Riphah International University

Masters Thesis

Machine Learning based Intrusion Detection System for Cyber Physical Power System

---

Author:

Abdul Wahab

Supervisor:

Dr. Muhammad Zubair

A thesis submitted in fulfillment of the requirements

for the degree of MS (IS)

in the

Riphah Institute of Systems Engineering

March 10, 2020

RIPHAH INTERNATIONAL UNIVERSITY

Abstract

Information Security

Riphah Institute of Systems Engineering

MS (IS)

Machine Learning based Intrusion Detection System for Cyber Physical Power System

by Abdul Wahab

Cyber-physical systems (CPS) are contributing significantly in improving efficiency and quality of human life. Modern electric power systems equipped with industrial control system (ICS) forms a full-fledged cyber-physical system (CPS), where power flows on the physical system and monitoring & control information flow in the cyber system. Electric power system or smart grid operations can be broadly divided into power generation, transmission, distribution and consumptions. Synchrophasor based wide area monitoring system (WAMS), a state of the art industrial control system for power transmission is the focus of this research study. WAMS enable real time monitoring and control of transmission subsystem, and gradually replacing traditional supervisory control and data acquisition (SCADA).

WAMS fundamental dependence on modern information communication infrastructure, introduces cyber security risks and threatens the entire power transmission network. Data injection, command injection and device setting change are the persistent threats to electrical power cyber-physical system. ICS intrusion detection systems (ICS-IDS) are generally deployed alongside network intrusion detection system (NIDS) and other security measures as defense in depth strategy in smart grids.

Probability based machine learning algorithms have demonstrated their capability for cyber-attack detection in electrical power systems. Their ability to model complex relationships between electrical quantities makes them a preferred choice for distinguishing between normal events and cyber attacks. In the power transmission system, intelligent electronic device (IED), an endpoint device hosting relay and phasor measuring unit (PMU), is potentially capable of detecting intrusions near to source and alarm at an early point to take corrective action. However, the amount of resources is a drag on the operational viability of ML based systems for IED. Furthermore, prevalent approaches to Machine learning in a supervised manner make such systems impractical.

This thesis investigates the development of, a plug and play anomaly detection system. Proposed system learns in an efficient online manner without human supervision. Oak Ridge National Laboratories (ORNL) power system datasets were used to demonstrate the capability and capacity of prototype system.

In this thesis, we present Flare, a lightweight cyber-physical intrusion detection system (CP-IDS). Flare's core algorithm is based on gaussian probability density modeling, for density estimation of multivariate WAMS data. Probability densities of various electrical quantities are aggregated into a single variable to construct the behavioral representation of synchrophasor in normal time. Correntropy measure is operated on the posterior probabilities of learning grace period to accurately establish the baseline for the normal system. Central limit theorem played an important role in Flare's development by serving as a tool to support the assumption that electrical quantities are distributed normally in WAMS.

To demonstrate the runtime performance and operational viability of Flare for both central deployment and substation deployment, two environments were used to perform the benchmarks. For substation intelligent electronic device (IED) deployment experiments were performed on a cheep single board computer (SBC), Raspberry Pi 3B. For central deployment, experiments were performed on Intel Core i7 3.07 GHz PC running *Windows* 10. Empirical evaluations demonstrate that Flare, is a practical and economic CP-IDS for both deployment strategies central and substation IED. The experiments showed clear superiority of proposed detection framework over other algorithms.

Chapter 1

INTRODUCTION

Evolution of conventional electric power grid into a flexible IP based smart grid introduces vulnerabilities that could bring catastrophic grid failure (Otuoze, Mustafa, and Larik, 2018). Complex information infrastructure presents a large plane for cyber-attacks and number of cyber-attacks targeting industrial control systems are increasing day by day (Sanjab et al., 2016). Cyber-attacks on control systems attempts to corrupt control signals or measurement signals and disguise as power system disturbances or control actions (Pan, Morris, and Adhikari, 2015a).

This research study is aimed at development of an efficient and cost effective cyber-physical intrusion detection system (CP-IDS) for rapidly transforming national electrical power system of Pakistan into a highly inter-connected modern smart grids. An industrial control system IDS (ICS-IDS) or CP-IDS, both terms are often used interchangeably, is a device or software which monitors electrical measurements and upon detection of malicious activity, alerts the operator. CP-IDS for electrical power systems can be deployed at substation site with PMU/ relay by implementing it on a FPGA (or any other smart board) or a stand-alone single central server at control room. (Hink et al., 2014).

This chapter sets out the context with a background by giving an overview of present day power system control systems and cyber security challenges of WAMS. Outlines the research questions and an overview of thesis organizations.

- **See section 1.1** for background
- **See section 1.2** for electrical power systems in present era
- **See section 1.3** for cybersecurity challenges of WAMS and intrusion detection technologies
- **See section 1.4** for research challenges/ questions
- **See section 1.5** for a brief on thesis organisation

1.1 Background

Tolerance for power system unavailability is often low due to high dependency of society and its critical role in country's economics. National Transmission and Dispatch Company (NTDC) is revamping its power transmission monitoring infrastructure to control expanding power grid system under Power Transmission Enhancement Investment Program II Tranche 3 (Asian Development Bank (ADB), 2018). National Power Control Center (NPCC) Islamabad recently upgraded its national power system with a state-of-the-art SCADA for transmission monitoring and an energy management system (EMS) for efficient frequency control, economic dispatch and optimal power flow.

Network intrusion detection system (NIDS) are generally part of a secure smart grid infrastructure, however CP-IDS, where input to the IDS, is the operational process data of the underlying system is an important and effective capability to detect malicious activities in the power system. Resilience of modern smart grid control systems

to cyber-attacks can significantly impact the decision for early adoption of state of the art electrical power system in future. Improved capability, to thwart cyber-attack attempts, of electrical power systems is essential for faster adoption of future electrical power system and smart grids.

1.2 Electrical Power Systems in Present Era

A typical electrical power infrastructure can be viewed as a system of four cascade processes: generation, transmission, distribution and consumption (Ali Abur, 2004). Transmission subsystem, is the vital constituent of a national power infrastructure and is responsible for transporting the bulk power over long distance from generation site to the distribution/ load centers. Transmission infrastructure is geographically spread out across large administrative regions throughout a territorial dominion. An industrial control system (ICS) for power system is an overlay system designed for monitoring, control and management of underlying industrial process such as power generation, transmission, distribution and consumption (Pan, 2014). ICS choices for power transmission subsystem are,a traditional Supervisory Control and Data Acquisition (SCADA) or prevalent Wide Area Monitoring System (WAMS).WAMS (Wide area monitoring system) is an evolved form of traditional SCADA of power transmission subsystem. Classical SCADA systems are not able to provide infrastructure's state information at sub-second time frame (*Advantages of Synchrophasor Measurements Over SCADA Measurements for Power System State Estimation*).

1.3 Cybersecurity challenges of WAMS and Intrusion Detection Technologies

Cyber physical electrical power system's fundamental dependence on information & communication technologies (ICT) inherits a large attack surface. In synchrophasor based WAMS, PMU communicate the synchrophasors using IEEE C37.118 protocol, attacker can be intercept and alter these measurements to influence the operator at control center to take wrong decision. Attacker can also modify the relay operating configurations by exploiting the remote relay setting change feature. Control actions can also be intercepted and modified to trip the breakers and cause the blackout or execute potentially more dangerous attacks.

This research addresses command injection, remote relay setting change and data injection attacks,

- Remote Tripping Command Injection Attack **See section 3.1**
- Relay Setting Change Attack **See section 3.2**
- Data Injection Attack **See section 3.2**

Numerous cyber-attack detection technologies developed in recent years for cyber-physical power systems ranging from traditional power state estimation using purely electrical and mathematical calculations to Machine learning (ML) based intrusion detection systems (IDS). (J. Valenzuela, 2013) proposed and developed a cyber attack detection technique based purely on electrical theories like power flow fluctuations. (Talebi, Wang, and Qu, 2012) developed a system that works on the principal of weighted power state estimations to detect the intrusions. (Foroutan et al.,

2017) developed attack detection system using relationship between angle and voltage changes.

Machine learning and data mining based systems attempts to model behaviors of components or communication pattern between components. (Hink et al., 2014) evaluated multiple ML algorithms to validate the usefullness of ML algorithms for cyber-physical power transmission systems, probabilistic classifier (Naïve Bayes), Rule based classifier (OneR, NNge, JRipper), Decision tree based classifier (Random Forests), Non-probabilistic binary classifier (SVM) and Boosting and a meta-algorithm classifier (Adaboost).

1.4 Research Challenges/ Questions

This research study investigates the development of a lightweight plug and play anomaly detection system for synchrophasor based WAMS. Developed CP-IDS can be embedded directly into Intelligent electronic device (IED) or can be deployed in an inexpensive commodity hardware to work with IED.

An effort is made to develop a light weight plug and play CP-IDS with high detection accuracy and low computing requirements.This is accomplished by developing an extension of gaussian mixture probability distribution modeling (Haider, 2018) by modifying it to run in online manner with human supervision.

The research questions pursued in an attempt to address the present-day open problems, at least to some extent are as follows:-

1.4.1 Research Question 1

CP-IDS can be deployed centrally or in the sub-station at IED location.In sub-station IED deployment, each IED will be equipped with a lightweight CP-IDS, so CP-IDS will not consider attacks that occur on a separate line. Operation viability of machine learning based CP-IDS is a challenge for sub-station IED deployment, also highlighted in Hink et al., 2014.

> ***How can a lightweight CP-IDS be designed, to embed directly in an Intelligent electronic device (IED), for sub-station IED deployment?***

Develop a light-weight CP-IDS for power transmission sub-system to embed directly in Intelligent Electronic Device (IED), a central component of WAMS. Considering the geographically scattered power transmission networks, IED deployment is a practical strategy to detect malicious activity near to source. Research and development with regards to CP-IDS deployment scenarios clearly need attention owing to paucity of literature in this regard.

1.4.2 Research Question 2

> ***Can a stream processing algorithm be developed to process and classify high-speed data of WAMS in an online fashion and detect complex unknown attacks?***

The computational complexity of a ML based IDS in handling high velocity synchrophasor data, 120 measurements per second challenges their operational viability for sub-station deployment. A PMU produce 29 measurements in a faction of second (i.e. 120th of second), it means a matrix of 120 x 29 is produced in a second.

Processing such a, high speed, heterogeneous data, and effectively learn the underlying patterns is out of question with prevalent ML algorithms without fair amount of computing resources. See section 2.3.2 for real-time intrusion detection algorithms. Tomlin, Farnam, and Pan, 2016, Adhikari, Morris, and Pan, 2018 and Adhikari, Morris, and Pan, 2017 specific to electrical power transmission systems and Mirsky et al., 2018 for network intrusion detection in general.

1.4.3 Research Question 3

> ***Prevalent ML based solutions are qualified under direct supervision, can we develop an algorithm that learn without supervision with the aim to reduce heavy costs of manual labour and increased automation?***

In conventional machine learning based decision engines subject input data from industrial control system is offloaded from the production, Copy of the data is transferred to the training facility, off-line model is built and trained model is loaded to the production decision engine. ML based IDS reported in literature are focuses on prevalent approach of offline learning and model building using high grade CPU or GPU. These approaches lower the operational viability of ML based intrusion detection implementations.

Typical approach to machine learning based CP-IDS encompass following steps,

- Create a dataset encompassing both benign and attack time measurements of various components of transmission control system.
- Train the machine learning algorithm using a high-end computing machine with GPUs to model the difference between the normal and attack behaviors.
- Transfer a copy of the trained model to the substation's IDS.

- IDS uses that built model to make decision on real-time measurements.
- Re-train the model to dispose of identified false positives and improve the knowledge with passage of time and repeat the step three.

1.4.4 Research Question 4

If Gaussian density estimation and Correntropy approaches can be used to model the multivariate data of synchrophasor based WAMS?

Establish the possibility of applying the GMM and Correntropy approaches (Haider, 2018), (Lagrange, Fauvel, and Grizonnet, 2017), (Queiroz et al., 2016) and (Liu et al., 2017) on wide area monitoring system (WAMS) of electrical power transmission sub-system.

1.4.5 Research Question 5

If gaussian modeling can achieve the detection score greater than 73.43% against zero day attacks, achieved by common path mining algorithm?

Existing IDS for cyber-physical power systems achieved low zero-day detection score of 73.43 by common path mining algorithm, possibility of improvement in zero day attack detection capability will be investigated. (Pan, Morris, and Adhikari, 2015b).

1.5 Thesis Organization

The thesis document is arranged into five chapters.

Chapter 2 presents a comprehensive literature review covering, i). Machine learning based intrusion detection systems for power systems with respect to their operational viability ii). Intrusion detection systems employing the online data processing techniques and iii). Decision engines based on Gaussian Probability based Modeling Techniques.

Chapter 3 discusses power transmission sub-system faults, control commands/ actions and cyber attack events. Cyber attacks against electrical power systems includes data injection, command injection and aurora (Relay setting change) attacks. Oak Ridge National Laboratory (ORNL) and Mississippi State University Power System Datasets are introduced.

Chapter 4 discusses some basic concepts of probability, central limit theorem, gaussian probability distribution, correlation entropy and detailed account of anomaly based intrusion detection system developed using these theories.

Chapter 5 presents the empirical results of experiments performed for evaluation of Flare's accuracy. Learning time, execution time and details of the environments used for experiments are discussed.

Chapter 6 summarize the findings of this investigation account and presents the research contributions with highlights to future research possibilities.

Chapter 2

Literature Review

Chief motivation for this research study is the open research on operational viability of ML based IDS in modern power system or smart grids. (Hink et al., 2014) highlighted the deployment issues and urged the future work on development of a practical ML based IDS for power systems. Secondly, the burdensome supervised learning process employed in existing ML based IDS for cyber physical power systems and thirdly low zero-day detection score 73.43% reported(Pan, Morris, and Adhikari, 2015b). For last 8 years adequate research conducted on cyber security concerns of power systems.

National Transmission and Dispatch Company Limited (NTDC), is upgrading its supervisory control and data acquisition (SCADA)/ energy management system (EMS) for better situation awareness, pro-active monitoring and improved control over the Pakistan national power grid (Asian Development Bank (ADB), 2017). Along with enormous benefits of latest technologies serious threats are posed. Pakistan's critical national infrastructure also faces threats from the cyber world (Rubab Syed, 2019). Country has faced multiple state level cyber breaches (*Pakistan's cyber security*).

Catastrophic consequences of BlackEnergy 33 to Ukraine power grid, Stuxnet to

Iran's nuclear program and continuously growing number of such attacks targeting critical national cyber-physical systems is alarming. "A successful cyber-attack on critical infrastructure could do as much damage as a natural disaster, bringing a whole country to a standstill." (*Cyber-Attack US Struggle Taken Offline Power Grid*).

This chapter provides the technical background on Industrial control systems of power systems, ML based Cyber-physical IDS (CP-IDS) for synchrophasor based WAMS and a comprehensive survey of CP-IDS or ICS-IDS found in literature.

