Imperial College of Science, Technology and Medicine
Department of Electrical and Electronic Engineering
Control and Power Research Group

Operational Moving Target Defences for Improved Power System Cyber-Physical Security

Martin Henry Higgins

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy and for the Diploma of Imperial College, October 2021

Abstract

In this work, we examine how Moving Target Defences (MTDs) can be enhanced to circumvent intelligent false data injection (FDI) attacks against power systems. Initially, we show how, by implementing state-of-the-art topology learning techniques, we can commit full-knowledge-equivalent FDI attacks against static power systems with no prior system knowledge. We go on to explore how naive applications of topology change, as MTDs, can be countered by unsupervised learning-based FDI attacks and how MTDs can be combined with physical watermarking to enhance system resilience. A novel intelligent attack, which incorporates dimensionality reduction and density-based spatial clustering, is developed and shown to be effective in maintaining stealth in the presence of traditional MTD strategies. In resisting this new type of attack, a novel implementation of MTD is suggested. The implementation uses physical watermarking to drive detection of traditional and intelligent FDI attacks while remaining hidden to the attackers. Following this, we outline a cyber-physical authentication strategy for use against FDI attacks. An event-triggered MTD protocol is proposed at the physical layer to complement cyber-side enhancements. This protocol applies a distributed anomaly detection scheme based on Holt-Winters seasonal forecasting in combination with MTD implemented via inductance perturbation. To conclude, we developed a cyber-physical risk assessment framework for FDI attacks. Our assessment criteria combines a weighted graph model of the networks cyber vulnerabilities with a centralised residual-based assessment of the physical system with respect to MTD. This combined approach provides a cyber-physical assessment of FDI attacks which incorporates both the likelihood of intrusion and the prospect of an attacker making stealthy change once intruded.

Statement of Originality & Copyright Declaration

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources has been acknowledged.

Acknowledgements

I would like to express my very great appreciation to Dr Fei Teng for his valuable and constructive suggestions during the planning and development of this research project. His dedication to the work and generosity with his time has been very much appreciated. His insights, attention to detail and unwavering work ethic have been incredibly valuable.

To Professor Thomas Parisini, your guidance as my senior professor has been excellent and well appreciated.

To Professor Keith Mayes, thank you very much for your invaluable advice and mentorship throughout. I immensely enjoyed our time working on our paper and the Royal-Imperial Black Box commercialisation project.

To all the CAP PhD students and postdocs and the members of the Smart Grid CDTs who helped me throughout my studies, I express my thanks.

To my mother and father, without whom this would not have been possible.

This work was supported by EPSRC Centre for Doctoral Training in Future Power Networks and SmartGrids (EP/L015471/1).

Dedication

To my wife Mackenzie and newborn son Henry. You have provided more motivation than I ever needed.

‘The art of war teaches us to rely not on the likelihood of the enemy’s not coming, but on our own readiness to receive him; not on the chance of his not attacking, but rather on the fact that we have made our position unassailable.’

Sun Tzu

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

The power system is arguably the most critical of all modern networked infrastructures. To some extent, water, communications, sanitation and defence systems are all dependent on a stable electricity supply. Cascading blackouts and system failures can be costly, both in financial and human life terms, and so it is no wonder that the power system has become a key target for hackers seeking to disrupt or destroy. In 2015, Ukraine experienced one of the first successful (and well-documented) cyber-attacks against a power system distribution operator. The consequences of this attack were significant: around 300,000 people were disconnected from their power supply [1]. Recently, cyber-attacks against power systems have, unfortunately, become a major concern for power system operators. The advantages of cyber-attacks, in particular, are clear. The anonymity and plausible deniability the attacks offer are invaluable to national players seeking to minimise the political fallout from such operations. From the perspective of the attacker, the low-risk, high-reward nature of cyber-attacks means hackers can be hired from as little as around 5 USD per hour [2].

The events in Ukraine demonstrate an appetite to commit cyber-attacks against power systems and other critical infrastructures. While it is true that the Ukraine attacks show it is possible to target and successfully compromised a SO directly. In the future, this route to compromise is less likely to be available to attackers. This is partially due to the Ukraine attacks themselves. The high profile

nature of this attack has meant that security for SOs has increased and the vulnerabilities which were exploited for Ukraine are unlikely to be exploitable again going forward. We consider that, in the future, attackers are likely to explore alternate routes to compromise. These might include bottom up routes to attack i.e. compromising poorly defended, distributed system assets as opposed to the central control systems. We also consider, that in some respect, the consequences of the 2015 Ukraine power grid attack were muted. Although the attacks were successful, they were discovered quickly and rectified (from a power supply perspective) within a few hours. By comparison, if the attackers had opted for a deception-style attack and remained hidden post-attack initiation, they may have been able to cause more damage to the system in the long run. With this in mind, we discuss two main research areas related to deception-style attacks. The areas we discuss primarily are false data injection (FDI) attacks (sometimes called FDI attacks or FDIAs) and Moving Target Defence (MTD). Through this work we seek to make contributions to both the attack and defence literature. In the following, we offer a brief description of each of these fields with an expanded literature review on each of these topics found in the background chapter.

1.2 A Brief Introduction to False Data Injection Attacks

FDI attacks are a form of deception attack which involve using the system assets (usually measurement infrastructure) against the system operator. FDI attacks were first outlined for power systems in [3]. The attacks chief aim is to deceive the SO without triggering underlying bad data detection (BDD) processes used by the central state estimator. To achieve this, the attackers will 'inject' false data into distributed system measurements to manipulate either the SO directly or general system processes to gain a certain outcome. The principles are quite straightforward. The attacker injects data into the individual systems measurements. In doing this they seek to replicate a scenario that might cause detrimental action to the system. These detrimental actions can either be by the SOs own action i.e. dispatch power where non is needed or automated control responses. These types of attacks can be applied in power systems, gas networks, water systems or essentially any networked system reliant on distributed monitoring. FDI attacks can generally be considered to have a high knowledge barrier in comparison to other attacks. This is because in addition to all the intrusion requirements normal

cyber-attacks have, FDI attacks also require a good understanding of the underlying system in order to bypass BDD and remain hidden. There are various strategies for doing this. The use of topology information to remain hidden is quite commonly suggested because it allows high levels of flexibility in how attackers can structure their attacks. Hackers can use topology information to stay hidden and in fact often a central assumption in FDI attack research is that the attackers possess good knowledge of the system topology. Given the requirement of topology knowledge to perform an FDI attack, a sub-field has emerged on preventing the use of topology information for FDI attackers to stay hidden (MTD). Further, stealthiness is critical for an FDI attack to be successful. If a system operator knows an attack vector is false he will discard the measurements and attempt to establish the true attack vector.

1.2.1 Potential Consequences of FDI attacks

The consequences of a successful FDI attack can be severe. When performed under a full knowledge and access assumption the attacker can effectively replicate the appearance of any system state to the operator. The resulting impacts on system stability impacts and operation can result in a number of undesirable outcomes. For example, a system which is suffering already from under-voltage. The attacker could replicate a measurement set that hides the under-voltage from the system operator. Further, the attacker could attempt to compound the under-voltage by representing a system state that might encourage disengagement of generating assets. We might consider this as a load-shedding style of attack. In this way, the SO actually attacks his own system by trying to rectify the perceived system state. The opposite approach is also possible. Looking for areas operating at capacity and presenting under-voltage style vectors will give the appearance that more generation dispatch is needed. If the operator responds in the expected manner i.e. by initiating generator dispatch we could easily see over-voltages, higher expected temperature and protection trips. Further, if left undetected for a long period lines could even be permanently damaged by run over capacity. Protection trips resulting from these attacks in turn could result in cascading blackout failures, particularly if aspects of the blackout are hidden using the compromised measurements. Further, it would also be possible to manipulate market pricing in short-term spot markets if the system state could be manipulated. An operator of

fast response generation could benefit greatly from a perceived regional lack of generation.

1.3 A Brief Introduction to Cyber-Defences

While cyber-defence is a large field with many sub sections. Broadly, we can characterise the available defences for SOs into a few key areas. Authentication defences which include encryption, password protection. Analytic or mathematical based detection which would include anomaly detection, state estimation process. We might also consider social engineering based defences, utilising hierarchy, internal checks and sign off. Unfortunately, all these forms of defences are susceptible to FDI style attacks. The distributed nature of the attack means that the attacker can compromise old, distributed assets, which can be installed pre-cyber-security considerations. Thus making much of the authentication strategies redundant. FDI attacks are also structured to look plausible to internal analytics processes and to the human SOs which means they can bypass error checks both mathematical and human. We must therefore, consider a tool with which to evaluate FDI attacks. One such method is 'Moving Target Defence' or MTD.

1.4 A Brief Introduction to Moving Target Defence

MTD is a catch-all term for a collection of technologies, techniques and protocols for cyber-defence. In contrast to other conventional sources of defences, which utilise system protocols or software to enhance security, MTDs increase cyber resilience by dynamically changing the underlying system in which they operate. In this work we look exclusively at the attack and defence of power systems with specific reference to the FDI attack. As discussed, FDI attacks usually require system knowledge (such as network topology) to be successful. It makes sense that if the system operator can render the attacker's information obsolete, the attacker will struggle to attack without creating gross errors that can be exposed. Changing the system in this way has often been referred to as a moving target in that it changes the underlying model (the target), and if the attacker does not update his model, his attack will likely fail as their attacking model will now be based on obsolete information. Within

a power system context, MTD usually refers to the use of network infrastructure such as distributed flexible AC transmission system (D-FACTS) or transmission switching to invalidate the attackers system knowledge. However, MTD is not merely limited to power systems. There is also a body of work related to MTD in cyber networks which usually involves using virtual private networks (VPNs) and circulation of network channels to expose attackers.

