Malicious False Data Injection Attacks and Defense Mechanisms in Smart Grids,” IEEE Transactions on Industrial Informatics, vol. 11, pp. 1198–1209, 10 2015.
[33] Y. Li and Y. Wang, “False Data Injection Attacks with Incomplete Network Topology Information in Smart Grid,” IEEE Access, vol. 7, pp. 3656–3664, 2019.
[34] S. Mousavian, J. Valenzuela, and J. Wang, “A Probabilistic risk mitigation model for cyber-attacks to PMU networks,” IEEE Transactions on Power Systems, vol. 30, no. 1, pp. 156–165, 2015.
[35] C. Tu, X. He, X. Liu, and P. Li, “Cyber-attacks in PMU-based power network and countermeasures,” IEEE Access, vol. 6, pp. 65594–65603, 2018
[36] R. Ma, S. Basumallik, S. Eftekharnejad, and F. Kong, “Recovery-based model predictive control for cascade mitigation under cyber-physical attacks,” 2020 IEEE Texas Power and Energy Conference, TPEC 2020, 2020.
[37] P. Gao, M. Wang, J. H. Chow, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, and M. P. Razanousky, “Identification of successive ’Unobservable’ cyber data attacks in power systems through matrix decomposition,” IEEE Transactions on Signal Processing, vol. 64, no. 21, pp. 5557– 5570, 2016.
[38] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, “Detecting false data injection attacks on power grid by sparse optimization,” IEEE Transactions on Smart Grid, vol. 5, pp. 612–621, 3 2014.

[39] Q. Yang, D. An, R. Min, W. Yu, X. Yang, and W. Zhao, “On optimal PMU placement-based defense against data integrity attacks in smart grid,” IEEE Transactions on Information Forensics and Security, vol. 12, no. 7, pp. 1735–1750, 2017.
[40] X. Lou, C. Tran, R. Tan, D. K. Yau, Z. T. Kalbarczyk, A. K. Banerjee, and P. Ganesh, “Assessing and mitigating impact of time delay attack: Case studies for power grid controls,” IEEE Journal on Selected Areas in Communications, vol. 38, no. 1, pp. 141–155, 2020.
[41] Y. Chen, L. Xie, and P. Kumar, “Dimensionality reduction and early event detection using online synchrophasor data,” in Proc. IEEE Power and Energy Society General Meeting, 2013, pp. 1–5.
[42] Y. Chi, Y. C. Eldar, and R. Calderbank, “PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations,” IEEE Trans. Signal Process., vol. 61, no. 23, pp. 5947–5959, 2013.
[43] N. Dahal, R. L. King, and V. Madani, “Online dimension reduction of synchrophasor data,” in Proc. IEEE PES Transmission and Distribution Conf. Expo. (T&D), 2012, pp. 1–7.
[44] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, “Missing Data Recovery by Exploiting Low-Dimensionality in Power System Synchrophasor Measurements,” IEEE Transactions on Power Systems, vol. 31, pp. 1006–1013, 3 2016.
[45] L. Xie, Y. Chen, and P. R. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” IEEE Transactions on Power Systems, vol. 29, pp. 2784–2794, 11 2014.
[46] M. Eissa and A. Kassem, “Hierarchical Clustering based optimal PMU placement for power system fault observability,” Heliyon, vol. 4, p. 725, 2018.
[47] M. Aoki, “Notes on economic time series analysis: System theoretic perspectives,” Springer, vol. 220, 1983.
[48] Y. Hao, M. Wang, J. H. Chow, E. Farantatos, and M. Patel, “Modelless data quality improvement of streaming synchrophasor measurements by exploiting the low-rank hankel structure,” IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6966–6977, 2018.
[49] M. V. Venkatasubramanian and R. Carroll, “Oscillation Monitoring System.” https://www.naspi.org/sites/default/files/2016-10/08 wsu oms synchrophasors mani 20081016%281%29.pdf/.
[50] V. Venkatasubramanian, “Real-Time Strategies for Unwrapping of Synchrophasor Phase Angles,” IEEE Transactions on Power Systems, vol. 31, no. 6, pp. 5033–5041, 2016