- **See section 2.1** for Industrial Control Systems for Power Transmission Subsystem
- **See section 2.2** for ML based Cyber-physical IDS (CP-IDS) for WAMS
- **See section 2.3** for Current CP-IDS or ICS-IDS
- **See section 2.3.1** for Data Mining and Machine Learning based IDS for Power Transmission System
- **See section 2.3.2** for Real Time Intrusion Detection in Electrical Power Transmission Systems
- **See section 2.3.3** for Intrusion Detection Systems based on Gaussian Naive Bayes Modeling Techniques

2.1 Industrial Control Systems for Power Transmission Subsystem

WAMS (Wide area monitoring system) is an evolved form of *traditional* ***SCADA*** of power transmission subsystem. Classical SCADA systems are not able to provide infrastructure's state information at sub-second time frame (*Advantages of Synchrophasor Measurements Over SCADA Measurements for Power System State Estimation*). State of the art wide area monitoring system, rely on synchrophasor technology (Mynam, Harikrishna, and Singh, 2011). A typical synchrophasor based WAMS is a networked system of phasor measurement units, phasor data concentrators, relays and other hardware modules (see Figure 2.1) such as protective relays, intelligent electronic device (IEDs), industrial computers, firewalls, network intrusion detection systems (NIDS), switches and routers. Cyber physical electrical power systems also contains various application softwares (Databases,energy management system and historians etc.), industrial PCs, industrial radios and numerous communication protocols (IEEE C37.118, IEC.61850 protocol stack, MODBUS and DNP 3.0 etc.). WAMS enables real time visualization capability of a power system, power system state estimation and event detection, and(Adhikari, 2015). Data acquisition technology of phasor measurements, enables monitoring of transmission system behaviors over large geographic areas to discover and deny grid volatility (Sanjab et al., 2016). The PMUs are deployed to monitor health of the electrical power transmission system by measuring relevant parameters, including phasor measurements. Phasor or synchrophasor are time-synchronized measurements of multiple eletrical quantities and numerical estimations that represent both the magnitude and phase angle of the sine waves found in electricity. Synchrophasors are time-synchronized using wireless communication technologies among geographically dispersed PMUs

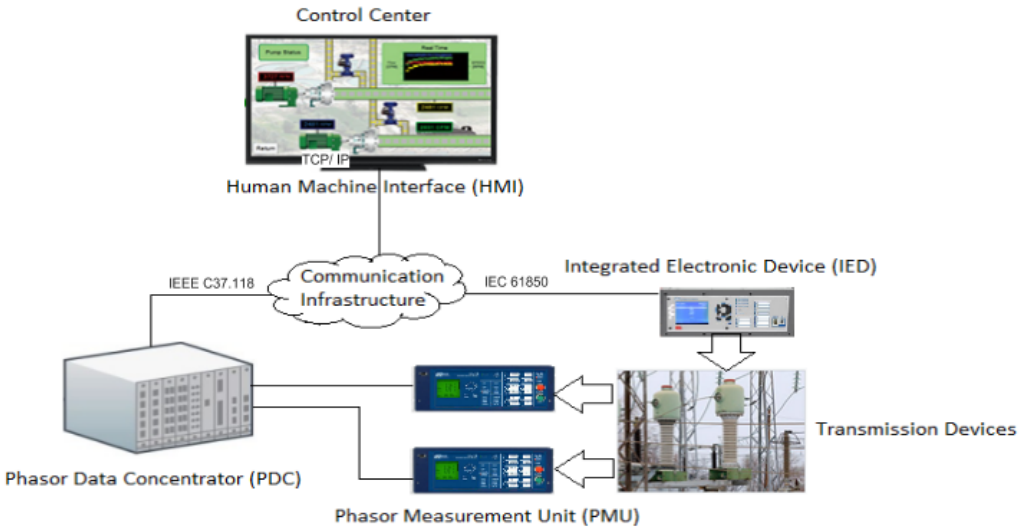


FIGURE 2.1: Typical architecture of synchrophasor based WAMS

for accuracy(*Synchrophasor Applications in Transmission Systems*).

SCADA ecosystem is composed of field sensors, actuators, RTU (Remote telemetry unit), MTU (Master telemetry unit) and HMI (Human Interface). Transmission subsystem SCADA field sensors measure various electrical quantities, voltage magnitudes, active and reactive power values of power transmission lines. (Mynam, Harikrishna, and Singh, 2011). RTUs installed at remote substation forward the collected quantities to MTU (Master Telemetry Unit) which is installed at control center. RTU is a real-time programmable logic controller (PLC) which converts the sensor's analog information into digital form before communicating to MTU. Control center is equipped with HMI to enable the remote monitoring, administration and ladder logic update. RTU also receives the control signals from master telemetry unit (MTU) in order to control the process equipment through switch-boxes, relays and breakers.

2.2 ML based Cyber-physical IDS (CP-IDS) for WAMS

Cyber-attack detection in power systems using machine learning received considerable attention in recent years. (Pan, Morris, and Adhikari, 2015c) proposed and developed a novel specification-based intrusion detection employing bayesian network to graphically encode the causal relations to act as , which are used as rules in the proposed intrusion detection framework. (Tomlin, Farnam, and Pan, 2016) developed a cluster analysis and classifying states using the Mamdani fuzzy inference. (Adhikari, Morris, and Pan, 2018) developed a system using Hoeffding Adaptive Trees (HAT) augmented with the drift detection method (DDM) and adaptive windowing (ADWIN). (Adhikari, Morris, and Pan, 2016) proposed State Extraction Method (STEM) for data preprocessing and applied Non-Nested Generalized Exemplars(NNGE) for classification for Cyber-Power Event and Intrusion Detection. (Pan, Morris, and Adhikari, 2015a) proposed a novel pattern mining approach to extract and discriminate the patterns of power-system disturbances and cyber-attacks. (Chen et al., 2018) selected five classifier as base, Bayesnet, OneR, Ripper, C4.5 and SVM to evaluate their performance with ensemble method, Adaptive boosting, Bagging, Majority voting and Random forest.

Generally, ML based Intrusion Detection System for power system as shown in Figure 2.2 consist of,

- Data Source
- Machine Learning (ML) Algorithm for learning and execution
- Data offloading from production systems
- Modeling building

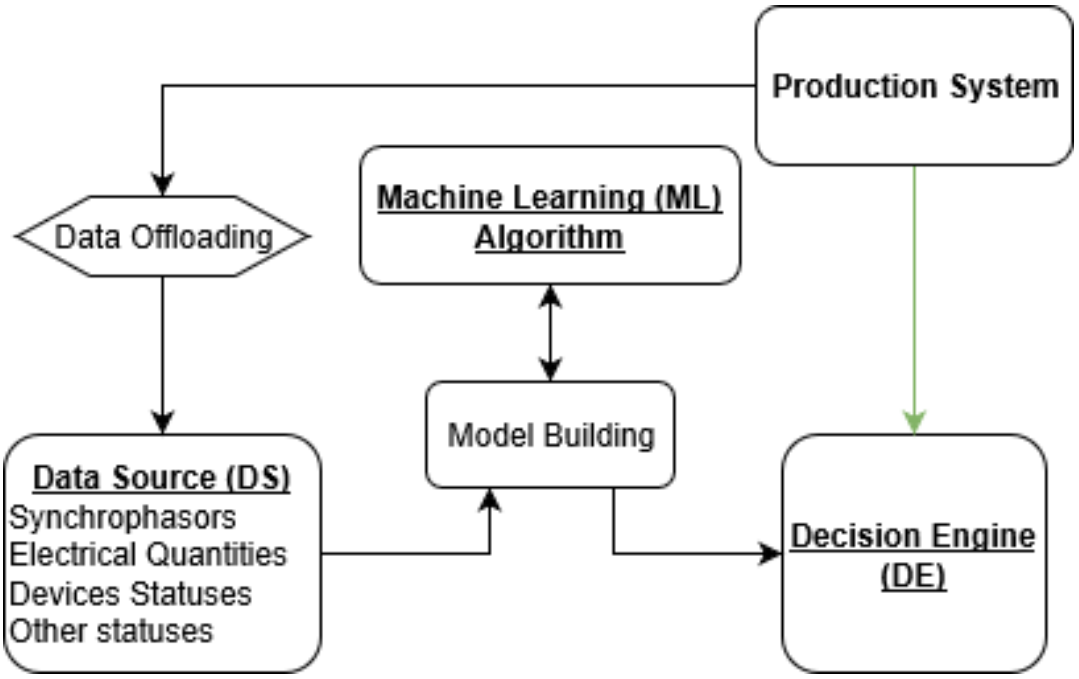


FIGURE 2.2: Typical CP-IDS architecture for synchrophasor based WAMS

- Re-train the model to dispose of identified false positives and improve the knowledge with passage of time and repeat the step three.

The Data Source (DS) reflects the activities occurring in an electrical power system equipped with synchrophasor based WAMS. 29 x electrical quantities (See Table **??**)are calculated by a phasor measurement unit, called synchrophasor. Synchrophasor, encapsulated in a IEEE C37.118 instance by PMU becomes the data source for CP-IDS. Three other types of information are also available, namely: (i) Snort, Network intrusion detection system (NIDS) alert; (ii) Control panel log file entries; and (iii) relay status. Machine Learning (ML) Algorithms are used to learn the complex relationship among features by creating the models on offloaded data. During execution phase decision engine (DE) is mandated to discriminate between the normal event and cyber-attack using learned model.

2.3 Current CP-IDS or ICS-IDS

During last decade various governments heavily funded universities to study the cyber security concerns of the modern cyber physical electrical power systems and develop effective counter measures. Notable work is undergoing at Idaho National Laboratory, Sandia National Laboratory, Critical utility infrastructural resilience (CRUTIAL), Royal Melbourne Institute of Technology and Lowa State University. Despite other cyber security measures (compliance to standards, policies, control system network isolation and firewalls) to protect the cyber physical electrical transmission system, network activity and operating process monitoring is fundamental to identify patterns related to cyber-security threats. Electrical power industrial control systems testbed at Mississippi State University created in collaboration with Oak Ridge National Laboratories (ORNL) focuses on emerging smart grid technology-synchrophasor based wide area monitoring systems (WAMS).

Intrusion detection systems, techniques found in literature are organized into three sections as,

- Data Mining and Machine Learning based IDS for Power Transmission System ref to **Table** 2.1
- Real Time Intrusion Detection in Electrical Power Transmission Systems ref to **Table** 2.2
- Intrusion Detection using Gaussian Naive Bayes Modeling Techniques ref to **Table** 2.3

2.3.1 Data Mining and Machine Learning based IDS for Power Transmission System

(Hink et al., 2014) evaluated multiple classification algorithms to determine machine learning methods ability to discriminate fault event against cyber-attack disturbances using electrical measurements in smart grids. Second purpose of the (Hink et al., 2014) was to weigh practical implications of deploying machine learning systems as an extension to existing power system architectures. Probabilistic classifier (Naïve Bayes), Rule induction (OneR, NNge, JRipper), Decision tree classifier (Random Forests), Non-probabilistic binary classifier (SVM) and a meta-algorithm for learning boosting (Adaboost) were evaluated using Weka (An Open Source Machine Learning Framework). Multiclass, Three-class and binary power system datasets provided by Mississippi State University were used in the experiments. Highest accuracy and F-measure score was reported by JRipper+Adaboost Table 2.1 summarizes the results. Operational deployment issues (acquiring and maintaining in-situ training data , the need for labeled training data, classification performance, learning system feedback and retraining criteria) of a classification system for smart grids and envisioned (Hink et al., 2014) future work turned into the cheif motivation of this research study.

(Pan, Morris, and Adhikari, 2015c) developed a specification-based intrusion detection framework by using bayesian network graphic encoding, a probabilistic network to specify system behaviors by the transitions of system states. two bus two generator transmission system is simulated with over current protection scheme and nine power system natural disturbance and cyber attack scenarios were implemented. Proposed IDS trained using datasets of the scenarios correctly classified all the test scenarios ref to Table 2.1 for details . (Pan, Morris, and Adhikari, 2015c) discussed

both deployment strategies i.e. central and substation but considered resource requirement a future work. Requirement of system expertise and buden some process of specification development is also discussed.

A novel pattern extraction methodology to classify power-system disturbances from cyber-attacks is proposed by (Pan, Morris, and Adhikari, 2015a). Labeled data is used to discover common path, discovered paths are used for classification. Three experiments were performed. The training and testing dataset used in experiment 1 were subdivided according to 1LG fault and command injection attack subtypes for experiment 2. Four types of symmetric and unsymmetrical faults and three types of cyber-attacks dataset was used for experiment 3. 10 round cross-validation used in all experiments and accuracy achieved were 95%, 87.6% and 93.21% respectively Table 2.1 summarizes the results for experiment1, experiment2 and experiment3.

(Pan, Morris, and Adhikari, 2015b) A common path mining based IDS prototype developed in Referene4 to classify disturbances, normal control operations, and cyber-attacks for the distance protection scheme for a two-line three-bus power transmission system. Dataset of 25 scenarios was generated by simulating a hardware in loop two-line three-bus power transmission system. 90.4% classification accuracy of 25 scenarios , 84.6% for SLG fault sub-groups by fault location and circuit load. Proposed system achieved average detection accuracy of 73.43 % for zero-day attack scenarios Table 2.1 summarizes the results.

(Chen et al., 2017) designed a Distributed Event and IDS. To model power system behaviors specification based methodology is adopted. Various feature selection methods such as mutual information, brute force and correlation are employed to evaluate classification capability of attributes of Mississippi State University Power System Dataset. NNGE, J48, IB1, NNGE-PSO, JRip and Random forest are trained

and tested on features selected using Joint Mutual Information(JMI), Joint Mutual Information Maximization (JMIM) and Correlation-Based Feature Selection. NNGE-PSO scored highest accuracy for 98.380%, 98.011% and 94.039% for Two class, Three class and multi class dataset respectively Table 2.1 summarizes the results.

(Adhikari, Morris, and Pan, 2016) evaluated the Non-Nested Generalized Exemplar classifier for feature states constructed through State Extraction Method (STEM). Mississippi State University Power System Datasets were used to extract the states information. Three experiments were performed to evaluate the classification accuracy of STEM-NNGE on multiclass, binary class and multiclass on IEEE 9 Bus system. Classification accuracy achieved with STEM-NNGE is 94 %, 98% and 99.54 respectively Table 2.1 summarizes the results. Experiments were performed with on a thirty two GBs of memory and 4 x core Intel® Xeon® with CPU speed of 3.70 GHz.

(Demertzis and Iliadis, 2018) recommends the employment of the An adaptive elitist differential evolution (AEDE) optimization for selection of weights and bias of the ELMs rather than random initialization. Performance of the proposed AEDE-ELM is evaluated and compared with RBFANN, GMDH, PANN and FNNGA learning algorithms. Proposed classifier achieved the maximum accuracy of 96.55% summary is presented in Table 2.1.

(Wilson et al., 2018) presents an autoencoder based deep learning framework to develop machine-learned features against power transmission attacks. The proposed stacked autoencoder (SAE) used for automatic and adaptive attack detection and unsupervised feature learning framework. This approach optimized machine-learned

detection accuracy. Experiments were performed on power system datasets of Mississippi state university. Proposed implementation learns offline and executes in online manner. Overall 99.67%, 98.68% and 97.99% accuracy is achieved with select 32, selected 64 and original 128 features Table 2.1 summarizes the results.

(Chen et al., 2018) evaluated multiple ensemble learning methods as cyberattack detectors for industrial control systems of electrical power systems. (Chen et al., 2018) discussed the practical considerations for designing and deploying ensemble learning methods such as Bagging, Adaptive boosting, Random forest and Majority voting. The investigation of deployment viability in a live operation environment is considered a future work.