1.5 Objectives

Our objective with this work is to innovate in a few areas. To start, we'd like to expand the literature on FDI attacks. In the past, we have seen FDI attacks primarily performed against static system with few defences considered. However, in practice, power systems themselves are quite dynamic and increasingly open to implementation of new defences. We intend to examine how to make FDI style attacks MTD-competitive i.e. how to make it so they can potentially bypass MTD defences. We intend to offer attackers intelligent solutions to attacks that takes advantage of state-of-the-art topology learning and data driven techniques. Almost all the current works on FDI attacks have been performed under the assumption of either full knowledge of the current system topology or (at least) a fixed underlying topology. Therefore we believe the relaxation of this assumption will provide an interesting research avenue and believe there is significant research value in exploring the question of how to circumvent MTD (as an attacker). In fact, in this work, one of the first things we will show is that attacks (even with limited knowledge) can use data to get limited topology information and circumvent naive applications of MTD.

We also note that there is a lack of operational discussion on the practical application of MTD and how it should be used. Papers focus on the evaluation of attacks and rarely on the implementation. MTD is costly both in terms of both infrastructure and ongoing operational costs. We also hope to explore methods of reducing the cost of MTD, particularly in the instance of no attacks present. While MTD has been shown in the past to be an effective method of counteracting FDI it also has its own drawbacks when applied in the power system. We intend to explore event-triggers as a method of reducing the overall application of MTD in the power system. We also wish to explore distributed

solutions detection solutions to the MTD application problem in order to enhance regional protection.

Finally, we would like to create a risk-based assessment method for system operators to assess the native risk of their system with respect to FDIAs and MTD. Much of FDI research has been focused on the ability to attack, and little focus has been put on actual advice for system operators in defending their systems. We are hoping to tie together these areas and produce a body of work that makes a real and applicable contribution to the defence of systems from FDI style attacks.

1.6 Contributions

- Our first contribution is to use state-of-the-art (SOA) topology learning algorithms to create a topology-learning-aided FDI attack capable of attacking power systems under a low knowledge assumption model.
  - In the past, the level of knowledge of power system topology has been an important assumption for attacking models. Our attacking model is committed against the AC power system and uses the latest state-of-the-art topology discovery techniques to build a model for the network.
  - As part of this topology learning attack, we introduce an attacker-side criteria assessment via a pseudo-residual calculation to allow the probabilistic assessment of attack success before any attack has been committed, allowing the attacker to ensure stealthiness. This contribution allows attackers to assess attack likelihood of success before engaging an attack. This is crucial as we consider the high costs of intrusion and the attackers aversion to detection. We also show regional pseudo-residuals can be used to verify local attacks, even in the presence of global topology errors.
  - Finally, for this topology learning attack, we demonstrate how quickly the attacker can develop full knowledge of system topology and parameters and effectively validate the full system knowledge assumptions in previous studies. As we have good data on prior intrusion times from events such as the Ukraine 2017 attack, we can utilize this contribution allows us to assess to likelihood of a blind attack being plausible.

- For our 2nd results chapter we provide a few contributions. On the attacking front, we explore the idea of counter-MTD or MTD resilient FDI attack techniques. With the success of these new attacks established. We introduce a new 'camouflaged' implementation of MTD to drive detection against traditional and intelligent FDI attacks.

  - Whereas previous FDI attacks have been designed against static systems, we seek to offer new attacking considerations in the presence of dynamic systems with MTD. This contribution allows for attacks in non-static systems with MTD in place providing an attacking counter to naive or simple applications of MTD.
  - The proposed intelligent attack works under zero system knowledge assumptions and combines dimensionality reduction and unsupervised learning to identify the underlying clusters associated with network topology and to design the corresponding attack vector.
  - This method is shown initially to be effective at discerning the underlying clusters associated with the different MTD configurations. It is then extended to a blind-ICA attack and shown to be effective and stealthy against traditional MTD.
  - The proposed defence strategy combines MTD and physical watermarking to enhance security via an added 'Gaussian' watermark into physical plant parameters. As the added watermark mimics the underlying noise of the system, the physical changes driven by MTD stay hidden.
  - The physical watermarking is combined with cumulative error monitoring to spot minor but sustained changes in the system to trigger alarms.

- For our 3rd results chapter, we explore event-triggered MTD as a method of reducing the overall application of MTD in the power system.

  - We also attempt to consider a more distributed approach to the use of MTD and propose a distributed/regional protection protocol. The protocol consists of an initial anomaly detection followed by MTD and traditional CSE.
  - The anomaly detection uses Holt-Winters seasonal forecasting distributed to the individual measurement. The seasonality captures the intra-day demand differences to minimise the

overall window for attack. If the anomaly detection is triggered, we then apply MTD implemented via inductance perturbation of D-FACTS devices, which will drive the residual errors in CSE even under the stealthy FDI attack.

- In our final results chapter we examine risk assessment of FDI attacks with MTD.
  - To start, our model provides a weighted graph assessment of the FDIA intrusion risk of the cyber components of the grid. In addition to FDI vulnerabilities or RTUs and meters, our model includes (for the first time) overlapping-style attack opportunities, i.e. not simply the choice of the RTU or the meter combinations for a given state but also some combinations of the two.
  - We also introduce an MTD (post-intrusion) effectiveness criteria that considers system capacity constraints in the context of an FDI attack and the required level of MTD to expose an attack for an overload-style attack. We model the level of divergence required to protect each bus and branch combination in the context of a minimum attack vector.

1.7 Publications

1.7.1 Journal Papers

M. Higgins, F. Teng, and T. Parisini, "Stealthy MTD Against Unsupervised Learning-Based Blind FDIAs in Power Systems," IEEE Transactions on Information Forensics and Security (Accepted, Published November 2020).

M. Higgins, K. Mayes, and F. Teng, "Enhanced Cyber-Physical Security Using Attack-Resistant Cyber Nodes and Event-Triggered Moving Target Defence," IET Cyber-Physical Systems: Theory & Applications (Accepted, Published March 2021).

M. Higgins, X. Wangkun, F. Teng, and T. Parisini, "Cyber-Physical Risk Assessment for Power System FDI Attacks with Moving Target Defences" (Pending submission).

1.7.2 Conference Papers

M. Higgins, J. Zhang, N. Zhang, and F. Teng, "Topology Learning Aided False Data Injection Attack without Prior Topology Information," Proceedings of the IEEE Power Energy Society General Meeting July 2021 (Accepted)

1.7.3 Commercial Ventures

On the back of the research performed in this thesis we have received 30K in funding from as a applicant on the Royal-Imperial Black Box (RIBB) project with an application via Innovate UK. Our device will combine physical verification via MTD with enhanced cyber nodes. The project is intended to be a commercialization of our IET Cyber-Physical Systems publication. The RIBB builds on the techniques outlined in [4] to propose a novel way of protecting the power system. The project has been successfully carried through and has won an additional 60k in funding to develop a prototype device which is now currently being developed.

1.7.4 Contributing Author Works

X. Wangkun, M. Higgins, F. Teng, "Blending Data and Physics Against False Data Injection Attack - An Event-Triggered Moving Target Defence Approach," (Pending submission).

Chapter 2

Background

This chapter provides the background knowledge for the areas approached in this work. Initially, we outline the state-of-the-art literature for each relevant sub-field in the thesis. This chapter provides background on FDI attacks, MTD and cyber-physical risk assessment. Following this, we provide preliminaries for areas that are common to the results chapter, which includes formulations on state estimation, FDI attacks and MTD.

2.1 Relevant Literature

The modern power system is dependent on state estimation processes for effective operation and control. The basis of power system state estimation is the estimation of power system angles and voltage magnitudes from observed real and reactive power flows. The seminal work [5] outlines much of the basis for the modern state estimation of power systems and is still relevant some 22 years after publication. Further, a survey on power system state estimation can also be found at [6].

2.2 False Data Injection Attacks

As discussed, an FDI attack is a form of deception attack that utilises the altering of power system measurements to deceive the system operator. First outlined in [3], FDI attacks involve altering power flow measurements to corrupt a network operator's state estimation processes and deceive the system operator into seeing a state that isn't correct. The consequences of a successful FDI attack can be severe. For example, in [7], it is shown that FDI attacks can be used to promote load shedding via the system operator. In [8], it was shown how FDI attacks can be used to mask transmission outages from the central system operator. Instigation of line outages was shown to be possible in [9]. While in [10] limited information FDI attacks are shown to be capable of strategic line overloading. Cascading blackouts were shown to be possible in [11] where it was shown to be possible to cause wide-scale blackouts across grids using FDI attacks. Market manipulation via FDI based attacks has also been shown to be possible in [12].