Appendix A

A1. Python Code:

```
#ts = time window start
# te = time window ends
lst = np.arange(ts, te)
err_f = list()
err_t = list()
err_n = list()
print(lst)
for rng in lst[0:]:
  data_s = rng
  data_e = rng+200
  data_m = rng+100
  ang11 = c632[data_s:data_e]
  ang501 = c632[data_s:data_e]
  ang301 = c632[data_s:data_e]

  ang51 = list()
  ang31 = list()
  for kin in np.arange(data_s,data_m):

    ang51.append(ang501[kin])
    ang31.append(ang301[kin])
  for kin in np.arange(data_m,data_e):
    if data_m < 12498:
      ang51.append(ang501[kin])
    else:
```

```
      ang51.append(ang501[kin] + 5)

  for kin in np.arange(data_m,data_e):
    if data_m < 12498:
      ang31.append(c632[kin])
    else:
      ang31.append(c632[kin+200])
 # print('31')
 # print(ang31)
 # print('51')
  #print(ang51)
  tim = np.arange(0,len(ang11))

#ang1 = list()
#ang5 = list()
#ang3 = list()
#angx = ang11.to_list()

#for i12 in range(len(ang11)):
  ang1 = ang11.to_list()
  ang3 = ang31
  ang5 = ang51
 # print(ang1[0])
  ang21 = list()
  ang41 = list()
  ang61 = list()
 # print(math.radians(ang1[1]))
  for ind in range(len(ang1)):
    ang21.append(math.radians(ang1[ind]))
    ang41.append(math.radians(ang3[ind]))
    ang61.append(math.radians(ang5[ind]))

  ang22 = np.unwrap(ang21)
  ang42 = np.unwrap(ang41)
  ang62 = np.unwrap(ang61)

  ang2 = list()
```

```
  ang4 = list()
  ang6 = list()
  for ind1 in range(len(ang22)):
    ang2.append(math.degrees(ang22[ind1]))
    ang4.append(math.degrees(ang42[ind1]))
    ang6.append(math.degrees(ang62[ind1]))

  Y1 = np.array(ang2)
  Y2 = np.array(ang4)
  Y3 = np.array(ang6)
#  print(Y1.shape)



  L =len(Y1)
  K = int(L/2)+1
  H = np.zeros((K,L-K+1))
  for i in range(K):
    for j in range(L-K+1):
      H[i,j] = Y2[i+j]



     u,s,v = svd(H)
  a = linalg.norm(H,'fro')
  norm_list = list()
  errt = list()
  for i1 in range(5):
    u1,s1,v1 = low_rank(u,s,v,i1)
    H2 = np.matmul(np.matmul(u1,s1),v1.T)

    a2 = linalg.norm(H2,'fro')
    norm_list.append(a2)
    if a2 != 0:
      errt.append((a-a2)/a)

  print(len(errt))
  t = np.arange(0,len(errt))
  plt.plot(t,errt,label = "timing attack")
```

```
H = np.zeros((K,L-K+1))
for i in range(K):
  for j in range(L-K+1):
    H[i,j] = Y1[i+j]



    u,s,v = svd(H)
a = linalg.norm(H,'fro')
norm_list = list()
errn = list()
for i1 in range(5):
  u1,s1,v1 = low_rank(u,s,v,i1)
  H2 = np.matmul(np.matmul(u1,s1),v1.T)

  a2 = linalg.norm(H2,'fro')
  norm_list.append(a2)
  if a2 != 0:
    errn.append((a-a2)/a)

print(len(errn))

plt.plot(t,errn,label = "Normal condition")
H = np.zeros((K,L-K+1))
for i in range(K):
  for j in range(L-K+1):
    H[i,j] = Y3[i+j]



    u,s,v = svd(H)
a = linalg.norm(H,'fro')
print(a)
norm_list = list()
errf = list()
for i1 in range(5):
  u1,s1,v1 = low_rank(u,s,v,i1)
  H2 = np.matmul(np.matmul(u1,s1),v1.T)
```

```
    a2 = linalg.norm(H2,'fro')
    print(a2)
    norm_list.append(a2)
    if a2 != 0:
      errf.append((a-a2)/a)

  print(len(errf))

  plt.plot(t,errf,label = "FDIA")
  plt.legend()
  plt.show()
  print('mean for FDIA')
  print(np.mean(errf))
  err_f.append(np.mean(errf)*100)
  print('mean for No attack')
  print(np.mean(errn))
  err_n.append(np.mean(errn)*100)
  print('mean for timing')
  print(np.mean(errt)*100)
  err_t.append(np.mean(errt)*100)
t1 = np.arange(len(err_f))
plt.plot(t1,err_f,label = "FDIA")

plt.plot(t1,err_t,label = "Timing Attack")

plt.plot(t1,err_n,label = "No Attack")
plt.legend()
plt.show()
```