TABLE 2.1: Comprehensive Comparison of ML based IDS for Electrical Power Transmission System

Ser	Algorithm(s)	Detection Capability	Results	Learning Methdology	Deployment
1	Naïve Bayes,OneR, NNge, Jripper,Random Forests,SVM,Adaboost (Hink et al., 2014)	Multiclass (37 event scenarios) Three-class (Attack, Natural Disturbance, and No Event) Binary (Attack, Normal)	JRipper+Adaboost having the highest overall F-Measure 0.955	Supervised ML	Further work
2	Novel Specification-based intrusion detection-Rules Bayesian network is used to graphically encode the causal relations, which are used as rules in the proposed intrusion detection framework.(Pan, Morris, and Adhikari, 2015c)	Nine power system scenarios (Data sets derived fro Power System Dataset)	100% for Nine scenarios	Supervised ML	Sub-Station
3	Sequential pattern mining approach to accurately extract patterns of power-system disturbances and cyber-attacks (Pan, Morris, and Adhikari, 2015a)	One Non-attack Fault and Three Cyber-attack datasets (Data sets derived fro Power System Dataset)	95 % for Experiment 1 87.6% fro Experiment 2 93.21 % for Experiment 3	Supervised ML	Central
4	Temporal state-based specifications Learning Data Mining - common path mining algorithm(Pan, Morris, and Adhikari, 2015b)	Twenty five scenarios consisting of stocktickerSLG faults, control actions, and cyberattacks	90.4 % accuracy 73.43 % For Zero-days	Supervised ML	Central

TABLE 2.1: Comprehensive Comparison of ML based IDS for Electrical Power Transmission System

Ser	Algorithm(s)	Detection Capability	Results	Learning Methdology	Deployment
5	Specification-Based Intrusion Detection Systems Brute force method, correlation based feature selection (CFS) and mutual information based feature selection with Particle Swarm Optimization Decision tree, instance-based learning and NNGE(Chen et al., 2017)	Binary and Multi Class datasets	NNGE+STEM algorithm provides 96% and 93% detection accuracy for Binary and Multi Class datasets HAT algorithm provides 98% and 92% for Binary and Multi Class datasets respectively	Supervised ML	Sub-Station
6	NNGE+STEM (Adhikari, Morris, and Pan, 2016)	Two Non-attack Contingency and Four Cyber-attack (Data sets derived fro Power System Dataset)	94.0% with NNGE for Multiclass Classification 98% with NNGE for Binary Class Classification 99.54% with NNGE for Multi Class Classification on the IEEE 9 Bus System	Supervised ML	Not Discussed
7	Extreme Learning Machine (ELM) model, which is optimized by the Adaptive Elitist Differential Evolution algorithm (AEDE)(Demertzis and Iliadis, 2018)	3 classes datasets (No Events, Normal Events, Attack)	96.55% with proposed AEDE-ELM	Supervised ML	Not Discussed

TABLE 2.1: Comprehensive Comparison of ML based IDS for Electrical Power Transmission System

Ser	Algorithm(s)	Detection Capability	Results	Learning Methdology	Deployment
8	Stacked autoencoders(Wilson et al., 2018)	Attack Vs Normal (Data sets derived fro Power System Dataset)	99.89 % for Normal Events 98.94% for Data Injection Attacks 97.34% for Command Injection Attack 98.53 % for Relay Setting Change Attack	Offline Supervised Learning and Online Execution	Not Disscused
9	Probabilistic classification (Bayesnet), Rule induction (OneR, Ripper), Decision tree learning (C4.5), Non-probabilistic binary classification (SVM), The Ensemble Methods (Adaptive boosting, Bagging, Majority voting, Random forest)(Chen et al., 2018)	Attack Vs Normal (Data sets derived fro Power System Dataset)	Bagging C4.5, Adaboost Ripper, Adaboost C4.5 and Random forest achieved more than 90 % accuracy	Supervised ML	Further work

2.3.2 Real Time Intrusion Detection in Electrical Power Transmission Systems

(Tomlin, Farnam, and Pan, 2016) developed an unsupervised anomaly-based intrusion detection system works by quantifying the degree by which an event is an attack. The clustering approach employed classifies states using the Mamdani fuzzy inference system. Subset of power system datasets of Mississippi State University consisting of 37 event scenarios was used to evaluate the proposed algorithm accuracy. 99.11 % overall accuracy is achieved Table 2.2 summarizes the results. (Tomlin, Farnam, and Pan, 2016) argued the IDS placement at electrical transmission substations as a single IDS instance for multiple substations may have large system bandwidth and hardware impact.

(Adhikari, Morris, and Pan, 2018) developed an event and intrusion detection system (EIDS) considering high speed live WAMS data. This research contributed towards realization of a practical EIDS based o Hoeffding Adaptive Trees (HAT) augmented with the drift detection method (DDM) and adaptive windowing (ADWIN). Implementation effectively classified cyber contingencies in real time with a reasonable accuracy of greater than 94% for multiclass and greater than 98% for binary class Table **??** summarizes the results. Datasets used consist of 45 classes of cyber-power contingencies. Deployment of HAT+DDM+ADWN based EIDS is earmarked future work. Possible deployment approaches are discussed, distributed/ substation EIDS. Second, a macro level EIDS to monitor grid subsystems.(Adhikari, Morris, and Pan, 2018) discussed the burdensome process of manual labeling, training time and their impact on operational viability.

(Adhikari, Morris, and Pan, 2017) evaluated three data mining and machine learning

algorithms mining common paths, NNGE, and HAT using data from the power system test bed described in the paper. 45 class labeled dataset ws used to evaluate the NNGE and HAT classification accuracy while seven class dataset used for evaluating the common path mining algorithm. Results reported for NNGE and HAT are 94% and 94% respectively. Common path mining algorithm achieved the 93% accuracy Table 2.2 summarizes the results.

(Mirsky et al., 2018) developed an Online, Unsupervised, Low Complexity using ensemble of autoencoders for online anomaly detection in computer networks. Experimental results on an operational IoT network demonstrate the algorithm's efficiency, and ability to run on a low end computing platforms. An online technique for automatically constructing the ensemble of autoencoders (i.e., mapping features to ANN inputs) in an unsupervised manner.

TABLE 2.2: Real Time Intrusion Detection in Electrical Power Transmission Systems

Ser	Algorithm(s)	Detection Capability	Results	Learning Methdology	Deployment
1	Cluster analysis and classifying states using the Mamdani fuzzy inference. Fuzzy c-means (FCM) algorithm combined with the fuzzy inference system (FIS)(Tomlin, Farnam, and Pan, 2016)	Attacks Vs Normal (Data sets derived fro Power System Dataset)	99.11 % Overall accuracy	Unsupervised ML	Sub-Station
2	NNGE HAT Common Path Mining(Adhikari, Morris, and Pan, 2017)	Multiple Data sets derived fro Power System Dataset	94 % with NNGE 94% with HAT 93% with Common Path Minng Algoithm	Supervised ML	Not Disscused
3	Hoeffding Adaptive Trees (HAT) augmented with the drift detection method (DDM) and adaptive windowing (ADWIN)(Adhikari, Morris, and Pan, 2018)	Two Non-attack Contingency and Four Cyber-attack (Data sets derived fro Power System Dataset)	The average kappa statistic was 94% for binary and 93% for multiclass classification.	Supervised ML	Central
4	Ensemble of Autoencoders with a custom feature extraction framework (Adhikari, Morris, and Pan, 2018)	Detect the IoT botnet by analyzing traffic traversing the TCP/ IP network	Area under the receiver operating characteristic curve (AUC) and the equal error rate (EER) of 0.9985 and 0.0032 respectively for mirai botnet	Unsupervised ML	Distributed deployment strategy

2.3.3 Intrusion Detection using Gaussian Naive Bayes Modeling Techniques

Gaussian mixture models (GMMs) have proved their worth for modeling text-independent speaker recognition systems, Linux host malware detection systems and various other modeling applications such as pattern recognition or classification problems in various fields (Reynolds, Quatieri, and Dunn, 2000), (He, Zheng, and Hu, 2010), (Queiroz et al., 2016), (Haider, 2018) and (Tapkir et al., 2018). (Reynolds, Quatieri, and Dunn, 2000) proposed a gaussian Mixture Model-Universal Background Model (GMM-UBM) for speaker verification/detection. (Haider, 2018) proposed the Gaussian mixture model (GMM) and Correntropy technique for anomaly detection in Linux-based hosts against unknown attacks. The posterior probabilities are used as input to the Correntropy to improve the efficiency as probabilities can more accurately estimate the dependencies of the prior and likelihood distributions (Haider, 2018). The proposed methodology well performed for detecting known and zero-day attacks.

TABLE 2.3: Intrusion Detection Systems based on Gaussian Naive Bayes Modeling Techniques

Ser	Algorithm(s)	Detection Capability	Results
1	Gaussian mixture with universal background modeling (Reynolds, Quatieri, and Dunn, 2000)	Gaussian mixture model (GMM)-based speaker verification system	Successfully recongnised all the cases
2	Semi parametric and non-parametric Gaussian distribution modeling decision Engine(Queiroz et al., 2016)	Healthy Vs Faulty Hard Disk Drives (SMART dataset provided by the Center for Magnetic Recording Research, University of California, San Diego)	92.21 % detection before some minitures of failure
3	Gaussian mixture model (GMM) and Correntropy technique with fuzzy rough set attribute reduction (FRAR) to identify hidden patterns (Haider, 2018)	Detect Anomalies in Linux-based Hosts (NGIDS-DS, KDD-98 and Kayacik datasets)	95.48 % 97.54 % and 99.55 % with proposed Corr-GMM and joint feature for NGIDS-DS, Kayacik and KDD-98 respectively
4	Gaussian Mixture Model (GMM) classifier for Power Normalized Cepstral Coefficients (PNCC) and Q-Log Normalized Cepstral Coefficients (QLNCC) (Tapkir et al., 2018)	Detect power networks anamolies of false data injection (Synthetic dateset generated by simulating IEEE 118-bus test system from MATPOWER)	Equal error rate of 12.98 % with proposed technique

Chapter 3

Power transmission system disturbances, control actions, cyber-attacks and ORNL datasets

Electrical power transmission networks may encounter various types of disturbances (faults and contingencies) and also susceptible cyber-attacks while in operation. A disturbance in transmission system can be characterised by over current, unbalanced phases, voltage dips, drastic changes of line impendence and fluctuations in frequencies. Transmission system faults can be classified into symmetric and asymmetric, and may be caused by natural disaster (i.e. lighting, wind, tree falling on lines) or equipment failure. Transmission sub-system contingencies includes transmission line loss, generator loss, and pre-scheduled line maintenance activities.

Cyber-attacks can cause power transmission system faults and contingencies. Assuming the attacker has gained the physical access to the network devices, cyber-attack can be realized by, **i)**. attacker injects false synchrophasor data to mimics a real fault, **ii)**. attacker injects control command to mimic the command from control

center, **iii)** attacker injects a false device setting change command to mimic a device setting change command from control center.

Due to confidentiality issues and proprietary nature of the data containing the power system fault events and cyber-attacks, research and development in ML based systems is relatively slow. Oak Ridge National Laboratory (ORNL) and Mississippi State University power system datasets were generated using three-bus two-line transmission system testbed with two zone distance protection scheme.

This chapter introduce the natural/ usual transmission system disturbance events, cyber-attacks on transmission line control system and the description of ORNL and Mississippi State University Power System Datasets. Subject datasets are a true representative of a contemporary electric power transmission system equipped with synchrophasor based WAMS, and available publicly.

- **See section 3.1** for Power System Faults and Line Maintenance Scenarios
- **See section 3.2** for Power System Cyber-attacks
- **See section 3.3** for ORNL and Mississippi State University Power System Datasets

3.1 Power System Faults and Line Maintenance Scenarios

3.1.1 Power System Faults

Electrical power transmission networks faces to two type of faults, symmetric faults and unsymmetrical faults. Transmission line measurements (Voltage, Current, Frequencies and Impedance) change to abnormal values. Control system is responsible

to communicate the fault condition to control room for an appropriate action by operator or the control logic. Fault events in power transmission sub-subsystem can be further divided into Line to Line (LL) faults, single line to ground (1LG) faults, double line to ground (2LG) faults and three lines to ground (3LG) faults.

Mississippi state university power system test bed implemented using hardware in loop (HIL) and was simulated to generate phase 'A' to ground faults for 1LG , phase 'A' 'B' to ground faults for 2LG faults, phase 'A' 'B' 'C' to ground fault for (3LG) faults, and phase 'A' to 'B' line to line fault for LL faults datasets. Distance and Over Current Protection scheme was configured in test-bed to provide a primary protection up to eighty percent of the line for Zone 1 protection while backup protection for Zone 2 protection of the line. The response time for Zone 1 was set to instantaneous and the response time for the Zone 2 protection was set to 20 cycles.

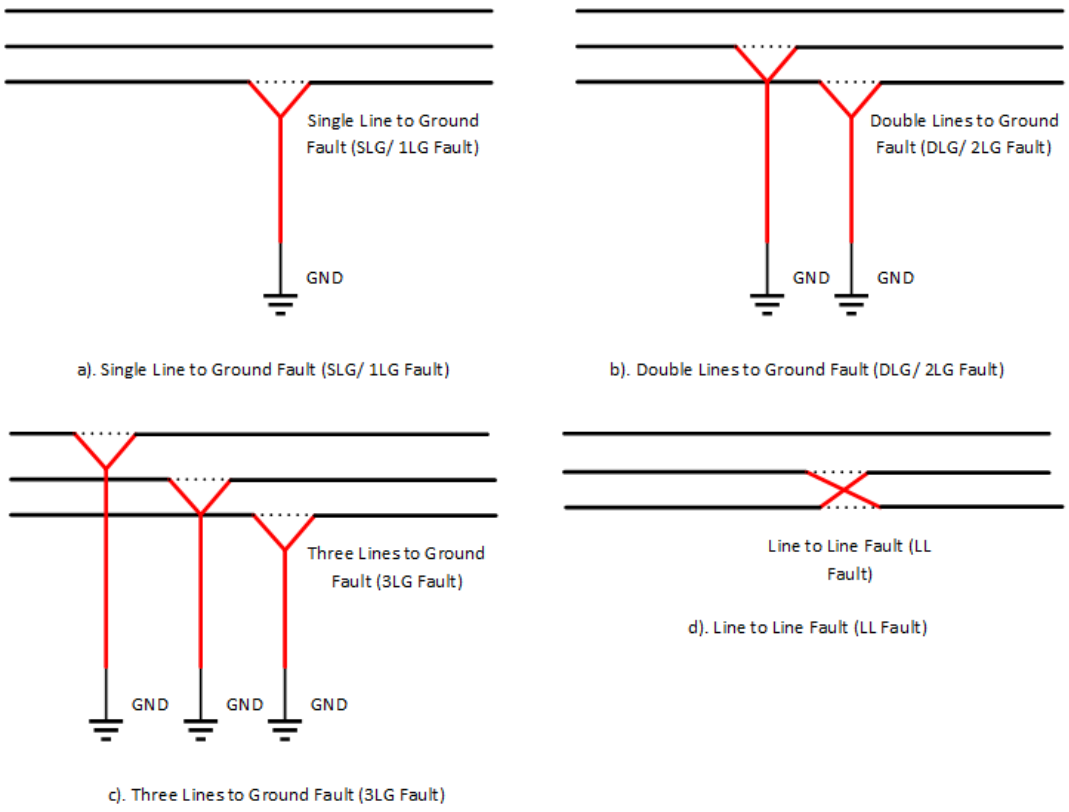


FIGURE 3.1: a). Single Line to Ground Fault (SLG/ 1 LG Fault) b). Double Line to Ground Fault (DLG/ 2LG Fault) c). Three Lines to Ground Fault (3LG Fault) d). Line to Line Fault (LL fault)

3.1.2 Power System Maintenance Events

Transmission line may need the ad-hoc maintenance after a fault or a scheduled one as a contingency plan. Transmission lines need to be disconnected from the grid to avoid any miss-happening. For this purpose support engineer command the relays to operate (open/ close the breakers) remotely at both ends of a line for line maintenance purpose. This control action can also be mimic by attacker.