FDI attacks use system information to remain hidden while attacking system measurement to misguide the WLS-based state estimation [5]. Although extensive research has been conducted to detect FDI attacks, the vast majority of papers operate on the assumption that residual testing is based on centralised state estimation (CSE). FDI attacks need to remain undetected by the network operator to be effective or the attack vector will simply be discarded. To this end, FDI attacks compete with bad data detectors (BDD) within state estimation processes. In a modern energy management system (EMS), the BDD at the power system level relies on weighted-least squares (WLS) and chi-squared error testing [13], meaning an attacker needs to structure the attack based on the system topology model to remain undetected. Initial models for FDI attacks assumed full knowledge of the system and full access to meter measurements within the system [3]. However, shortly afterwards works began to reduce this knowledge requirement. An incomplete knowledge attack was introduced in [14], which showed that a system could be attacked with only partial knowledge of the system topology and a subset of meter measurements. In [15] the blind FDI attack is introduced, which requires no system knowledge, provided the attacker has access to all meters within the attacked grid system. The blind FDI attack uses independent component analysis (ICA) to map the inter-correlations of the visible meter measurements to create an approximation for the power-flow model. A more effective

version of the attack that utilizes partial susceptance knowledge was developed in [16], allowing an islanded approach where the visible or 'high knowledge' parts of the system could be attacked by the standard-FDI attack, whereas low information areas could be attacked by the blind approach. Some recent studies enhanced FDI attacks by combining them with other forms of attack, such as denial of service (DoS) attack [17]. Other state-of-the-art works have now suggested methods of attacking the power system with limited system knowledge. In [18] a data driven approach for blind attacks in the AC power system is proposed. Jiao *et al*. utlise generalised adversary networks to create an effective attack model with limited information. Data-driven approaches with reference to FDI attacks are mostly from the defenders perspective (anomaly detection) [19][20] [21] [22]. In [23], an algorithm-utilizing principle component analysis (PCA) is developed for anomaly detection. By contrast, [19] employs margin-setting algorithms (MSA) for the detection of FDI attacks and compares their performance with support vector machines (SVM) and artificial neural networks (ANN). In [20], isolation forests are used for the detection of stealthy FDI attacks. In [24] a generative online network (GAN) is proposed as a defence against FDI attacks. The work incorporates a physical model which captures ideal measurements in combination with the data driven GAN approach to capture deviations from ideal measurements. In [25], singular-value decomposition is used to construct attack vectors without knowing the underlying system measurement matrix. In [26], two strategies using sub-space separation are suggested: one aims to use estimated system subspace to hide attack vectors and another aims to mislead BDD so that non-attacked measurements are removed. These methods allow for admittance values to be estimated but require a large number of historical measurements. In [27], sparse FDI attacks against wide area measurement systems and defence methods are explored. Using historical data to mount FDI by using multiple linear regression model was outlined in [28]. In [29] the FDI attack problem is treated as a dynamic game utilising a Bayesian approach. In [30] the degrees of freedom available to the attack vector are considered. These are used to constrain the attacker in how he may manipulate the system and provide an optimal defence set for the system operator.

A few papers recently started to tackle FDI attacks from a distributed perspective. In [31], [32] & [33] FDI attacks against distributed Kalman-style or extended Kalman-style filtering are explored to show how dynamic-style estimation (at the individual state or measurement) can still have potential susceptibilities to attack. One of the most under-served areas with respect FDI attacks is in the field

of post attack discovery. As we have seen in previous literature, innovations to attack and data driven methods of attack detection are quite well served. However, what to do post these attacks is a field rarely discussed. Jorjani *et al*. attempt to rectify this. In [34] they offer an optimisation based approach to recover correct system measurements post attack. This research area is one of the crucial going forward for the field of FDI attacks. While there are many works now operating on the detection and execution of FDI attacks few have focused on what actions to take after an attack has successfully been launched. Even in commercial realm, the vast majority of ICS or SCADA security solutions are focused on pre-intrusion detection with few companies covering post-intrusion recovery solutions. A comprehensive review of FDI attacks can be found in [35] & [13].

2.3 Moving Target Defences

We note that almost all the current literature on FDI attacks, focus on fixed network topology in their attacking model. Whether these data-driven approaches can be applied to design FDI attacks under intentional or unintentional topology changes has not yet been investigated. In fact, because FDI attacks are dependent on the characteristics of the physical system, it is essential to understand whether and how the physical system can be used to actively defend against attacks. We also consider that to be secure at the physical layer, the system must be protected against these types of attacks. As a result, a body of work has emerged to utilize the physical system to actively defend against attacks. Topology-driven MTD has been previously shown to be effective against FDI attacks. In particular, MTD is proposed through either transmission switching [36] or admittance perturbation via D-FACTS devices [37]. Very few papers have seriously explored transmission switching as a legitimate method of applying of MTD in the power system. This is because, although in theory it could be effective at driving increased residuals in the case of an attacker unaware of MTD, as we showed in [38], the impacts of such an approach on a power system control and the cost to an operation would be undesirable. Also, from the attacker's perspective, identifying these changes and countering are much easier in the case of transmission switching. The first to suggest reactance-style perturbations (via D-FACTs) as a form of MTD was Morrow *et al* in [37]. They used a proposal of randomised key-spaces to select MTD configurations. By comparison to transmission switching,

the cost of D-FACTS-based applications is lower and commiserate with less overall impact to power system stability. Indeed, although it has been shown that MTD can be effective, limited work has addressed how it should be implemented in power systems. Questions around the frequency and when MTD should be applied remain. Continuous application of MTD will result in significantly higher costs and time-based application will be susceptible to in-cycle attacks.

We outline some of the keys works in MTD for power systems here. In [39], a reactance perturbation scheme for evaluating FDI attacks is offered. They utilize a secure reactance perturbation optimization heuristic to maximise the detection probability of FDI attacks against power systems. This offers a direct improvement on the work outlined in [37] because it provides a structured way of maximising the detection chance of MTD. One of the most crucial papers in this sub-field, from the operational perspective of MTD, has been [40]. In prior works, the cost of applying was often ignored in favour of addressing problems on detection and optimal use. Lakshminarayana *et al* offered the first look at a functional cost of MTD given the sub-optimal use of assets to provide the system changes. They have also made a further contribution to the MTD field with a paper on applying a game theoretic approach to MTD [41], wherein they use a Nash equilibrium solution to minimise the defensive costs during the systems operational time. Another paper that provides important considerations from an operational perspective is [42]. In this work, Zhang *et al* prove the susceptibility of isolated state measurements and design an algorithm for branch perturbation selection. This proof has interesting implications for MTD against very sparse networks (such as distribution networks), namely that busbar points with no interconnections will not be defensible via MTD. Some limitations of MTD were explored in [43], which included consideration for isolated measurement buses. From the perspective of MTD planning, in [44], a heuristic is developed for a near optimal solution for D-FACTs deployment. The author also provide a coordinated perturbation scheme design to improve MTD performance in the detection of FDI attacks. In [45] another systemic approach for MTD planning is outlined. In this work Liu *et al*. outline a graph theory based planning metric for optimal distribution of MTD throughout a power system. They also demonstrate the need for an upper bound on MTD effectiveness to protect a system. Relatively few works have approached MTD from the perspective of the distribution network or three-phase systems. In [46] metrics are provided for the performance of MTD under the more realistic three-phase scenario. Optimisations are made between hiddenness and effectiveness

while attempting to maintain system stability. In [47], an event-driven MTD protocol is proposed for control systems.

With attackers' increasing capability, there are growing interests in the research community to design new forms of MTD that can hide its existence to the attacker. One of the key state-of-the-art papers in this field is [48], which presents an enhanced hidden MTD model to make the topology change invisible to an attacker via identifying alternative topology and state combinations under the same power flow profile. Whilst this method is clearly effective, it relies on being able to find an alternative topology and alternative states to maintain constant power flows, which can be computationally expensive and even infeasible in a system with limited acceptable state ranges. In some ways, this work was a solution in search of a problem. Up until this point, there has been very little work on MTD-resilient attacks. In [38], we provide both the problem for Tian *et als* solution and propose an alternative to their direct solution. In our work, MTD is camouflaged as system noise rather than hidden within the profile of the system. In [4], we argue that MTD should only be applied if there is a creditable suspicion that the system is under attack to minimise this cost.

Similar principles have also been applied to control systems in [49], where utilizing packet drops for MTD was explored in [50]. These forms of MTD don't typically use the physical system in the same manner as power system MTD. Where power system MTD involves altering the topology of the underlying physical system in general, control system based MTD generally involves using system dynamics from the result of dropped packets.

2.4 Cyber-Physical Risk Assessment

Some works have already attempted to tackle risk assessment with respect to FDI attacks. For example, Hug *et al.* explored this in the perspective of the weakest node attack point [51], where they perform node-based target selection using a minimum meters criteria to compromise the node or state angle. The number of meters required for each node are evaluated for the AC and DC models, and an alternate meter conquering strategy is proposed using the RTUs rather than the individual meter measurements. This work offers these strategies separately, i.e. the attacker can capture the meter

or upstream RTU. In practice, however, overlapping vulnerabilities will exist such that an attacker could use combinations of compromising RTUs or individual meters to minimise the overall cost of intrusion. A similar methodology is explored in [52] where security indices are developed based on the physical topology of the power system with specific reference to the FDI attacks. In this case, the security index is defined by the minimum meter change potential with an aim of finding the sparsest possible attack. However, again, the work makes no reference to the RTU or combination-style vulnerabilities and are firmly rooted in individual meter capturing. Similar index-based approaches are also applied in [53]. In [54], Pan *et al*. offer one of the first risk assessments of FDI attacks with cyber considerations for telemetered measurements. The attack combines standard FDI-style attack vectors with DoS-style attacks aimed at reducing the number of meters required to compromise a state. The DoS attacks take the targeted meter out of service and are combined strategically to force the system operator to replace uncontrolled meter measurements with pseudo-measurements. The 'security index' employed refers to the number of meters required to compromise a given node (which can be reduced by incorporating the DoS component). This type of risk assessment is comparable to a meter-only attack (with lower requirements due to the DoS) because it does not consider the RTUs or any components upstream of the meter. Although these frameworks offer some interesting risk perspectives on the FDI attack, they can be improved in a number of ways. Consideration of overlapping-style attacks, which combine RTU and meter-style intrusions, would better represent the risk to a power system.