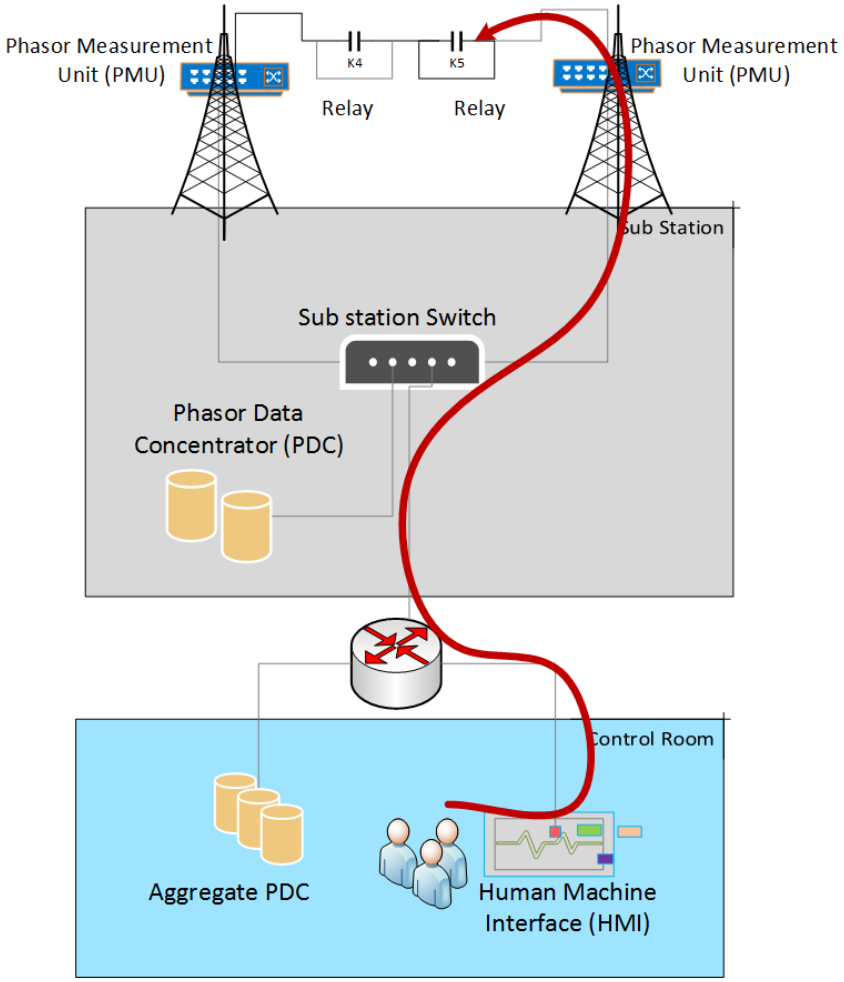


FIGURE 3.2: Line Maintenance Event Scenario

3.2 Power System Cyber-attacks

3.2.1 Remote tripping command injection (Attack)

Command injection attacks can be performed to cause contingencies. Relays can be tripped by injecting the crafted packets or replaying the sniffed/ captured remote relay trip command traffic. Aurora is an example of such attack, repeated command injection attacks are carried on relays to open breakers. This attack methodology was

employed to open breakers at the end of transmission lines L1 and L2 by sending repeated trip commands to relays. Measurements from both attack scenarios are available in the power system datasets.for command injections

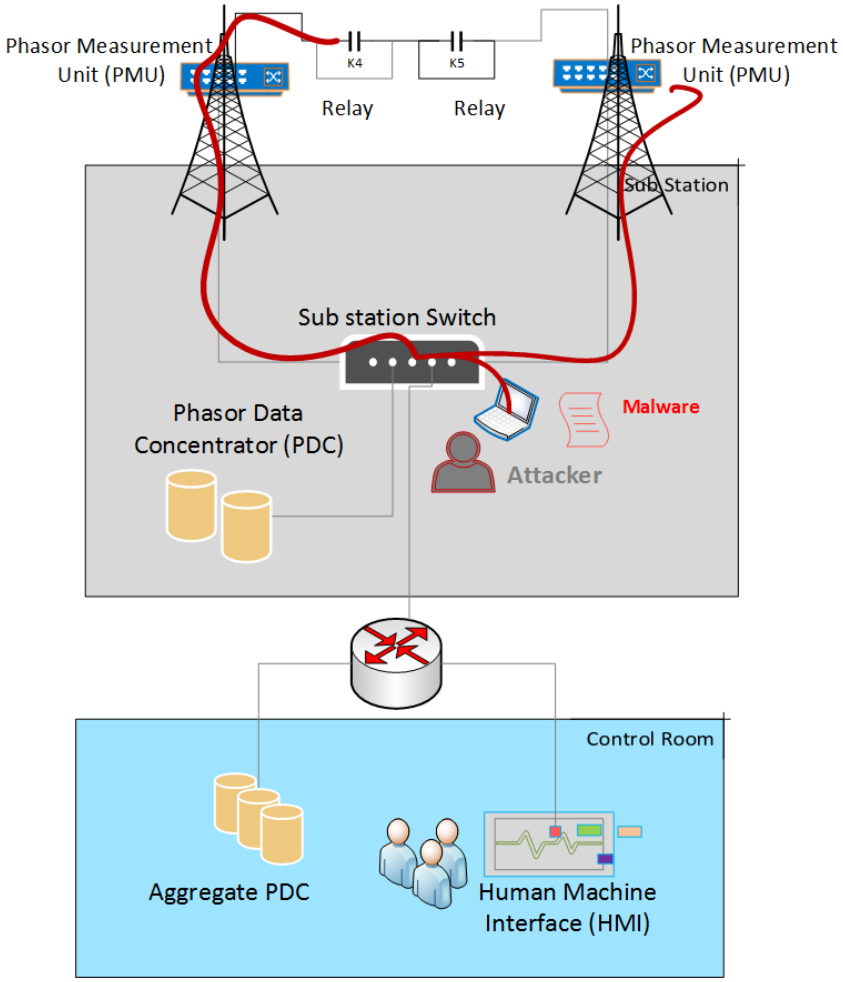


FIGURE 3.3: Command Injection Attack scenario in Power Transmission WAMS

3.2.2 Relay setting change (Attack)

Relay is protective device operates on occurrence of fault in transmission line to initiate the opening of circuit breaker and disconnect the faulty segment of the system from rest of the system. Relays are configured with protection schemes for protecting

transmission lines. Mississippi State University power system dataset have the measurements of the wide area monitoring system (WAMS) under relay setting change attack. Three-bus two-line transmission was simulated with over current protection setting of relays.

Major configurations of relay are that an attacker can change are, tripping threshold, operating time parameters and switch ON/ OFF (Adhikari, 2015). Attacker can perform relay setting change attack in following ways,

- Relay settings change from relay faceplate by gaining the physical access of the relay site.
- A malware that have successfully compromised the control room software can also manipulate control devices by exploiting the remote setting change capability of control roon software

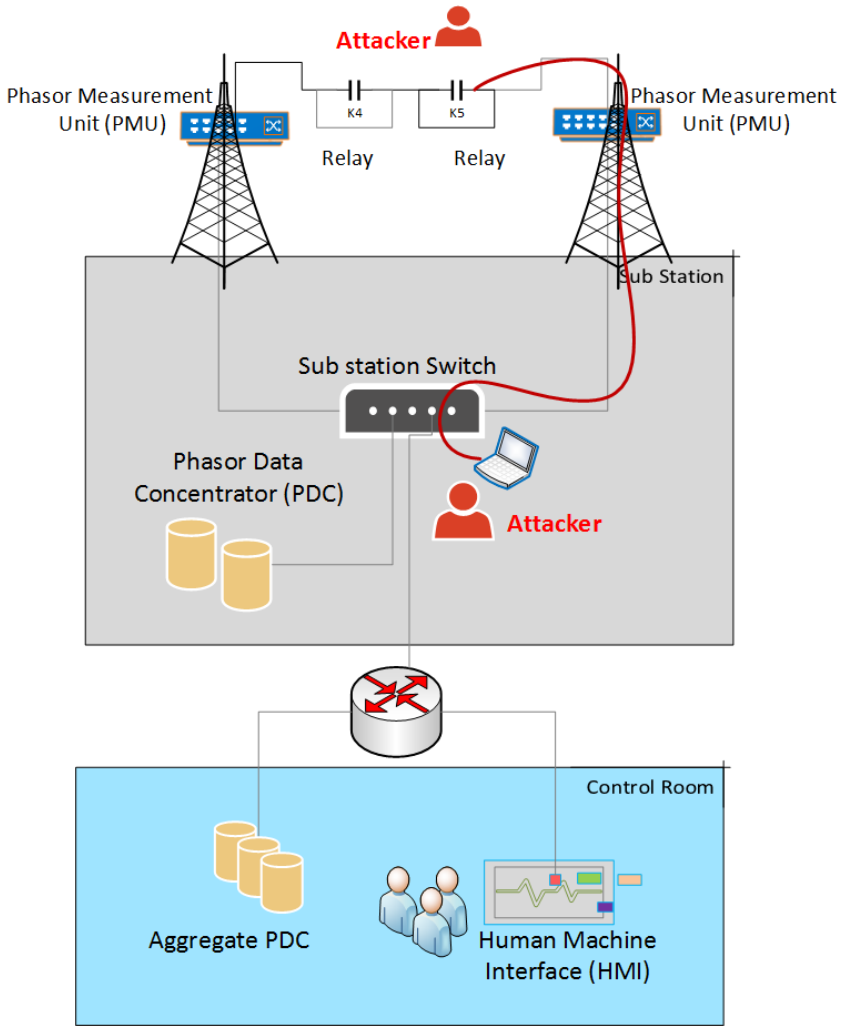


FIGURE 3.4: Relay Setting Change Attack scenario in Power Transmission WAMS

3.2.3 Data Injection (Attack)

False data injection (FDI) attack in electrical power industrial control systems can be disastrous and may cause widespread disturbance to the smart grid. To realize the false data injection (FDI) , attacker exploit the miss-configurations of networking devices, communication protocols, sensor devices, physical security and policies of smart grid cyber physical system. FDI attack of Ukraine power grid in 2015 caused

unavailability of electricity for several hours. Data Injection attack in the cyber physical systems is generally realized with following steps,

- Attacker get the physical access to the sub station switch
- Plugs His/ Her laptop and redirect all the traffic coming from the Intelligent Electronic Device (IED) to pass through His/ Her laptop
- At this stage attacker may have two course of actions,
  - Passive sniffing with data replay - Extract the measurements from traffic of normal time and save for future. When the attacker or malicious code is performing another action of bigger task, recorded measurements are reported to control to hide the activities.
  - Active sniffing with false data injection - Attacker inject the synthetic false data to emulate power system faults.

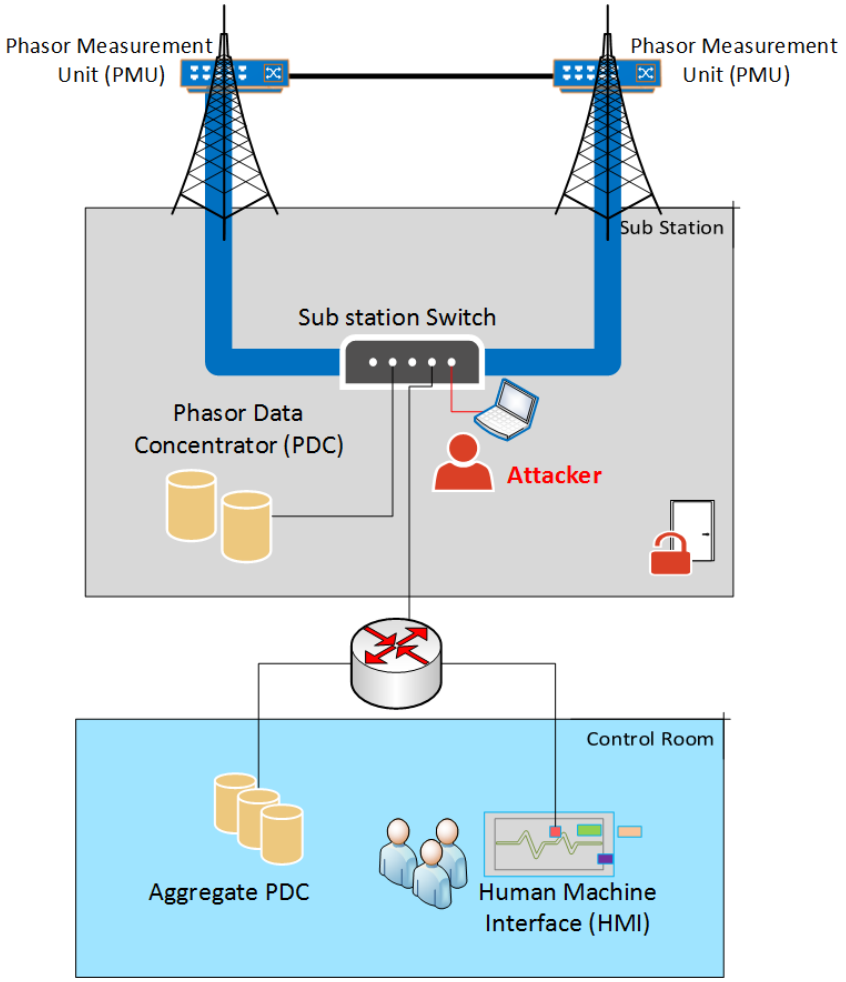


FIGURE 3.5: False Data Injection attack in Power Transmission WAMS

3.3 ORNL and Mississippi State University Power System Datasets)

Power system datasets produced by Mississippi State University and Oak Ridge National Laboratory (ONRL) consists of 15 sets with thirty seven power system event scenarios in each. Eight scenarios are related to natural transmission subsystem

events and twenty eight scenarios related to cyber-attacks on transmission subsystem. Power system test bed simulated using a real-time power system simulator with hardware in loop design using commercial protective relays, PDCs and PMUs. Three-bus two-line transmission system was configured with two zone distance protection scheme. Mississippi State University and Oak Ridge National Laboratory (ONRL) power system datasets are comprehensive and true representative of contemporary power transmission system.

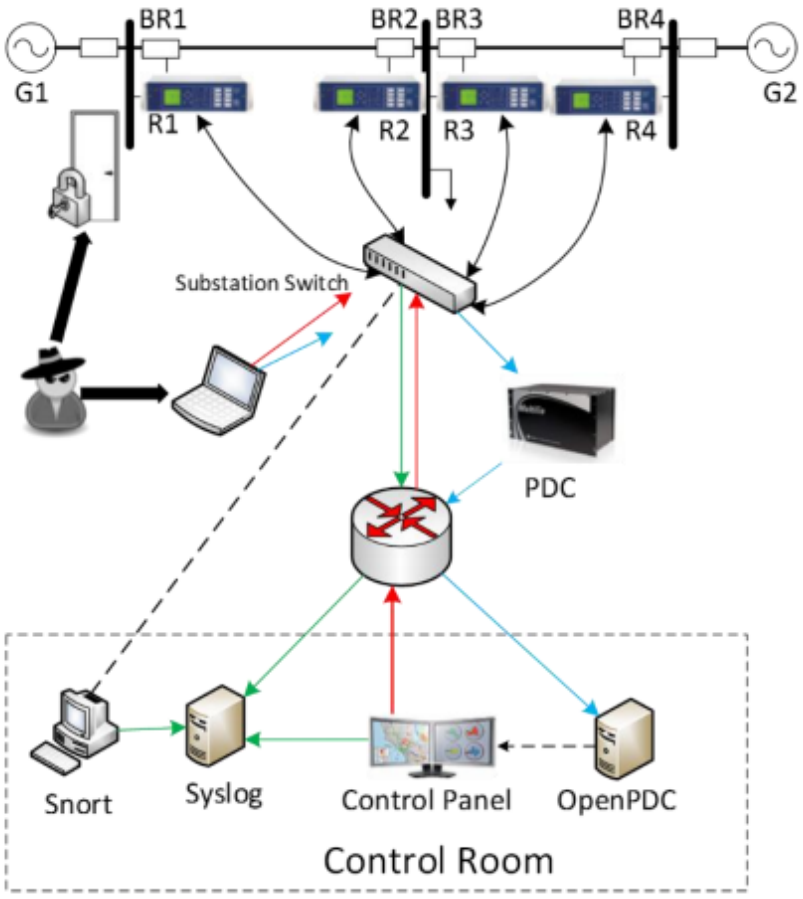


FIGURE 3.6: Power System Testbed used to generate Datasets

The figure 3.6 depicts an abstract design diagram of the testbed used to generating

these scenarios. There are several elements in the network diagram, first of all G1 and G2 are energy generators. R1 through R4 are Intelligent Electronic Devices (IEDs) capable of switching on or off the breakers. The labeling of these breakers is BR1 through BR4. Line One extends from breaker one (BR1) to breaker two and line two from breaker three (BR3) to broker four (BR4). Each IED controls one breaker. BR1 is controlled by R1 while BR2 is controlled by R2 respectively and so on accordingly. The IEDs were configured with distance protection scheme to trips the breaker to identified faults, whether valid or false, as they do not have an inner validation to identify the distinction. Operators can also issue control commands to the IEDs, R1 through R4 to trip the breakers BR1 through BR4 manually. For each phase eight quantities of voltage and current are measured making total of 24 measurements. Frequency, Frequency Delta, Apparent Impendence, Apparent Impendence Angle and Status Flag for each line makes total of 29 electrical quantities measured from a line by phasor measurement units (PMU). There are 4 x PMUs producing total of 116 measurement columns in datasets. Table 3.9 presents a short description of each electrical quantities in datasets along with attributes respective position for each IED. Along with synchrophasors columns, twelve columns are for relays logs, snort alert logs, and control panel logs. The last column is titled 'marker' is actually the label for each fault and attack event scenarios (ref to Table 3.1,Table 3.2,Table 3.3,Table 3.4,Table 3.5,Table 3.6,Table 3.7 and Table 3.8) discussed earlier in this chapter,

- **Short-circuit fault** – A short circuit in a power lines. Please see Figure 3.1
- **Line maintenance** – Relays are operated/ switched ON/ OFF on a specific line for maintenance purpose. Please see Figure 3.2
- **Remote tripping command injection (Attack)** – A malicious command sent by attacker to a relay to causes a breaker to open/ close. Please see Figure 3.3

- **Relay setting change (Attack)** – The attacker changes the settings of a relay or disable the relay function so that relay will not trip for an actual fault or a valid command. Please see Figure 3.4
- **Data Injection (Attack)** – Attacker mimic an actual power system fault by changing current, voltage, sequence components values to blind the operator or the automated response and causes an unintended reponse. Please see Figure 3.5

Label	SLG Faults Events
1	Natural - SLG faults - 10 to 19% on L1
2	Natural - SLG faults - 20 to 79% on L1
3	Natural - SLG faults - 80 to 90% on L1
4	Natural - SLG faults - 10 to 19% on L2
5	Natural - SLG faults - 20 to 79% on L2
6	Natural - SLG faults - 80 to 90% on L2

TABLE 3.1: SLG Fault Events

Label	Line Maintenance Events
13	Natural - Line maintenance - Line L1 maintenance
14	Natural - Line maintenance - Line L2 maintenance

TABLE 3.2: Line Maintenance Events

Label	Data Injection Attack - SLG Fault Replay Events
7	Fault from 10 to 19% - L1 with trip command
8	Fault from 20 to 79% - L1 with trip command
9	Fault from 80 to 90% - L1 with trip command
10	Fault from 10 to 19% - L2 with trip command
11	Fault from 20 to 79% - L2 with trip command
12	Fault from 80 to 90% - L2 with trip command

TABLE 3.3: Data Injection Attack - SLG Fault Replay Events

Label	Command Injection Attack - Single Relay Remote Trip Events
15	Command Injection to Relay1
16	Command Injection to Relay2
17	Command Injection to Relay3
18	Command Injection to Relay4

TABLE 3.4: Command Injection Attack - Single Relay Remote Trip Events

Label	Command Injection Attack - Two Relay Remote Trip Events
19	Command Injection to R1 and R2
20	Command Injection to R3 and R4

TABLE 3.5: Command Injection Attack - Two Relay Remote Trip Events

Label	Relay Setting Change Attack - Two Relay Disable and Line Maintenance
39	L1 maintenance with R1 and R2 disabled
40	L2 maintenance with R3 and R4 disabled

TABLE 3.8: Relay Setting Change Attack - Two Relay Disable and Line Maintenance

Label	Relay Setting Change Attack - Single Relay Disable and Fault
21	Fault from 10 to 19% - L1 with R1 disabled & fault
22	Fault from 20 to 90% - L1 with R1 disabled & fault
23	Fault from 10 to 49% - L1 with R2 disabled & fault
24	Fault from 50 to 79% - L1 with R2 disabled & fault
25	Fault from 80 to 90% - L1 with R2 disabled & fault
26	Fault from 10 to 19% - L2 with R3 disabled & fault
27	Fault from 20 to 49% - L2 with R3 disabled & fault
28	Fault from 50 to 90% - L2 with R3 disabled & fault
29	Fault from 10 to 79% - L2 with R4 disabled & fault
30	Fault from 80 to 90% - L2 with R4 disabled & fault

TABLE 3.6: Relay Setting Change Attack - Single Relay Disable and Fault

Label	Relay Setting Change Attack - Two Relay Disable and Fault
35	Fault from 10-49% on L1 with R1 and R2 disabled & fault
36	Fault from 50-90% on L1 with R1 and R2 disabled & fault
37	Fault from 10-49% on L1 with R3 and R4 disabled & fault
38	Fault from 50-90% on L1 with R3 and R4 disabled & fault

TABLE 3.7: Relay Setting Change Attack - Two Relay Disable and Fault

Eletrical Quantity		R1	R2	R3	R4
Voltage Phase Angle	For phase A	0	29	58	87
	For phase B	2	31	60	89
	For phase C	4	33	62	91
Current Phase Angle	For Phase A	6	35	64	93
	For Phase B	8	37	66	95
	For Phase C	10	39	68	97
Pos. – Neg. – Zero Voltage Phase Angle	For Phase A	12	41	70	99
	For Phase B	14	43	72	101
	For Phase C	16	45	74	103
Pos. – Neg. – Zero Current Phase Angle	For Phase A	10	47	76	105
	For Phase B	11	49	78	107
	For Phase C	12	51	80	109
Appearance Impedance Angle for relays	ZH	27	56	85	116
Frequency for relays	F	24	53	82	111
Frequency Delta (dF/dt) for relays	DF	25	54	83	112
Status Flag for relays	S	28	57	86	115
Appearance Impedance for relays	Z	26	55	84	113
Voltage Phase Magnitude	For Phase A	1	30	59	88
	For Phase B	3	32	61	90
	For Phase C	5	34	63	92
Current Phase Magnitude	For Phase A	7	36	65	94
	For Phase B	9	38	67	96
	For Phase C	11	40	69	98
Pos. – Neg. – Zero Current Phase Magnitude	For Phase A	13	42	71	100
	For Phase B	15	44	73	102
	For Phase C	17	46	75	104
Pos. – Neg. – Zero Voltage Phase Magnitude	For Phase A	19	48	77	106
	For Phase B	21	50	79	108
	For Phase C	23	52	81	110

TABLE 3.9: Electrical quantities measured by each Intelligent electronic device (IED) in a reference 2-Bus 2-Generator power transmission system deployed with four IEDs

Chapter 4

Flare - A lightweight CP-IDS based on Gaussian Mixture Modeling and Correntropy-induced metric (CIM)

Flare is a plug-and-play CP-IDS for WAMS based electric power transmission system. Flare's core, is an un-supervised anomaly based decision engine (DE), which achieves optimal accuracy in discriminating the normal vs attack time measurements of WAMS. It has small memory footprint and is capable to process high speed WAMS data at real-time. Flare can be deployed centrally or on a single IED, operational viability of both deployment scenarios is demonstrated in Chapter 5, by evaluating both versions of core algorithm.

For central deployment flare shall process the measurements of multiple IEDs, while, in distributed or sub-station deployment, it need to process the measurements of a single IED. Flare can be embed directly into IED or can be deployed on a separate commodity hardware such as raspberry pi to detect attack against that IED.

This chapter give an overview of central limit theorm, properties of normal distribution and probability density function. Recap of related concepts and theories, employed in design and developement of Flare, considered vital and presented in section 4.4 .

- **See section 4.1** for WAMS Data Charateristics and Analysis
- **See section 4.2** for Probability Distribution (PD) of Power System Electrical Quantities
- **See section 4.3** Gaussian Function and Gaussian/ Normal Distribution
- **See section 4.4** System Design - Development of Flare
- **See section 4.5** Flare - Learning and Execution

4.1 WAMS Data Characteristics and Analysis

4.1.1 Central Limit Theorem and centrality of electrical quantities in WAMS

Central limit theorem – Evolution to present day form of the 'central limit theorem' can be credited to four prominent mathematicians Abraham de Moivre, Pierre-Simon Laplace, Carl Friedrich Gauss and finally George Polya.

Central limit theorem states,

> *"If $\overline{X}$ is the mean of a random sample of size n drawn from a population with mean μ and variance σ^2, the sampling distribution of $\overline{X}$ approaches the gaussian distribution with mean μ and variance σ^2/n as the sample size n increases without limit"*

Central limit theorem tells that whatever be the form of the parent population, the sampling distribution of mean is approximately normal provided n is sufficiently large. Figure 4.1 depicts the frequency histograms of sampling means, of 29 x synchrophasors quantities, which, clearly tends to be normally distritbuted. In a nutshell central limit theorem enables us to approximate the sampling distribution of means with a normal distribution and one of the most important concepts of statistics and probability.

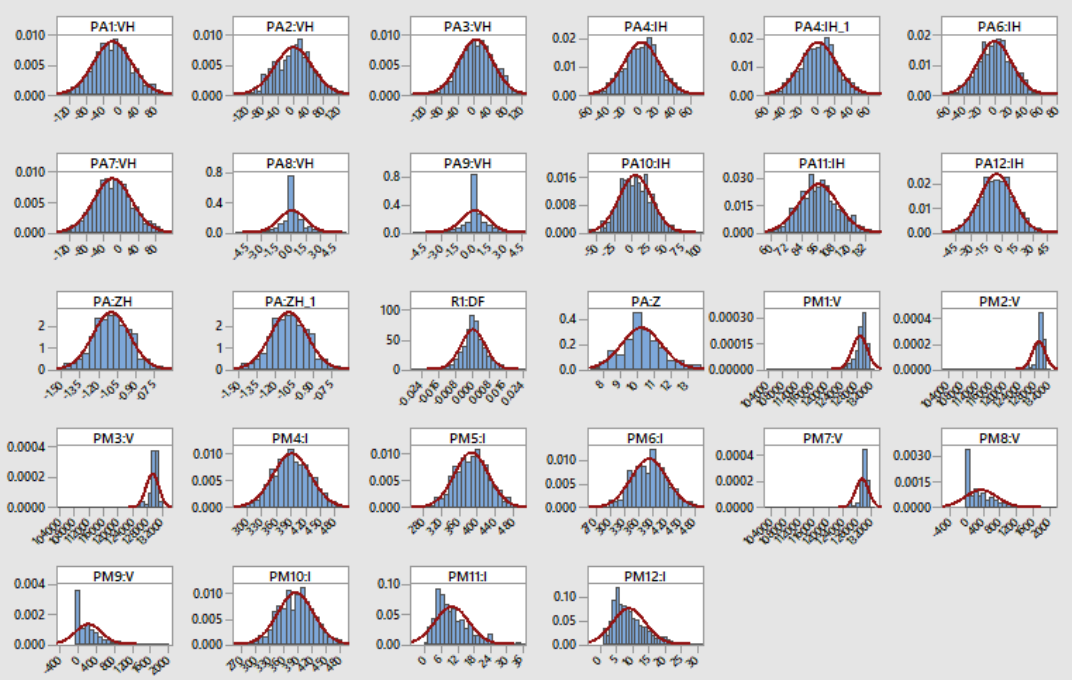


FIGURE 4.1: Centrality of electrical quantities of WAMS

4.1.2 Empirical evidences of synchrophasors' Normality

Normality test results helps in rejecting or failing to reject the null hypothesis that the data come from a normally distributed population. Table 4.1 shows the results of well known Anderson-Darling, Ryan-Joiner and Kolmogorov-Smirnov normality test.

Anderson-Darling Test This test uses the ECDF (empirical cumulative distribution function) with the distribution expected if the data under test is normal. Test yield a number, which helps to accept or reject the null hypothesis of population normality.

Ryan-Joiner Normality Test This test is based on correlation between test data and the normal scores of test data. If the correlation coefficient is near 1, the population is likely to be normal. This test is similar to the Shapiro-Wilk normality test.

Kolmogorov-Smirnov Normality Test This test is also based on ECDF (empirical cumulative distribution function) of test sample data with the distribution expected if the data were normal. Calculated difference is adequately large, the test will reject the null hypothesis of population normality. If the p-value of this test is less than chosen α, it may lead to reject or accept null hypothesis of population normality.

Significance level α of 0.05 is chosen for all normality tests. Significance level of 0.05 means, risk of concluding the data do not follow a normal distribution-when, actually, the data do follow a normal distribution-is 5%.