In addition, previous works have assumed that a stealthy FDI attack is possible once the attacker gains access to the required combination of meters or RTUs. However, this fails to consider the post-intrusion system defences such as MTD. Thus, it is important to expand upon these works by redefining a successful FDI attack by explicitly considering the existence of MTD. Our combined, overlapping vulnerability attack type takes some inspiration from [55]. In this work, *Barrere et al* outline a cyber-physical assessment framework that features combined-style attacks for industrial control systems (ICSs). In this paper, the concept of the risk assessment of ICS systems is explored via the use of AND/OR graphs to capture overlapping vulnerabilities. The AND/OR graphs allow the identification of critical sets of components and combinations of attacks that might minimise the attacker cost in attacking. This allows assessment in the context of overlapping physical and cyber

vulnerabilities, which may allow for a lower cost of capture. For example, instead of attacking ICS sensors directly, a combination of agents or sensors could be used to complete the attack and reduce the overall cost. We take a similar approach for our network intrusion model herein. Other works have approached the subject of cyber-physical vulnerabilities. In [56], a probabilistic risk assessment model is introduced. The model uses acyclic digraphs to represent the inter-dependencies between different components in a cyber-physical system. They also quantify risk in terms of attack impact and attack likelihood to succeed. However, the work is decoupled from power system analysis. Similarly, in [57], a probabilistic risk approach is used but with an attack focusing on the removal of graph nodes or edges and the effects they have on the network. The paper is cyber-physical in that it overlays a SCADA network graph over the power system topology. They simulate random and targeted attack styles, but similarly to [56], the work is done from a graph-theoretic perspective (node isolation) as opposed to a power system perspective of practical outages or overloads. In [58], a framework for the cyber-physical modelling of power grid infrastructure is outlined. The attack focus in this paper is around circuit breaker control and de-energising certain areas of the grid. They combine upper-level RTU modelling with a lower-level telemeter network model. The work is cyber-physical in that it considers RTU placement and underlying power system topology. The authors use partially observable Markov decision processes (POMDP) to critical path mapping and identify potential disconnection loops within the system. Potential targets are weighted to reflect the difficulty in capturing them. One of the earliest relevant works in the field, Bargiela *et al* explore network observability as a function of network topology in [59]. The work also proposes an optimal protection graph that satisfies the spanning tree. This graph can then be used to return a set of optimal buses to protect and guarantee reliable state estimation. In [60], an integrated model-based approach for cyber-physical risk assessment is used that outlines the vulnerabilities of specific controllers into an industrial testbed. Initially, physical threats are derived to assess risk outcomes with the cyber vulnerabilities analysed to find potential path mapping to these outcomes. The work takes a system-specific approach with testbeds for analysing oil and gas systems. In [61], a vulnerability assessment framework for systematically evaluating SCADA vulnerabilities is proposed. The method can be used to model access points for SCADA networks, construct a model for intrusions, simulate cyberattacks and suggest security improvements. The model is cyber-physical in that it analyses the cyber-net model and a corresponding

power flow simulation to assess outage damage. Impact factors are associated with each substation to rate the loss-of-load (LoL) associated with each potential attack. In [62], a meshed network framework that considers both power system features and bi-directional communication flows is presented. They also model the impact of automated control systems with respect to load shedding and relay protection triggering. The paper performs further analysis on the prospect of cascading failures and the overall fragility of the system.

In the next session we outline some shared fundamentals between each results chapter in this thesis. Namely we outline, the derivation for the state estimation problem, the standard standard FDI attacking model and the MTD impact on residual derivation which is common to many of the results chapters herein.

2.5 Fundamentals

2.5.1 State Estimation

As we perform FDI attacks against a power system, it makes sense to first outline the state estimation and error detection processes with which they are competing. We therefore consider a standard AC power system with real power flow measurements under the non-linear expression defined by

$$P_{ij} = V_i^2 g_{ij} - V_i V_j g_{ij} \cos \triangle \theta_{ij} - V_i V_j b_{ij} \sin \triangle \theta_{ij}. \tag{2.1}$$

and reactive power flows by

$$Q_{ij} = -V_i^2 (b_{ij} + b_{ij}^{sh}) + V_i V_j g_{ij} \cos \triangle \theta_{ij} - V_i V_j b_{ij} \sin \triangle \theta_{ij}. \tag{2.2}$$

V and θ are the system states, whilst P and Q are the power measurements. This system is measured by estimating a set of n state variables $\mathbf{x} \in \mathbb{R}^{n \times 1}$ estimated by analysing a set of m meter measurements

$\mathbf{z} \in \mathbb{R}^{m\times 1}$ and corresponding error vector $\mathbf{e} \in \mathbb{R}^{m\times 1}$. The non-linear vector function $\mathbf{h}(.)$ relating meter measurements $\mathbf{z}$ to states $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), h_2(\mathbf{x}), ..., h_m(\mathbf{x}))^T$ is shown by

$$\mathbf{z} = \mathbf{h(x)} + \mathbf{e}. \tag{2.3}$$

The state estimation problem is to find the best fit estimate of $\hat{\mathbf{x}}$ corresponding to the measured power flow values of $\mathbf{z}$. Under the most widely used estimation approach, the state variables are determined by the minimisation of a WLS optimization problem as

$$\min_x J(\mathbf{x}) = (\mathbf{z} - \mathbf{h(x)})^T \mathbf{W} (\mathbf{z} - \mathbf{h(x)}). \tag{2.4}$$

This is done using iterative processes, usually the Newton-Raphson [5], utilising the Jacobin $\mathbf{J}$ of partial derivatives

$$\mathbf{J} = \begin{vmatrix} \frac{\triangle h_1}{\triangle x_1} & \dots & \frac{\triangle h_1}{\triangle x_n} \\ \dots & \dots & \dots \\ \frac{\triangle h_1}{\triangle x_m} & \dots & \frac{\triangle h_m}{\triangle x_n} \end{vmatrix}.$$

The aim with these iterative processes is to minimise the difference between the individual estimated values of power flows and the measured ones, where the error (or line residual) r_p for real power is defined by

$$r_p = -P_{ij}^m + V_i^2 g_{ij} - V_i V_j g_{ij} \cos \triangle\theta_{ij} - V_i V_j b_{ij} \sin \triangle\theta_{ij}. \tag{2.5}$$

Iterative State Estimation

The iterative processes are performed as follows:

1. An initial guess for the current state is made (either using a flat start or based on using historical

data).

2. The power balance equations are solved using the guess and the difference between calculated and measured values in the residual error.

3. The system model is linearized around guessed states with this model used to calculate the next guess.

4. The power balance equations are recalculated using this next best guess, and the residual error is recorded again.

5. This is repeated until the stopping condition is reached. This condition is usually either an overall error level between measured and estimated values or an arbitrary number of iterations.

At the system level, the error check for final decision-making is based on the absolute value of the sum of errors known as the 2-norm difference between measured and estimated power flows, which is defined by

$$r = ||\mathbf{z} - \mathbf{h}\hat{\mathbf{x}}||_2. \tag{2.6}$$

Where the alarm limit τ is defined using engineering judgement, usually based on chi-squared testing criteria based on a 95% or 99% confidence interval derived via regression of previous residual values, such that an alarm is raised if $r > \tau$. Appropriate selection of the alarm limits and finding different ways to provide justification for weightings could provide a substantial body of research on its own. Generally, the measurement weighting comes from the variance and expected error of the telemetered measurement. We used a fixed variance and equal weighting across all our measurements. This allows us to basically ignore the problem. However, in practice system operators will likely consider a few factors. Standard error of the telemetered measurement vs the average size of the measurement. Timeliness of the measurement, i.e. how often do we get updates from this section of the grid, with upto date measurements being more valuable. Also, it is likely deciding the alarm limits themselves will not be arbitrary. The literature overwhelming supports using a 2-norm, 2 SD assumption representing a 95% confidence interval of system residual. However in practice this seems like a low bar for error

acceptance. For one, this means about 5% of measurement sets will be discarded or overwritten with pseudo-measurements. When you consider that also that occasionally the received measurement set will result in a non-convergent solution this means upwards of 5% non-state visibility on any given day. I think in practice operators will accept a larger bound for error although for our simulations we operate under this 2-norm assumption.

2.5.2 Linear Approximation

Occasionally, AC system models can suffer from issues of time complexity and non-convergence. Consequently, some of the simulations performed herein have been performed on the DC, linear approximation (with model extensions offered later). Therefore, we also derive here the linear approximation. The linear model can often be used for operation in real power systems due to its close approximation [63] under a few assumptions. These assumptions are as follows:

1. Line resistances are considered negligible compared to line reactance, i.e. $R_L << X_L$
2. The voltage profile is flat, and voltage amplitude is equivalent across all nodes
3. The voltage angle differences are small, which results in a linearisation of the sine/cosine elements in power flow equations

Under these assumptions $V_i \approx V_j$, the small-angle differences result in $\sin \triangle\theta_{ij} \approx \theta_{ij}$ and g_{ji} is much smaller than b_{ji}, leading to the linear approximation for line power flows of

$$P_{ij} = -b_{ij}\theta_{ij}. \tag{2.7}$$

For the power system as a whole, we consider the matrix formulation, represented by a linear regression model as a function of the Jacobian $\mathbf{H} \in \mathbb{R}^{m \times n}$ matrix and the state vector $\mathbf{x}$. Which (at a system level) translates to the linear approximation

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}. \tag{2.8}$$

Similar to the AC model, the state estimation problem is to find the best-fit estimate of $\hat{\mathbf{x}}$ corresponding to the measured power flow values of $\mathbf{z}$. Under the most widely used estimation approach, the state variables are determined by the minimisation of a WLS optimization problem as

$$\min_{x} J(\mathbf{x}) = (\mathbf{z} - \mathbf{H}\mathbf{x})^T \mathbf{W} (\mathbf{z} - \mathbf{H}\mathbf{x}). \tag{2.9}$$

$\mathbf{W}$ is a diagonal $m \times m$ matrix consisting of the measurement weights.

These weights can represent meter accuracy, reliability or simply engineering judgment about the relative importance of that particular measurement. Usually, the inverse of measurement error σ_1^{-2} is used

$$\mathbf{W} = \begin{vmatrix} \sigma_1^{-2} & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^{-2} & 0 & 0 & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & \sigma_m^{-2} \end{vmatrix}.$$

A solution for a minimal **J(x)** can be analytically obtained by taking the 1st derivative with respect to $\mathbf{x}$ and solving for 0, yielding $\hat{\mathbf{x}}$ defined by

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{z}. \tag{2.10}$$

The current approach in power systems operation for bad data detection is to use the 2-norm of the measurement residual with a detection threshold η [5]. The residual $\mathbf{r}$ is defined by the difference between the measured power flow values of $\mathbf{z}$ and the value calculated from the estimated state values $\hat{\mathbf{x}}$ and the known topology matrix $\mathbf{H}$

$$r = ||\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}||_2. \tag{2.11}$$

Similarly to the AC model an assumption is made that the errors of state variable $\mathbf{x}$ are random, independent and follow a normal distribution with mean zero and unit $\mathcal{N}(0,\sigma^2)$, a chi-squared distribution model $\chi^2_{m-n,\alpha}$ with $m-n$ degrees of freedom and confidence interval α (typically 0.95 or 0.99). can be used to define the detection threshold as

$$\eta = \sigma\sqrt{\chi^2_{m-n,\alpha}}. \tag{2.12}$$

If $r_t > \eta$, BDD alarms will trigger, and the system operator will discard the result, removing the elements from the residual calculation with large values and replacing them with an appropriate pseudo-measurement based on historical data.

2.5.3 Constructing Attack Vectors

In the case of an infinitely resourced and knowledgeable attack, the attacker has full access and control of the metering infrastructure and can change the measured power flows in almost any desired manner. In this case, it is trivial to design the attack to maintain a residual at a given value. The attacker can simply choose any linear combination of $\mathbf{Hc}$ (in the DC model) where $\mathbf{c} \in \mathbb{R}^{n\times 1}$ and provided $\mathbf{a} = \mathbf{Hc}$ will pass BDD. The vector $\mathbf{c}$ can be selected to have the desired impact on the state vector $\mathbf{x}$:

$$\mathbf{z}_a = \mathbf{z} + \mathbf{a} = \mathbf{z} + \mathbf{Hc}. \tag{2.13}$$

When this vector is injected, the 2-norm residual is shown as below:

$$r_a = \|(\mathbf{z}+\mathbf{a}) - \mathbf{H}(\hat{\mathbf{x}}+\mathbf{c})\|_2 = \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\|_2. \tag{2.14}$$

The residual under attack, r_a, will be equal to r as $\mathbf{a}$, and the $\mathbf{Hc}$ components will cancel because $\mathbf{a} = \mathbf{Hc}$.

These attacks can also be performed in the non-linear model. If the attacker has knowledge of how

the non-linear function is formed $\mathbf{h}(.)$, they can define a set of $\mathbf{x}$ and use the same iteration state estimation the system operator will use to replicate a plausible measurement set. In short, the attack vector will be defined by

$$\mathbf{z}_a = \mathbf{h}(\hat{\mathbf{x}} + \mathbf{c}). \tag{2.15}$$

Where $\mathbf{c}$ is an $n \times 1$ matrix denoting the desired bias injected into the system states (usually voltage angles) by the attacker and $\mathbf{z}_a$ denoting the desired attack vector profile of measurements. The residual under such an attack will therefore be defined by

$$r = ||\mathbf{z}_a - \mathbf{h}(\hat{\mathbf{x}} + \mathbf{c})||_2. \tag{2.16}$$

The attacker can ensure this value is close to 0 because the injected (measured) value has been designed specifically to equal the one estimated using these flows.

2.5.4 MTD through Topology Changes

Under AC state estimation, system measurements will consist of real power flows defined by (2.1) and reactive power defined by (2.2)

For the real power residual, error at the individual measurement level will be the difference between the measured flows and the estimated value from the system model such that the real power residual can be expressed as

$$r_{ij}^P = -P_{ij}^m + V_i^2 g_{ij} - V_i V_j g_{ij} \cos \triangle\theta_{ij} - V_i V_j b_{ij} \sin \triangle\theta_{ij}. \tag{2.17}$$

and reactive power flow residual can be expressed as

$$r_{ij}^Q = -Q_{ij}^m - V_i^2 (b_{ij} + b_{ij}^{sh}) + V_i V_j g_{ij} \cos \triangle\theta_{ij} - V_i V_j b_{ij} \sin \triangle\theta_{ij}. \tag{2.18}$$

In the AC state estimation model, MTD can employ resistive as well as inductive components to introduce change. We don't explore resistive MTD in this work (and we have yet to see it suggested seriously as a method of MTD).

2.5.5 Linear Approximation for MTD

We derive here the analytical expression of the impact on the residual of the topology change for a linear system under attack vector $\mathbf{a} = \mathbf{Hc}$. Using the WLS formulation, r_a can be expressed as

$$r_a = \|(\mathbf{z} + \mathbf{Hc}) - \mathbf{H}(\mathbf{H}^T\mathbf{WH})^{-1}\mathbf{H}^T\mathbf{W}(\mathbf{z} + \mathbf{Hc})\|_2. \tag{2.19}$$

The attacker is assumed to have static topology knowledge and to construct the injected attack vector $\mathbf{z}_a$ as a function of the original topology $\mathbf{H}_o$. The new topology with MTD applied is $\mathbf{H}_n$, which is only known by the SO. As a result, the measurement vector under attack $\mathbf{z}_a$ will be

$$\mathbf{z}_a = \mathbf{z} + \mathbf{H}_o\mathbf{c}. \tag{2.20}$$

The SO estimates $\hat{\mathbf{x}}$ via the WLS minimisation using the visible $\mathbf{z}_a$ and $\mathbf{H}_n$. The min error estimate of $\hat{\mathbf{x}}_n$ will utilize the new topology $\mathbf{H}_n$, whereas the attack vector is developed based on the old topology $\mathbf{H}_o$. Consequently, the new residual will be a product of the attack vector based on old topology $\mathbf{H}_o\mathbf{c}$, and the WLS estimation will be based on the new topology as

$$r_n = \|\mathbf{z} + \mathbf{H}_o\mathbf{c} - \mathbf{H}_n(\mathbf{H}_n^T\mathbf{WH}_n)^{-1}\mathbf{H}_n^T\mathbf{W}(\mathbf{z} + \mathbf{H}_o\mathbf{c})\|_2. \tag{2.21}$$

Defining the WLS minimisation factor for the new topology as $\mathbf{F}_n$, which is fixed for a given topology as $\mathbf{F}_n = (\mathbf{H}_n^T\mathbf{WH}_n)^{-1}\mathbf{H}_n^T\mathbf{W}$, the residual 2-norm can be rewritten as

$$r_n = \|\mathbf{z} + \mathbf{H}_o\mathbf{c} - \mathbf{H}_n\mathbf{F}_n(\mathbf{z} + \mathbf{H}_o\mathbf{c})\|_2. \tag{2.22}$$

Considering the old topology $\mathbf{H}_o$ as a function of the new and system change $\mathbf{H}_n + \triangle\mathbf{H}$, the residual in terms of the new topology can hence be calculated as

$$r_n = \|\mathbf{z} + (\mathbf{H}_n + \triangle\mathbf{H})\mathbf{c} - \mathbf{H}_n\mathbf{F}_n(\mathbf{z} + (\mathbf{H}_n + \triangle\mathbf{H})\mathbf{c})\|_2. \tag{2.23}$$

$\mathbf{H}_n\mathbf{F}_n\mathbf{H}_n$ is the idempotent matrix of $\mathbf{H}$ and therefore $\mathbf{H}_n\mathbf{F}_n\mathbf{H}_n\mathbf{c} = \mathbf{H}_n\mathbf{c}$ the expression can be rearranged into

$$r_n = \|(1 - \mathbf{H}_n\mathbf{F}_n)\mathbf{z} + (1 - \mathbf{H}_n\mathbf{F}_n)\triangle\mathbf{H}\mathbf{c}\|_2. \tag{2.24}$$

With the WLS min factor in terms of the original topology

$$\mathbf{F}_n = ((\mathbf{H}_o - \triangle\mathbf{H})^T\mathbf{W}(\mathbf{H}_o - \triangle\mathbf{H})^T)^{-1}((\mathbf{H}_o - \triangle\mathbf{H}))\mathbf{W}. \tag{2.25}$$

The weighting factor $\mathbf{W}$ cancels and thus expands to

$$\mathbf{F}_n = (\mathbf{H}_o^T\mathbf{H}_o - 2\triangle\mathbf{H}^T\mathbf{H}_o + \triangle\mathbf{H}^T\triangle\mathbf{H})^{-1}(\mathbf{H}_o - \triangle\mathbf{H})^T. \tag{2.26}$$

Because $\triangle\mathbf{H}$ will be significantly smaller than the overall system such that $\triangle\mathbf{H} << \mathbf{H}$, we can approximate the gain matrix of $\triangle\mathbf{H}$ as $\triangle\mathbf{H}^T\triangle\mathbf{H} = 0$ and consequently $\mathbf{F}_n \approx (\mathbf{H}_o^T\mathbf{H}_o - 2\triangle\mathbf{H}^T\mathbf{H}_o)^{-1}(\mathbf{H}_o - \triangle\mathbf{H})^T$, resulting in the expanded formulation

$$\begin{aligned} r_n = \|\mathbf{z} + \mathbf{H}_o\mathbf{c} - (\mathbf{H}_o - \triangle\mathbf{H}_o)(\mathbf{H}_o^T\mathbf{H}_o \\ - 2\triangle\mathbf{H}^T\mathbf{H}_o)^{-1}(\mathbf{H}_o - \triangle\mathbf{H})^T(\mathbf{z} + \mathbf{H}_o\mathbf{c})\|_2. \end{aligned} \tag{2.27}$$

As shown in (2.24), any $\triangle\mathbf{H}$ will change the residual value $\mathbf{r}_n$. The aim of the defender is to select a value for $\triangle\mathbf{H}$ such that under attack vector $\mathbf{H}_o\mathbf{c}$, the new residual exceeds the alarm criteria (usually chi-squared criteria) $r_n > \sigma\sqrt{\chi^2_{m-n,\alpha}}$. The question of how to do this in the most efficient manner