- P-value $\leq \alpha$: The data do not follow a normal distribution
- P-value $> \alpha$: Cannot conclude the data do not follow a normal distribution

4.2 Probability Distribution (PD) of Power System Electrical Quantities

Expressing the PMU's measured quantities as random variables as $X = [x_1, x_2, x_3 ..., x_m]$, where m is the number of features that is 128. As possible outcomes of a continuous

Ser	Synchrophasor	Anderson-Darling	Ryan-Joiner	Kolmogorov-Smirnov
		P-Value	*P-Value*	*P-Value*
1	**PA1:VH**	0.616	>0.100	>0.150
2	**PA2:VH**	0.087	>0.100	0.087
3	**PA3:VH**	0.273	0.024	>0.150
4	**PA4:IH**	0.253	>0.100	>0.150
5	**PA5:IH**	0.253	>0.100	>0.150
6	**PA6:IH**	0.79	>0.100	>0.150
7	**PA7:VH**	0.614	>0.100	>0.150
8	**PA8:VH**	<0.005	<0.010	<0.010
9	**PA9:VH**	<0.005	<0.010	<0.010
10	**PA10:IH**	0.034	0.048	0.031
11	**PA11:IH**	0.006	0.021	0.025
12	**PA12:IH**	0.724	>0.100	>0.150
13	**PA:ZH**	0.483	>0.100	>0.150
14	**R1:DF**	<0.005	<0.010	<0.010
15	**PA:Z**	0.353	>0.100	>0.150
16	**PM1:V**	<0.005	<0.010	<0.010
17	**PM2:V**	<0.005	<0.010	<0.010
18	**PM3:V**	<0.005	<0.010	<0.010
19	**PM4:I**	0.49	>0.100	>0.150
20	**PM5:I**	0.261	>0.100	0.072
21	**PM6:I**	0.32	>0.100	>0.150
22	**PM7:V**	<0.005	<0.010	<0.010
23	**PM8:V**	<0.005	<0.010	<0.010
24	**PM9:V**	<0.005	<0.010	<0.010
25	**PM10:I**	0.353	>0.100	>0.150
26	**PM11:I**	<0.005	<0.010	<0.010
27	**PM12:I**	<0.005	<0.010	<0.010

TABLE 4.1: Normality tests of WAMS electrical quantities

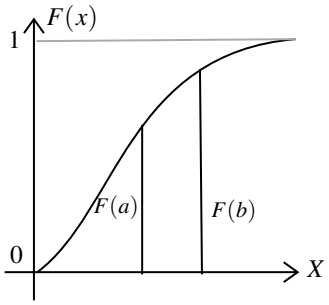
random variables are uncountable, it is difficult to write down the probabilities of all possible events. Therefore, the PDF is a function which gives the probability of one event.

A random variable X is defined to be continuous if it can assume every possible value in an interval [a,b], a<b, where a and b may be -∞ and +∞ respectively. A r.v. X may also be defined as continuous if its distribution function (d.f.) $F(x)$ has no jumps or steps but is a continuous function for all x. Let the derivative of $F(x)$ be denoted by $f(x)$, i.e.

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x) \tag{4.1}$$

Since $F(x)$ is a non-decreasing function of x, we have The function $f(x)$ is called

$$\begin{cases} \textbf{i).} f(x) \geq 0 \\ \\ \textbf{ii).} F(x) \;=\; \int_{-\infty}^{\infty} f(x)dx \;, \textit{ for all x.} \end{cases}$$



the probability density function, abbreviated to p.d.f., or simply density function of the r.v. X. A p.d.f. has the following properties:

4.3 Gaussian Function and Gaussian/ Normal Distribution

Gaussian function is an exponential function with a concave quadratic function and is used to describe the gaussian distributions. In other words, gaussian function is

i). $f(x) \geq 0$, for all x

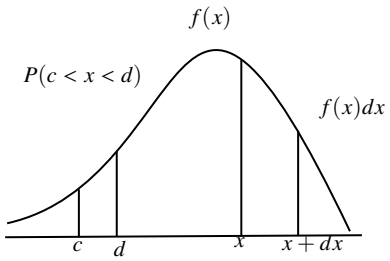
ii). $\int_{-\infty}^{\infty} f(x)\,dx = 1$

iii). *The probability that X takes on a value in the interval [c,d] , c< d is given by*

$$P(c < x \geq d) = F(d) - F(c)$$

$$= \int_{-\infty}^{d} f(x)dx - \int_{-\infty}^{c} f(x)dx$$

$$= \int_{c}^{d} f(x)dx \text{ ; which is the area under curve y=f(x) between X=c and X=d}$$



the probability density function (p.d.f.) of a gaussian distribution.

A gaussian distribution is defined by the p.d.f.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[(x-\mu)^2/\,2\,\sigma^2\right]}\ ,\ for\ -\infty < x < \infty,\ and\ \sigma > 0 \qquad (4.2)$$

- μ is the mean.
- σ is the standard deviation.
- π is a constant, approx. 3.141
- e is a constant, approx. 2.7183

As obvious from the equation 4.3 a gaussian distribution is characterized by two parameters μ and σ, its mean and standard deviation.

4.3.1 Properties of Gaussian/ Normal Distribution

1. The function $f(x)$ defining the gaussian normal distribution is aproper p.d.f. i.e f(x) $\geq$ 0 and the total area under the curve is unity.
2. The mean and variance of the gaussian distribution are μ and σ respectively.
3. The median and mode of the gaussian distribution are each equal to μ, the mean of the distribution.
4. The mean distribution of the gaussian distribution is approximately $\frac{4}{5}$ of its standard deviation.
5. The gaussian curve has points of inflection which are equidistant from the mean.
6. For the gaussian distribution, the odd order moments about the mean are all zero and the even order moments are given by

$$\mu_{2n} = (2n-1)(2n-3)...\ 5.3.1\ \ \sigma^{2n} \tag{4.3}$$

7. if X is $N(\mu, \sigma^2)$ and if $Y = abX$, then Y is $N(a + b\mu, b^2\sigma^2)$
8. Sum or addition of autonomous gaussian variables yields a gaussian/ normal variable.
9. Irrespective of values of μ and σ , areas under normal curve shall remain fixed in certain proportions within a defined number of standard deviations on both side of μ.

10. The Quartile deviation, Q, is found as

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu-Q}^{\mu+Q} e^{-\left[(x-\mu)^2/\ 2\ \sigma^2\right]}\,dx\ =\ \frac{1}{2} \tag{4.4}$$

4.4 System Design - Development of Flare

This section presents the theories behind the proposed cyber-physical intrusion detection system.

Flare starts processing the measurements as it is plugged into the production system. Flare process a single instance of measurements received, this is simulated by streaming the PMU measurements stored in a CSV file as it coming from a real PMU using powerful python object iterator, $next()$, an iterator is an object which implements the iterator protocol, return the next item from the container (Van Rossum and Drake, 1995). Flare process algorithm 1 aggregate the measurement instances before the training grace period ends. As the training grace period finishes, data-frame is passed to training sub-section, where upper correntropy and lower correntropy is computed as explained in section 4.4.2.

Algorithm 1 Flare Measurements Instance Processing Algorithm

```
1: instance_window=120
2: ex_counter
3: trg_counter
4: n_trained                                   ▷ The number of instances so far
5: training_grace_period
6: procedure FLARE_PROCESS(X) ▷ Live data from PMU, X, X = [x1, x2, x3 ..., xm]
7:     if n_trained ≥ training_grace_period then
8:         ex_counter += 1
9:         if ex_counter ≤ window then
10:            Ex_Mat.append(X)
11:        else if ex_counter = window then
12:            ex_counter=0
13:            return execute(Ex_Mat)
14:        else
15:            n_trained
16:    else
17:        trg_counter += 1
18:        if trg_counter < training_grace_period then
19:            train_Mat.append(x)
20:        else if trg_counter = training_grace_period then
21:            n_trained += 1
22:            return train(Train_Mat)
23:    n_trained += 1
```

Central limit theorem (CLT) played a central role in development of Flare's core algorithm. Gaussian probability density estimation is employed for efficient modeling of multivariate electrical data of power transmission system WAMS. A high level diagram is shown in figure 4.2. Anomaly based DE, developed as an outcome of this research is based on modeling the benign behavior of the WAMS based power transmission system using mixture of gaussains and Correntropy-induced metric (CIM). CIM or correlation entropy is a concept from information theory Liu, Pokharel, and Príncipe

explained its characteristics and implementation for non-Gaussian signal processing.(Liu, Pokharel, and Príncipe, 2007)

Electrical measurements from the power transmission system are expressed as random variables and each variable's probability distribution is modeled using gaussian probability density function (PDF). To fuse the multivariate distributions of the synchrophasor quantities into one variable concept of mixture modeling employed in order to specify the exact boundaries of normal data in real-time processing. Secondly, a Corrector method is used to design an autonomous baseline that adaptably identifies the precise bounds of legitimate patterns and consider any variations outside this baseline as anomalous(Haider, 2018).

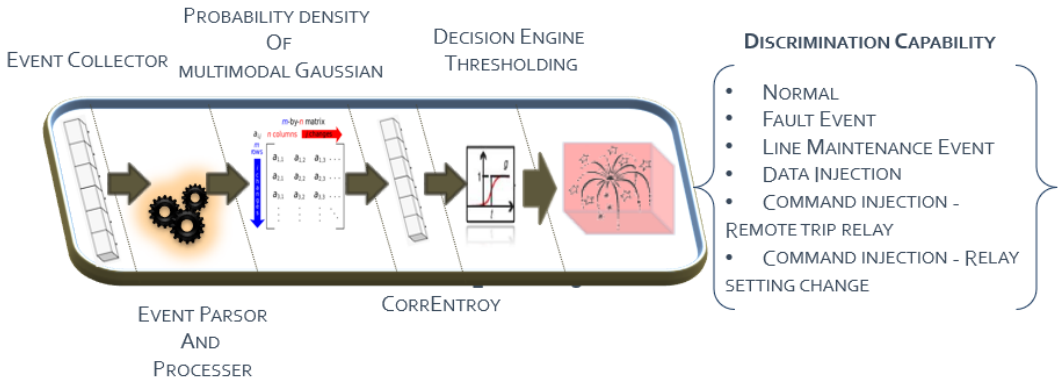


Figure 4.2: Proposed DE based on Gaussian Mixture Modeling and Correlation Entropy(Corr-GMM)

4.4.1 Gaussian Mixture Modeling (GMM)

Flare DE models the probability distributions of each feature variable using gaussain distribution. Expressing the PMU's measured quantities as random variables as $X = [x_1, x_2, x_3 ..., x_m]$, where m is the number of features. Each feature variable (x_j), where $(j = 1, ..., m)$, can be modeled by a Gaussian Distribution (GD) (Haider, 2018). The PMU's quantities comprise n measurements, where the *ith* reading from the m-dimensional X variable space is expressed as $X_i = [x_{i1}, ..., x_{ij}, ..., x_{im}]$ and its

p.d.f. is estimated by:

$$p(x_j|\mu_j,\sigma_j^2) = \frac{1}{\sigma_j\sqrt{2\pi}} e^{-\left[(x_j-\mu_j)^2/\ 2\ \sigma_j^2\right]} \tag{4.5}$$

where μ_j and σ_j^2 denote the mean and variance of variable x_j along the measurement axis or precisely along the row number of the record matrix, respectively.

Considering the multivariate data from the power system WAMS, all the variables are estimated using p.d.f. in order to best fit the normal data and detect deviations as abnormal.A GMM for multiple variables is a combination of m components, where m denotes m relevant features, with the posterior probability of taking the i_th measurement value (X_i) formulated by:

$$p(X_i|\theta_i) = \sum_{k=1}^{m} \omega_k p(x_{ik}|\theta_k) \tag{4.6}$$

where $(\omega_1,....,\omega_m)$ are the probabilistic proportions/weights of the components and θ_k includes μ_k and σ_k. $p(x_ik|\theta_k)$ refers to the probability of taking the value of the i_th sample by the feature variable (x_j) which can be computed by equation 4.5. posterior probabilities of the measurements, that is, supposing that there are n sequences, the feature variables in which are identically and independently distributed (iid), the posterior probabilities of all the whole measurements can be estimated by:

$$p(X|\theta) = \prod_{i=1}^{n} \sum_{k=1}^{m} \omega_k p(x_{ik}|\theta_k) \tag{4.7}$$

The parameters of the GMM, i.e., the mean (μ_k), variance (σ^2) and weight (ω_k), for all the measurements (X) are estimated using the following three equations.

$$\mu_k = \frac{\sum_{i=1}^{n} p(x_{ik}|\psi).y_{ik}}{\sum_{i=1}^{n} p(x_{ik}|\psi)} \tag{4.8}$$

$$\sigma_k^2 = \frac{\sum_{i=1}^{n} p(x_{ik}|\psi).y_{ik}^2}{\sum_{i=1}^{n} p(x_{ik}|\psi)} - \mu_k^2 \tag{4.9}$$

$$\omega_k = \frac{1}{m}\sum_{i=1}^{n} p(x_{ik}|\psi) \tag{4.10}$$

where ψ denotes the data collected during the training grace period and used to successfully build the GMM model. During training period, the GMM parameters (μ,σ and ω) are estimated from the normal data using equations 4.8, 4.9 and eq:9 by letting $X_{normal} = x_1, ..., x_n$, where X_{normal} is a set of normal measurements. Parameters obtained are used to compute the posterior probabilities $(P_posterior)^normal$ from the normal measurements using equations (4.5 and (4.7) with the training profile containing the GMM measures $(\mu, \sigma, \omega, P_{posterior})^normal$

Probabilities based corelation entropy can more accurately estimate the dependency of the prior and likelihood distributions of the observations (Haider, 2018).

4.4.2 Correntropy-induced metric (CIM)

Correntropy, is employed to measure the local similarity between any two random variables, (Haider, 2018) introduced the notion of correntropy of probabilities, this theory is utilized here for probabilities obtained from equation (4.7). The interdependency of PMU's measurements can be more reliably estimated based on their

distribution of prior and likelihood data, i.e., posterior = likelihood x prior. The Correntropy of the two consecutive observations $A = p(Yi|\theta)$ and $B = p(Yi+1|\theta)$ in terms of posterior probabilities is defined as:

$$V_\sigma(A,B) = E[k_\sigma(A-B)] \tag{4.11}$$

where k_σ is a kernal function given by,

$$g(X) = \exp(-\frac{X^2}{2\sigma^2}) \tag{4.12}$$

Correntropy-induced metric (CIM) employyed by (Haider, 2018) to measure the local similarity between to measurements, for any two observations (A and B) as:

$$Corr(A,B) = (g(0) - \frac{1}{n}\sum_{i=1}^{n} g(A-B))^{\frac{1}{2}} \tag{4.13}$$

Correntropy for normal data $Corr^normal$ is estimated from $(P_posterior)_normal$ using equation (4.13), a measure that can more accurately specify the lower and upper boundaries of a normal profile which were computed using equations (4.14) and (4.15), respectively.

$$lower(Corr^{normal}) = min(Corr^{normal}) \tag{4.14}$$

$$upper(Corr^{normal}) = max(Corr^{normal}) \tag{4.15}$$

4.5 Flare - Learning and Execution

Flare, decision engine (DE), starts learning as it is plugged into the live/ production system. Flare assumes that all measurements are benign when it is plugged into the system till train grace period ends. Therefore, an attack already in its execution phase may be able to evade Flare's detection.

4.5.1 Flare Learning

Flare collect and store the data from the live system and estimate the parameters θ (μ, σ and ω) using the equations 4.8, 4.9 and 4.10 as explained in learning algorithm 2. With these parameters (μ, σ and ω) in hand, probability density function is estimated with 4.6. Inter-feature relation is captured by operating the 4.7, to combine the effect of variables. Corr-entropy measure is employed to operate on mixture of posterior probabilities rather than raw value to find the upper corr-entropy and lower corr-entropy. Two value $LowerCorr$ and $UpperCorr$ become the decision thresholds for execution time. These two values are calculated from benign measurements so become the numerical representatives of normal measurements. After calculating the threshold all the memory is made free.

Algorithm 2 Flare Training algorithm

1: **procedure** TRAIN($Train\_Mat$)
2: $\mu \leftarrow Train\_Mat$ ▷ For each column seperately
3: $\sigma \leftarrow Train\_Mat$ ▷ For each column seperately
4: $\omega \leftarrow Train\_Mat$ ▷ For each column seperately
5: $var \leftarrow Train\_Mat$ ▷ For each column seperately
6: **for** `X in Train_Mat` **do**
7: Compute $Posterior^{normal}$ using equation 4.5, 4.6 *and* 4.7
8: Compute $lower(Corr^{normal})$ and $upper(Corr^{normal}$ using equations 4.13, 4.14 *and* 4.15
9: **return** $upper(Corr^{\text{normal}})$ *and* $lower(Corr^{\text{normal}}$ ▷ The Normal Profile

4.5.2 Flare Executing

Execution mode starts with accumulating the number of measurements, specified by user (through *window* parameter) as per the configurations of the PMU. If the PMU is set to sense and transmit the measurements at rate of 120 measurements per second, Flare's *window* parameter will be set to 120. If the PMU is set to operate on 80 measurements per second, *window* parameter will be set to 80. During execution time Flare, accumulates the number of measurements specified in *window* variable and estimate the probabilities and correntropy measure as explained in the algorithm **3**.

Finally, this corr-entropy measure is checked against the decision boundaries set during training using $LowerCorr^{normal}$ and $UpperCorr^{normal}$. If the *corr* measure of the measurements received at point time is greater than *LowerCorr* and $UpperCorr$, it is considered normal otherwise a cyber-attack. After that memory holding the data is made free to hold the measurement of next time window. The instance is immediately discarded after execution of the model with an instance, as done in stream clustering.

Algorithm 3 Flare Execution algorithm

```
1: procedure EXECUTE(Ex_Mat)
2:     for Y in Ex_Mat do
3:         Estimate Posterior^execute                ▷ using equation 4.5, 4.6 and 4.7
4:         Estimate (Corr^execute)                   ▷ using equations 4.13, 4.14 and 4.15
5:         if (Corr^execute) ≥ lower(Corr^normal) or (Corr^execute) ≤ upper(Corr^normal) then
6:             return normal
7:         else
8:             return attack
```

Chapter 5

Evaluation and Results

This chapter aims to demonstrate the capacity and capability of the Flare to act as a practical anomaly detection system in a WAMS based power transmission system. To evaluate the performance of Flare, and establish the operational viability of Flare, independent experimentation performed for both central and distributed deployment.

This chapter presents the criteria used for evaluation of Flare, datasets used for experimentation and empirical results for both deployment strategies. Section 5.2 describe the experimentation and results obtained using Raspberry PI 3B for distributed deployment while Section 5.3 report the results obtained using Intel Core i7 CPU, 2.93 GHz with 8GM RAM for central deployment.

- **See section 5.1** for Performance Evaluation Method
- **See section 5.2** for EXPERIMENTATION-I
- **See section 5.3** for EXPERIMENTATION-II
- **See section 5.4** Comparative Analysis - WAMS based Intrusion Detection Systems

5.1 Performance Evaluation Method

Flare's performance reliability is assessed using three major metrics, the Detection Rate (DR), False Alarm Rate (FAR) and processing time (Haider, 2018), with the DR and FAR defined, respectively, as:

$$DR = \frac{Numberof DetectedAttacks}{NumberofAttacksPresent} \text{x}100\% \tag{5.1}$$

$$FAR = \frac{NumberofFalseAlerts}{NumberofTraces/SequencesinValidationdata} \text{x}100\% \tag{5.2}$$

The DR, FAR represents the accuracy and error of Flare, respectively. The FAR is the average of two errors, the False Positive Rate (FPR) and False Negative Rate (FNR). FAR can well indicate the capacity of Flare to incorrectly detect normal behavior as abnormal and abnormal behavior as normal, respectively (Haider, 2018). Measures of the processing times with respect to training and testing using the experimental datasets are also provided to fully assess the reliability of Flare.

5.2 EXPERIMENTATION-I - Sub-Station Deployment

5.2.1 Experiments and Results For IED Deployment

First experiment was performed to benchmark Flare's capacity to deploy at substation on a commodity hardware or imbed in Intelligent Electronic Device (IED). Flare achieved the satisfactory performance reliability along with lower memory and computation complexity. Table **??** shows the time consumed in seconds for training and execution phase on raspberry pi, which is near real-time. Perfectly real-time operations of Flare can be achieved with implementation in C++. Training and execution

times reported in Table **??**re recorded against the datasets described earlier in this chapter. ONRL and Mississippi state university power system datasets were used to create the datasets as described in Table **??**. Magnitude quantities were used to create the features with a combine effect of respective quantity of each phase (as shown in Table 5.1). Selection of features yeild a significant drop in processing time.

Ser	**Feature Name**
1	Averaged Phase A - C Voltage Phase Magnitude
2	Averaged Phase A - C Current Phase Magnitude
3	Averaged Pos. – Neg. – Zero Voltage Phase Magnitude
4	Averaged Pos. – Neg. – Zero Current Phase Magnitude
5	Frequency for relay
6	Frequency Delta (dF/dt) for relay
7	Appearance Impedance for relay
8	Appearance Impedance Angle for relay

TABLE 5.1: Averaged Magnitude Quantities measured by PMU

Power System Natural Event / Cyber Attack	**Nomral Training Instances**	**Normal Test Instances**	**Attack Test Instances**
Command Injection	3524	879	8739
Data Injection	3524	879	9582
SLG Fault	3524	879	14243
Line Maint	3524	879	3041
Relay Setting Change	3524	879	36159

TABLE 5.2: Datasets created for experimnet 1 evaluations

Measure	False Positive Rate	Error Rate	False Negative Rate
Comd. Inj.	0	0.75	1.51
Data Inj.	9.2	0.75	4.6
SLG Fault	0	1.62	3.25
Line Maint.	0	1.62	3.25
Relay Setting Change	0	4.6	9.2

TABLE 5.5: False Positive Rate, Error Rate and False Negative Rate measured on Raspberry Pi

Power System Natural Event / Cyber Attack	**Taining Time (Seconds)**	**Execute Time (Seconds)**
Command Injection	43	137
Data Injection	43	127
SLG Fault	43	192
Line Maint	43	44
Relay Setting Change	43	142

TABLE 5.3: Training and Execution time for Raspberry Pi

Attacks	**Command Injection**	**Data Injection**	**SLG Fault**	**Line Maint**	**Relay Setting Change**
Detection Rate	98.48	90.79	96.74	96.74	90.72

TABLE 5.4: Detection Rate against attacks using Raspberry Pi

5.3 EXPERIMENTATION-II - Central Deployment

5.3.1 Experiments and Results For Central Deployment

These experiments were conducted to benchmark Flare's capacity for central deployment. Central deployment is aimed at collecting the measurements from the

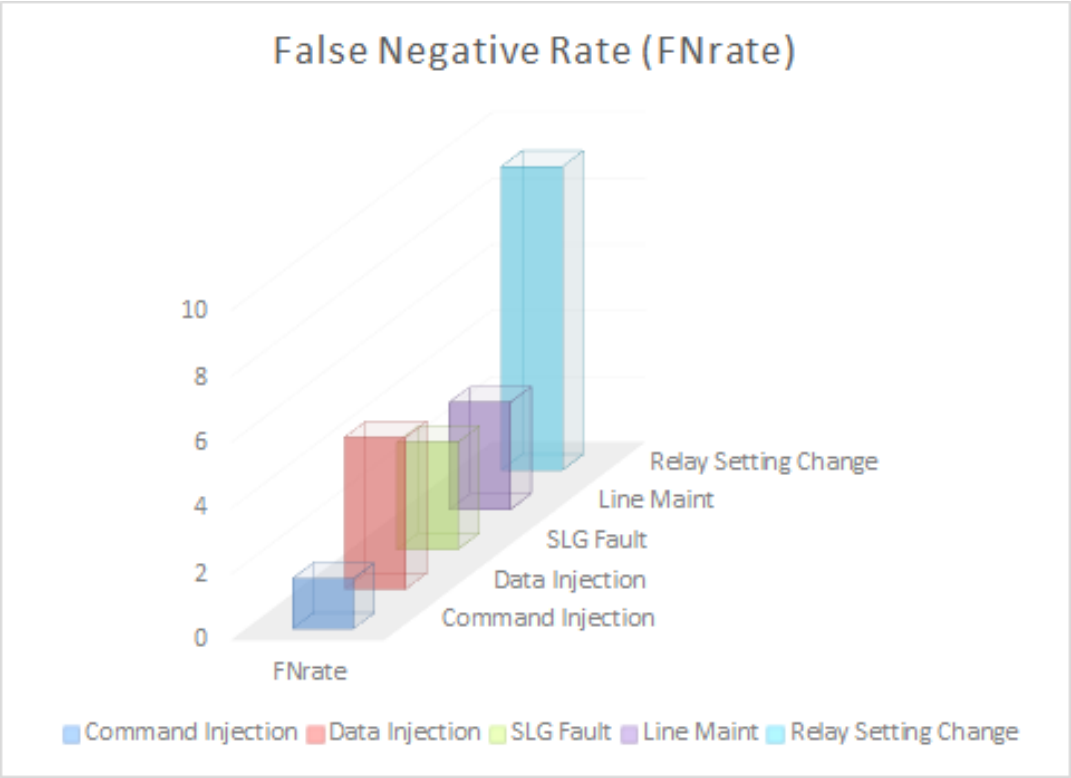


FIGURE 5.1: False Negative Rate

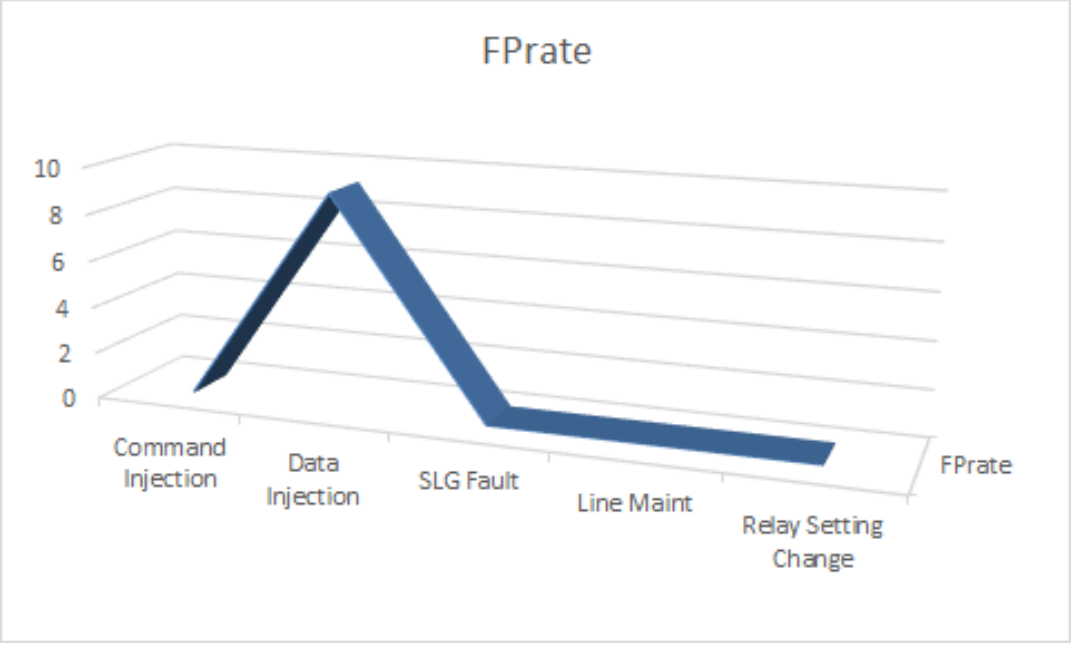


FIGURE 5.2: False Positive Rate

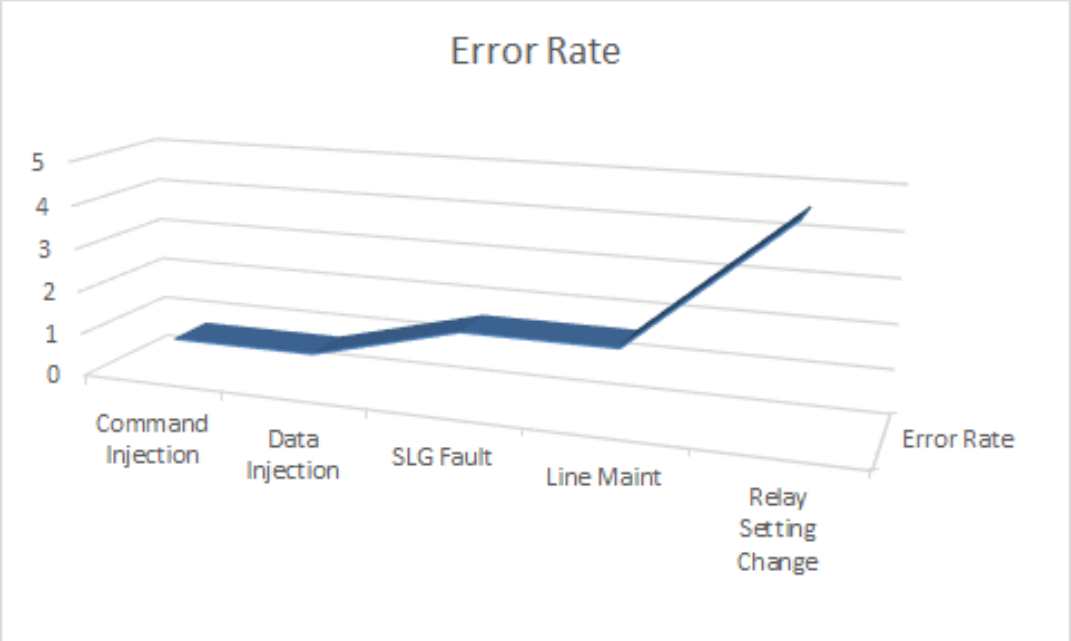


FIGURE 5.3: Error Rate on Raspberry Pi

complete power tranmission network. In our case the measurements from all four IEDs, 29 measurements from each IED making total of 116 and 12 x observations from other components (Snort, a NIDS, Relay Status, Control Panel logs) of the whole system. Flare achieved the satisfactory performance reliability.

ONRL and Mississippi state University power system contains a column 'marker' which, indicates the scenario (Table 5.6 shows the high level grouping of events against the scenarios. Experiment for central deployment were conducted in six phases, description of each phase is as **??**. Normal meauremernts were sliced into 55 % as normal train and remaining 45 % for test normal or validation dataset.

Table tab:Own presents the complete summary results while Figure fig:exp2drate, fig:fnrate and fig:fnerrate presents the graphical view of the results.