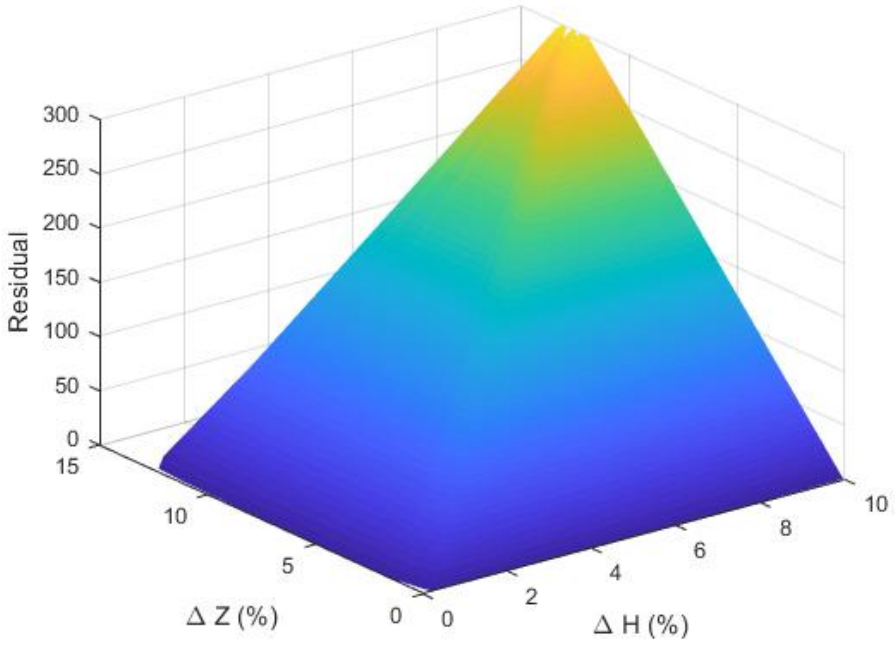

Figure 2.1: Residual values generated in MATLAB under WLS state estimation for as a function of the change applied to power flow $\mathbf{z}$ and overall change in topology $\triangle\mathbf{H}$ for a single branch.

is not always clear. The post-MTD residual is dependent on the system topology, size of the attack vector and current residual level of the power flow set. The impact of this is shown as a single line in Figure 2.1. Essentially, both an attack vector change and MTD are needed to create a residual change. Larger attack vectors require smaller MTD implementations to be effective. Some papers explore the concept of 'complete' MTD. However, it is quite difficult to assess whether an MTD is complete without the attack vector being known beforehand.

In the next chapter, we explore topology learning aided FDI attacks and show how a 'hidden topology' knowledge assumption can be invalidated.

Chapter 3

Topology-Learning-Aided False Data Injection Attack without Prior Topology Information

In this chapter, we develop a topology-learning-aided FDI attack that allows stealthy cyber-attacks against an AC power system state estimation without prior knowledge of system information. The attack combines topology-learning technique, based only on branch and bus power flows, and an attacker-side pseudo-residual assessment to perform stealthy FDI attacks with high confidence. This paper, for the first time, demonstrates how quickly the attacker can develop the full knowledge of grid topology and parameters and validates the full knowledge assumptions in the previous work.

The work outlined in our published conference paper [64] made up the basis of this chapter.

3.1 Introduction

3.1.1 Topology Discovery

Attacks that can learn the underlying system topology offer more flexibility in how targets are chosen. There have been attempts to develop topology discovery-style attacks. However, these have largely been done under the assumptions of the linear model, such as in [65]. This may be because up until recently, the capabilities for full branch topology discovery in the AC model have not been sufficient to capture both the network and branch incidence without assistance from PMU measurements. While PMUs are becoming more common in the power system, they are still relatively costly and comparatively sparse compared with power flow and voltage measurements. Voltage measurements themselves can be used to evaluate certain areas of the grid. In [66], voltage correlations are used to identify bus incidence. However, branch values are not calculated by the authors. This leaves a large portion of the required topology matrix unknown, and in practice, this is insufficient information for an FDI attack. In [67], a test is developed for estimating the dynamic Jacobin in the presence of topology changes. However, this method requires PMU measurements that are not always available to the attacker. Similarly, in [68] and [69], models for network parameter estimation are suggested. But again, these require PMU data to build an accurate model of the power system. As shown in [70], it possible to evaluate network branch parameters without PMU data. An initial approximation can be made using regression via matrix operations that give a quick approximation of the per unit network topology. This is then used as a starting point over which a fine identification is run. The fine identification uses a modified Newton-Raphson to get high-quality per-unit estimations of network topology. The closest paper in terms of contribution to this work is [65] which maps topology with FDI attacks for the DC model. However, we improve on this work in a few ways which we outline below.

3.1.2 Novel Contributions

The topology-learning technique is combined with an attacker-side pseudo-residual assessment to create a topology-learning-aided FDI attack (TL-FDIA), which has the capabilities of a full knowledge attack with no prior system knowledge requirements. Our contributions are outlined as below:

- A TL-FDIA capable of attacking power systems under a blind assumption model (no branch or network incidence information available). The attack is committed against an AC power system and uses the latest state-of-the-art topology discovery techniques to build a model for the network.
- We introduce an attacker-side criteria assessment via a pseudo-residual calculation to allow a probabilistic assessment of attack success before any attack committed, allowing the attacker to ensure stealthiness. We also show that regional pseudo residuals can be used to verify local attacks, even in the presence of global topology errors.
- We demonstrate how quickly the attacker can develop the full knowledge of system topology and parameters and effectively invalidate the full system knowledge assumptions in previous studies.

3.2 Topology Learning Aided FDI Attacks

3.2.1 Full Knowledge AC FDI Attack

If an attacker has knowledge of how the non-linear function is formed $\mathbf{h}(.)$ they can define a set of $\mathbf{x}$ values to achieve his stated aims in terms of P and Q such that

$$\mathbf{z}_a = \mathbf{h}(\hat{\mathbf{x}} + \mathbf{c}). \tag{3.1}$$

Where $\mathbf{c}$ is an $n \times 1$ matrix denoting the desired bias injected into the system states (usually voltage

angles) by the attacker and $\mathbf{z}_a$ denoted the desired attack vector profile of measurements. The residual under such attack will therefore be defined by

$$r = ||\mathbf{z}_a - \mathbf{h}(\hat{\mathbf{x}} + \mathbf{c})||_2. \tag{3.2}$$

The attacker can ensure this value be close to 0 as the injected (measured) value has been designed specifically to equal the one estimated using these flows. In practice however, it is unlikely that an attacker will have the required knowledge for FDI attack as this information will rarely be available publicly or intentionally hidden (in fact it is possible the system operator themselves may not always have a perfect picture of the underlying system).

3.2.2 Attack Assumption and Overview of TL-FDIA

For TL-FIDA, we operate from the assumptions usually present in the blind attack models [71] as below:

- The attacker has reading access to all measurements and can alter full or certain real and reactive power measurements in the system.
- The attacker has no knowledge of system interconnection or branch admittance/resistance values.

Keeping these assumptions in mind, the attacker will need to create a model of the power system only from the available power flow measurements. Once the attacker has gained access to the system, the algorithm enters a period of data collection. When sufficient data are received, the attacker attempts to perform the topology-learning step of the attack based on the received data. This allows the attacker to subsequently perform an attacker-side state estimation to verify the accuracy of the model using the derived topology. If the proposed vector passes the pseudo-state estimation residual check, the attacker can then proceed to attack phase. If not, the attacker waits for additional data and reruns the topology-learning step of the attack. This proposed flow is outlined in figure 3.1.

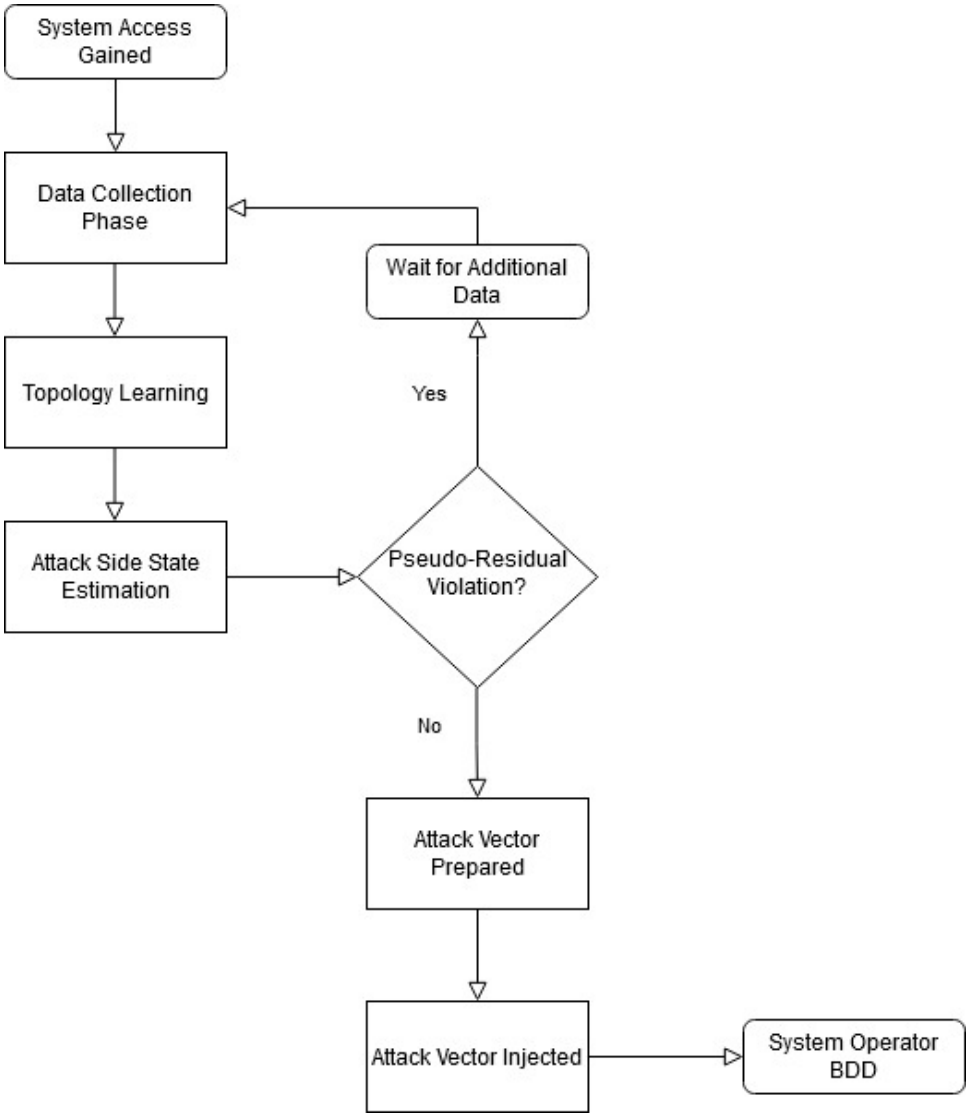

Figure 3.1: Overview of the Topology-Learning FDI Attack Algorithm Implementation.

3.2.3 Attacker-Side Verification

Compared with the full knowledge attack, an important consideration for TL-FDIA is to know when they have collected enough data and are ready to attack. This can be difficult because the attacker has zero prior information and no access to the system operator residual data, so any indication as to whether the proposed attack vector may pass BDD is based only on inbound new measurements. Consequently, we propose an attacker-side pseudo-residual calculation as an assessment on whether the attack can proceed based on

$$r_p = ||\mathbf{z}_a - \hat{\mathbf{h}}(\hat{\mathbf{x}} + \mathbf{c})||_2. \tag{3.3}$$

Where $\hat{\mathbf{h}}$ is the estimated non-linear transformation function based on the estimated topology values and state measurements themselves.

3.2.4 Sub-Graph Residual

In fact, even in the presence of global residual errors, the attacker may be able to identify sub-graphs within the network where he can attack without altering other regions with poor residual performance.

In practice, this will be similar to the incomplete information-type attacks in [14]. Areas of high regional residual will be assumed to have incomplete knowledge, and other lower-error regions can be attacked. Therefore, the attacker can use the regional meter error given by a triggered alarm defined by

$$r_p^m = \mathbf{z}^m - \mathbf{z}_{est}^m > \tau^m. \tag{3.4}$$

With respect to the FDI attack, the sub-graphs are given by the number of non-zero terms in the column vector of the topology matrix $\mathbf{H}$ for a given node n or the Jacobian $\mathbf{J}$ in the non-linear model. The attacker can then identify the corresponding sub-graphs related to this branch using the network incidence matrix $\mathbf{I}$. For meter number m,

$$col_n(\mathbf{I}_m) = \begin{cases} 1 & \text{meter } m \text{ is part of bus } n \text{ subgroup} \\ 0 & \text{meter } m \text{ is not part of the subgroup.} \end{cases} \tag{3.5}$$

The attacker then can structure the state adjustment vector $\mathbf{c}$ such that $\mathbf{I}_{n,m} = 0$. These regional residual calculations offer an opportunity to the attack, especially considering the context of localised MTD. If a local attack is being shown to increase the residual in some areas and not others the attacker can target his attack point based on the lower residual increase and hence increase the chance of staying hidden.

3.3 Blind Topology Identification

For the initial topology identification, we employ the method outlined in [70]. The aim of topology identification is to identify the network incidence as well as the branch values for the conductance and susceptance matrices $[G_{ij}]$ & $[B_{ij}]$. The technique we have employed here was originally used to map distribution networks for system operators. However, we turn this method against the system operator for purposes of the FDI attack where the attacker has limited system information. The method outlined by Zhang *et al* utilises a two-step identification process to identify per-unit branch topology information. This process proceeds as follows:

1. To start the algorithm takes as its inputs the nodal active/inactive power loads in addition to the nodal voltage magnitudes.

2. The algorithm then performs and initial linear regression to calculate approximations of conductance and susceptance.

3. Various noise reduction algorithms are then used which prune proposed branches which are unlikely to be part of the system topology.

4. The fine identification, pseudo-power power with integrated branch values then begins which takes the initial step approximation as a starting point in order to reduce processing time.

5. This outputs a highly accurate presentation of the line parameters and the voltage angles.

The initial regression uses a linearised approximation of the relationship between branch parameters, voltages and real and reactive powers to create a basis of initial approximation.

We can consider the matrix formulation for this in the terms

$$[P/V] = G^{\#}_{ij}[V], \tag{3.6}$$

$$[Q/V] = -B^{\#}_{ij}[V]. \tag{3.7}$$

$B^{\#}_{ij}$ & $G^{\#}_{ij}$ are approximations of the real and imaginary branch elements. These are based on the assumption of small state angle differences under the standard equations for real and reactive power injection. Given this approximation, the branch components can be extracted using matrix operations from which a solution for the approximate is the following:

$$G^{\#}_{ij} = [P/V][V]^T([V][V]^T)^{-1}, \tag{3.8}$$

$$B^{\#}_{ij} = [Q/V][V]^T([V][V]^T)^{-1}. \tag{3.9}$$

These initial steps are basic matrix operations. This means they can be performed quickly and with limited computational power. Under a DC approximation, these can give reasonable approximations of the network branches incidences on their own. Under the method proposed by Zhang *et al*, they provide an initial approximation for the network topology that is used as the starting point for the fine identification stage. This is then followed by a modified Newton-Raphson, that incorporates the branch topology values to refine the approximation. Given the power system bus injections under polar coordinates as

$$\begin{bmatrix} \triangle p \\ \triangle q \end{bmatrix}_{1\times 2n} = \begin{bmatrix} \frac{\triangle p}{\triangle g} & \frac{\triangle p}{\triangle g} & \frac{\triangle p}{\triangle g} \\ \frac{\triangle p}{\triangle g} & \frac{\triangle p}{\triangle g} & \frac{\triangle p}{\triangle g} \end{bmatrix} \cdot \begin{bmatrix} \triangle g \\ \triangle b \\ \triangle \theta \end{bmatrix}_{1\times(2m+n-1)} . \tag{3.10}$$

where g & b are conductance and susceptance of m branches. A pseudo-power flow calculation is performed where the generalised inverse is then applied to solve for the difference in both topology and state angle, such that

$$\begin{bmatrix} \triangle g \\ \triangle b \\ \triangle \theta \end{bmatrix} = \begin{bmatrix} \frac{\triangle P}{\triangle G} & \frac{\triangle P}{\triangle G} & \frac{\triangle P}{\triangle G} \\ \frac{\triangle P}{\triangle G} & \frac{\triangle P}{\triangle G} & \frac{\triangle P}{\triangle G} \end{bmatrix}^{+} \cdot \begin{bmatrix} \triangle P \\ \triangle Q \end{bmatrix} . \tag{3.11}$$

This is used in the usual additive process to derive an estimation for the topology.

$$\begin{bmatrix} g \\ b \\ \theta \end{bmatrix}^{(k+1)} = \begin{bmatrix} g \\ b \\ \theta \end{bmatrix}^{k} \cdot \begin{bmatrix} \triangle g \\ \triangle b \\ \triangle \theta \end{bmatrix} . \tag{3.12}$$

A full outline of the applied algorithm for the topology discovery component can be found in [70].

3.4 Results & Analysis

This section assesses the performance of the proposed attack on the IEEE 14-bus system. To replicate the real-time operation of a power system as close as possible, system loads have been simulated using mock load profiles. An example of variation across the sample base can be seen in 3.2. All simulations were implemented using the MATPOWER toolbox in MATLAB [72]. The Newton-Raphson AC state estimation module was used from MATPOWER to replicate the system operator state estimation. Simulations were performed using Intel Core i7-7820X CPU with 64 GB of RAM running on a Windows 10 system.

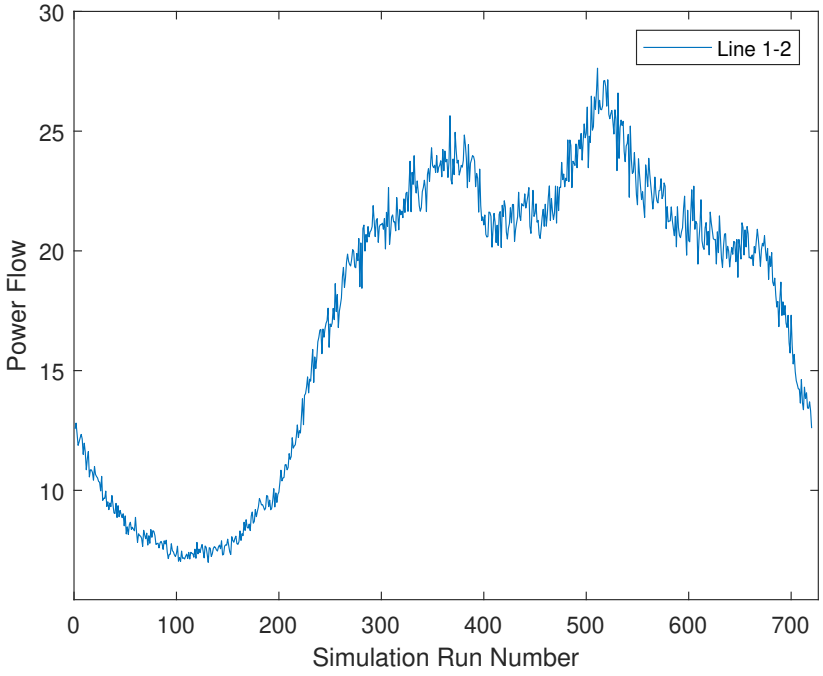

Figure 3.2: Power flow profile across system measurements.