Ser	Label	Event (Normal/ Natural Fault Event/ Cyber Attack)
1	41	Normal
2	1-6	Natural Fault Events (Various locations along line)
3	13-14	Natural Line Maintenance Events
4	7-12	Data Injection Attack
5	15-20	Command Injection Attack
6	21-40	Relay Setting Change Attack

TABLE 5.6: ORNL and Misstep State University State Power System Datasets Labels

Ser	Phase	Testing
1	Phase 1	Normal Vs. Attack
2	Phase 2	Normal Vs. Natural Fault Event
3	Phase 3	Normal Vs. Line Maintenance Event
4	Phase 4	Normal Vs. Data Injection Attack
5	Phase 5	Normal Vs. Command Injection Attack
6	Phase 6	Normal Vs. Relay Setting Change Attack

TABLE 5.7: Phase wise testing scenarios and scenario details

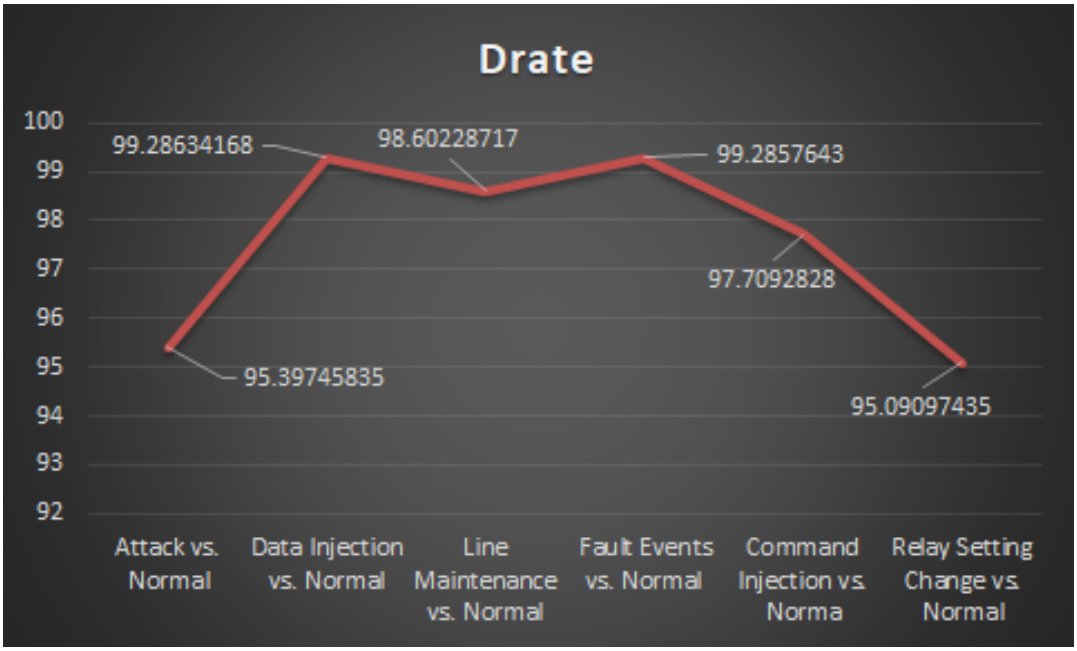


FIGURE 5.4: Detection Rate

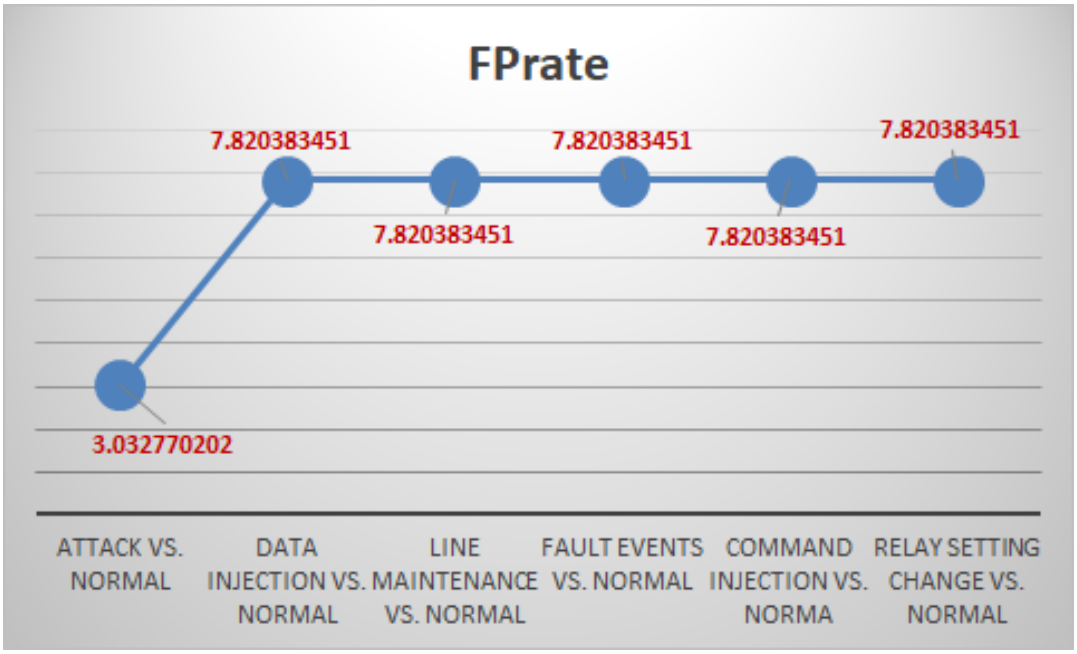


FIGURE 5.5: False Negative Rate

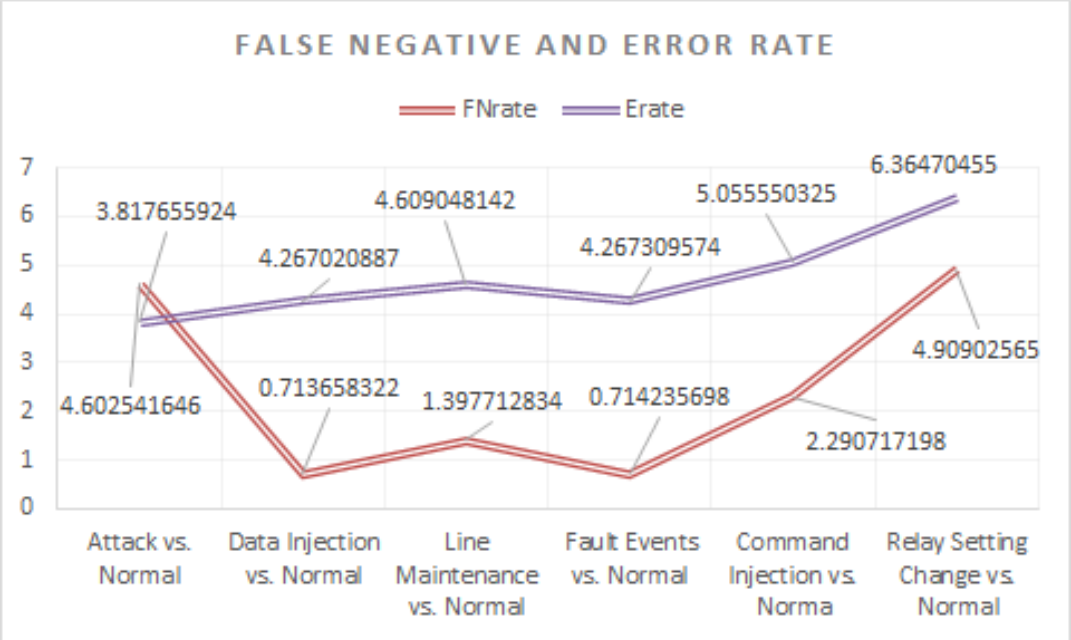


FIGURE 5.6: False Negative Rate and Error Rate

Ser	**Scenario**	**FPrate**	**FNrate**	**Erate**	**Drate**
1	Attack vs. Normal	3.0328	4.603	3.818	95.39745835
2	Data Injection vs. Normal	7.8204	0.714	4.267	99.28634168
3	Line Maintenance vs. Normal	7.8204	1.398	4.609	98.60228717
4	Fault Events vs. Normal	7.8204	0.714	4.267	99.2857643
5	Command Injection vs. Norma	7.8204	2.291	5.056	97.7092828

TABLE 5.8: Detection Rate (DR), False Positive Rate (FPR), Error Rate (ER) and False Negative Rate (FNR) measured on Core i7 with 8 GB RAM

5.4 Comparative Analysis - WAMS based Intrusion Detection Systems

This section is dedicated to compare Flare's performance with other algorithms discussed in chapter 2. Table tab:comparison summarizes the results achieved by other algorithm while Table tab:own summarizes the results achieved by Flare.

Ser	**Algorithm**	**Result**
1	Naïve Bayes,OneR, NNge, Jripper, Random Forest, SVM, Adaboost (Hink et al., 2014)	JRipper+Adaboost achieved F-Measure 0.955
2	Novel Specification-based Bayesian network (Pan, Morris, and Adhikari, 2015c)	100% for Nine scenarios
3	Sequential Pattern Mining (Pan, Morris, and Adhikari, 2015a)	95 % for Experiment 1 87.6% for Experiment 2 93.21 % for Experiment 3
4	Temporal state-based specifications Learning (Pan, Morris, and Adhikari, 2015b)	90.4 % accuracy 73.43 % for Zero-days
5	Fuzzy c-means (FCM) with fuzzy inference system (FIS) (Tomlin, Farnam, and Pan, 2016)	99.11 % Overall accuracy
6	Hoeffding Adaptive Trees (HAT) improved by the drift detection method and adaptive windowing (Adhikari, Morris, and Pan, 2018)	Kappa statistic is 94% for binary and 93% for multiclass classification.
7	Specification-Based and Decision trees, instance-based learning and NNGE (Chen et al., 2017)	NNGE+STEM achieves 96% and 93% and HAT achieves 98% and 92 % for Binary and Multi Class
8	Non-nested generalized Exampler with State Extraction Method (STEM) (Adhikari, Morris, and Pan, 2016)	94.0%, 98% and 99. 54% for Three, Binary and Multi Class on the IEEE 9 Bus System
9	Extreme Learning Machine (ELM) model optimized by the Adaptive Differential Elitist Evolution (Demertzis and Iliadis, 2018)	96.55% with proposed AEDE-ELM
10	Stacked Autoencoders (Wilson et al., 2018)	99.89 % for Normal Events 98.94 % for Data Injection 97.34% for Command Injection 98.53 % for Relay Setting Change
11	Probabilistic classification (Bayesnet), Adaptive boosting, Ripper, Bagging, SVM OneR, C4.5, Majority voting, Random forest (Chen et al., 2018)	Adaboost C4.5 and Random forest achieved more than 90 % accuracy
12	NNGE, HAT and Common Path Mining (Adhikari, Morris, and Pan, 2017)	94 % with NNGE 94% with HAT 93% with Common Path Mining

TABLE 5.9: Accuracy acheived by algorithms in litrature

Chapter 6

Conclusion

This research work made a notable contribution to research in the field of Cyber-physical Intrusion Detection Systems (CP-IDSs) by introducing a lightweight plug and play CP-IDS technology that can be deployed at substation level. As a result, a new approach for designing a reliable ML based CP-IDS that can distinguish fault events, line maintenance events and cyber-attack attempts to mimic legitimate behavior of the system in real-time.

An essential extension of gaussian mixture modeling is proposed to develop a threshold based decision engine capable of learning in an unsupervised manner. In this study, the central component of CP-IDSs technology, decision engine (DE) have been addressed.

This chapter is organized as follows: Section 6.1 details the major contributions of this research endeavor; Section 6.2 summarizes important limitations of the developed system; Section 6.3 outlines some open questions and suggestions for future research; and Section 6.4 presents concluding remarks.

6.1 Contributions of this Study

This section provides an overview of the contributions resulted from this research endeavour. These contributions can be summed into three, first and second contribution is presented to industry and third one is presented to academia by crafting a novel extension of probability density based classifier capable of processing real time streaming data.

(Addressing questions 1 given in Section 1.4.1) ***How can a lightweight CP-IDS be designed, to embed directly in an Intelligent electronic device (IED), for sub-station IED deployment?***: successfully developed a light-weight CP-IDS to embed in an IED. Developed algorithm can also be deployed independently on an inexpensive commodity hardware. Developed algorithm can be deployed centrally or in the sub-station at IED location, in sub-station IED deployment, each IED will be equipped with a lightweight CP-IDS, so CP-IDS will not consider attacks that occur on a separate line. Considering the geographically scattered power transmission networks, IED deployment is also a practical strategy to detect malicious activity near to source.

(Addressing questions 2 given in Section 1.4.3) ***Can a stream processing algorithm be developed to process and classify high-speed data of WAMS in an online fashion and detect complex unknown attacks?***: Developed execution algorithm accumulates the measurements of a second (i.e. 120 x 29) calculate the correntropies for this instance and compare the computed correntropies with decision correntropies computed earlier in training stage and alarm in case of anomaly at real-time. At this stage instance is discarded from memory to hold the data of next instance.

(Addressing questions 3 given in Section 1.4.3) ***Prevalent ML based solutions are***

qualified under direct supervision, can we develop an algorithm that learn without supervision with the aim to reduce heavy costs of manual labour and increased automation?: Developed algorithm starts learning as it is plugged into the system and fed with WAMS data. Developed algorithm do not require the labels, specifying an instance of measurements a malicious or benign. Ideal training is done on first time installation of IED. Training algorithm presume that while in train mode all data is benign. Learning and execution phases are explained in section 4.5).

(Addressing questions 4 given in Section 1.4.4) ***If gaussian density estimation and Correntropy approaches can be used to model the multivariate data of synchrophasor based WAMS?***: Gaussian probability estimation successfully achieved the purpose of mixture modeling of synchrophasors, while correlation entropy theories are used to accurately construct decision boundaries. Central limit theorem has been instrumental in revealing the latent statistical distribution of the synchrophasors.

6.2 Limitations of Proposed System

At its current development stage, the main limitations of the Flare decision engine, training and execution algorithms are described below.

Flare's assumption that all measurements of WAMS are benign during train-mode. During training, an attack already in its execution phase will cause the DE to train on malign measurements, which will badly impact its decision capabilities. This risk can be minimized by ensuing the system is not under attack when the Flare is deployed.

The dataset used for evaluation of the Flare, is generated using a power transmission network test-bed The ORNL and Mississippi State university power system datasets, discussed in Chapter 3 are generated by simulating a reference 2-Bus 2-Generator

power transmission system. Reported results and performance capability of Flare may change in real world deployment.

Flare is not immune to adversarial machine learning Advance adversary can craft malicious inputs to mislead the Flare's core algorithm model (Kurakin, Goodfellow, and Bengio, 2016). During re-training of Flare learning algorithm, an attacker may inject skillfully designed samples to compromise the learning process.

6.3 Open Questions and Future Work

Although this study has created a significant amount yield by developing an online gaussian mixture modeling based decision engine, further work is required to improve the Flare's capabilities to fulfil limitation highlighted in section 6.2. The key open questions are summarized below.

- How the risk of ill-learning can be minimized in case of an attack already in execution phase during training?
- Develop the capability of Flare to counter the adverbial machine learning.
- How the accuracy of the developed Gaussian Mixture Modeling based DE discussed in Chapter 5 can be improved?
- What are the best features of power systems for online classification?
- Evaluate the Flare's performance capability in real-world production power transmission system.

6.4 Final Remarks

Continuous emergence of new hacking paradigms, variety of available attack techniques and vulnerabilities of current systems are the noteworthy risks, that are perceived by society, from broad viewpoint. There is no silver bullet to combat all the possible attacks. Profound defense in depth strategy demands, adding security in layers to improve overall security posture of a system as a whole.

Complete code of designed decision engine (DE), Flare, discussed in the chapter 4 and 5 of this thesis are released publicly, which is a significant contribution allowing CP-IDS researchers to extend Flare's capability or reuse the Flare's core concepts to design better system.

To conclude, security of cyber-physical power transmission subsystem is an evolutionary process and significance of the cyber security will continue to grow. Timely detection of novel attacks and high-impact new threats enables application of corrective measures ideally, as fast as possible before attackers are successful. The significance of defensive security technologies is established to identify the novel attacks and threats, importance of in-house penetration testing and pre-deployment security evaluation of devices cannot be ignored. In future, researchers are encouraged to incorporate other matured technologies to identify the potential system and network flaws in development and deployment of protective measures to detect and deny cyber attacks.