3.4.1 Effectiveness of TL-FDIA

We attempt to model herein the FDI attack as applied to the IEEE 14-bus power system. We attempt a branch loading change attack. In figure 3.3, we model the system residual as we apply the TL-FDIA using a voltage angle bias of around 15% to bus 1. This results in a power flow changes across bus 1-2 and 1-5. In blue, we see the occasional residual spike past that of the acceptable alarm limit. In green, we show the attacker's pseudo-residual calculation, which, as expected, mimics the SO residual. As we shown previously, an attacker is likely to be measured in how he attacks a system and will try and leverage his resources to remain hidden. If he has sufficient time it makes sense that the attack will employ pseudo-residual style error checker (on the attackers side) in order to minimise the chance of discovery of their attack vector. We implement a pseudo-residual style attack to reflect his. As discussed, this pseudo residual is used by the attacker as attacker-side assessment criteria. Without the pseudo residual, we see around 80% success in attack with 20% of values exceeding the SO residual. However, by applying the residual check, we decrease this error significantly. In figure 3.4, we implement the pseudo residual as a decision statistic with the attacker choosing not to attack if they believe the residual will be violated. We note that the residual stays below the acceptable level

and hence avoids detection. This implies that system operators should be cognizant that motivated and intelligent attackers will likely be able to know when an attack is likely to be unsuccessful. Plausible, attacker side FDI assessment criteria have implications for long term intrusions. The attacker no longer has to take a risk on detection on the belief that his model is 'good enough' to pass BDD. He can simply wait until his shows adequate levels of risk acceptability and then launch his attack. Going forward, system operators should consider how then can invalidate the attackers knowledge. Some works have already started to address this. Tian *et al* offer a solution which hides system changes in the power profile [48]. Methods such as these could potentially mitigate the attackers ability to make a pseudo-residual calculation.

3.4.2 Data Requirement for TL-FDIA

The previous successful attacks were performed with the assumption that 720 pieces of measurement data are available. However, the attacker may want to be ready to attack as soon as possible to avoid accidental exposure, and hence, it is critical to understand the minimum of measurement data that allows such an attack. In figure 3.5, we show how the pseudo residual decreases with additional available data points. It is clear that with the increasing amount of measurement data, the system residual declines quickly. In the case of the 1% noise alarm and 14-bus system, the attacker will likely wish to wait until at least 200 points are available before attempting to build a map of the system because otherwise, the model is likely to contain enough gross errors to flag an alarm.

To further illustrate the data availability and readiness to attack, we simulate a time-domain attack scenario starting from the moment the attack gain access to the full meter measurement to when the attack is ready to launch. The timing of attacks is a crucial consideration. The longer an attacker spends in the system the higher his chance of discovery and expulsion from the system will be. There is also a chance the vulnerabilities the attacker is using will be patched/removed. Consequently, the attacker will seek to minimise his time in the system before attacking. In Figure 3.6, we show how as individual measurements come in, the system residuals from both operator and attacker changes over time as additional measurements are available to the attacker. We assume the attacker has access to measurements in a similar frequency to the central system operator with state estimation and mea-

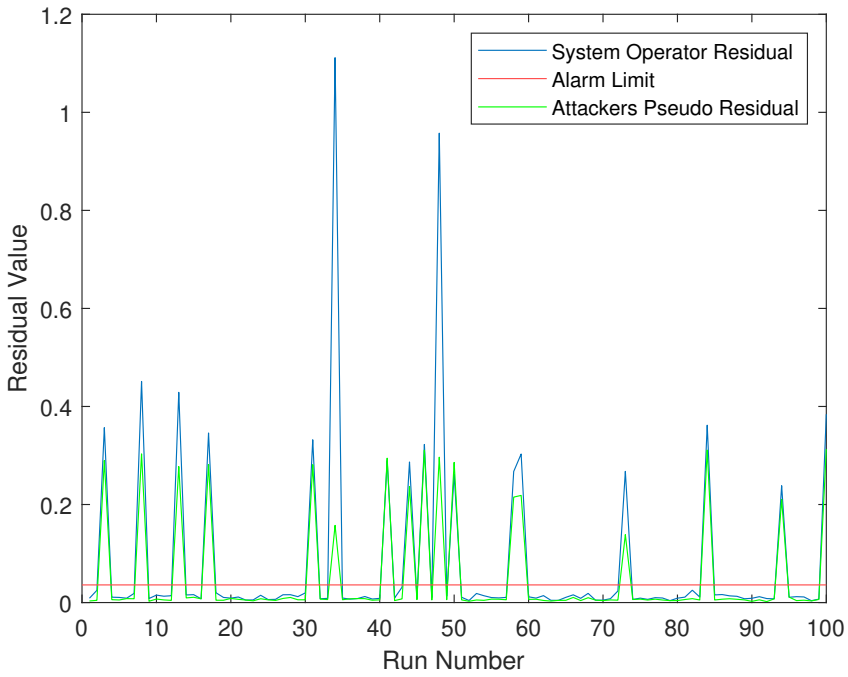

Figure 3.3: Residual value measured by system operator and attacker in presence of TD FDI attack against 1% equivalent alarm.

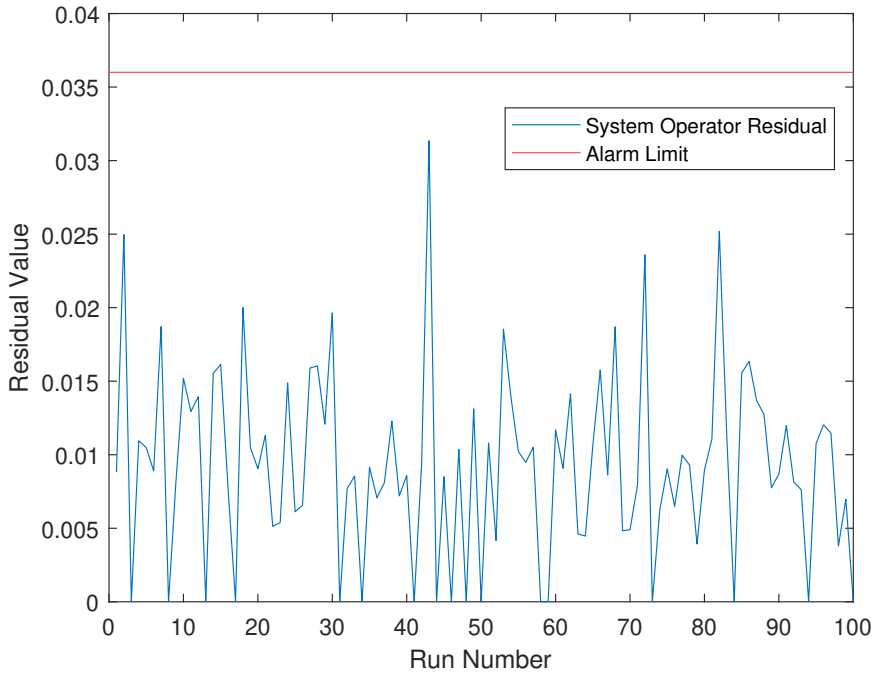

Figure 3.4: Residual value measured by system operator TD FDI attack with a pseudo-residual decision statistic considered against 1% equivalent alarm.

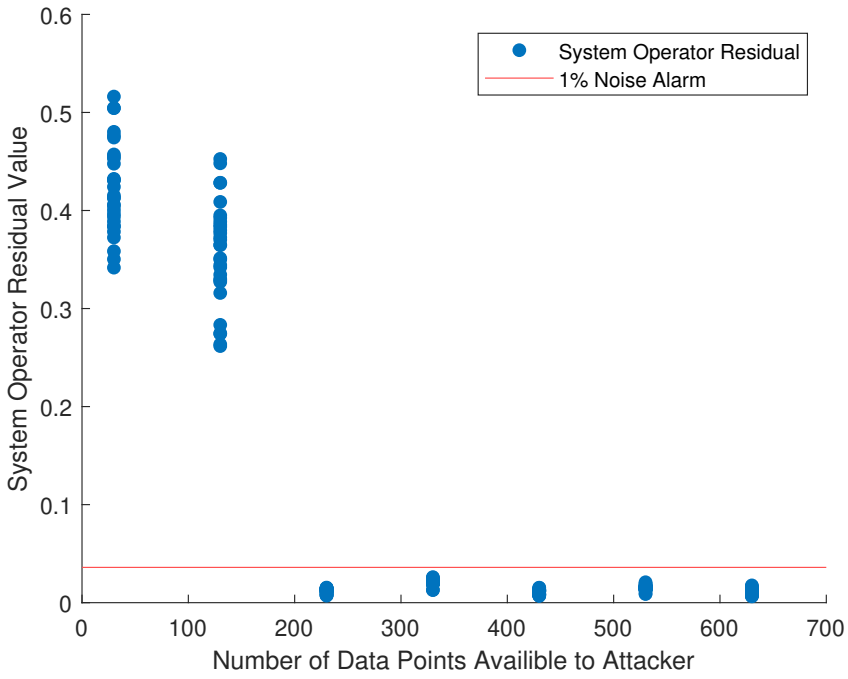

Figure 3.5: Residual value measured by the system operator in the presence of TD-FDI attack with an increasing number of available data points.

surement set received once every minute. First, it is important to note that once the attacker's pseudo residual converges to an acceptable level, the operator residual is also approaching to the level that will pass BDD, which demonstrates the effectiveness of an attacker-side pseudo-residual assessment. In addition, the figure suggests that the attacker will need about 3-4 hours of data collection before they can initiate the attack without detection. Compared with the length process to reconnaissance and penetrate the system, such duration is almost negligible, which validates the full knowledge assumptions in many previous works. Because there is a cost associated with MTD, the system operator may want to consider this min attack time in the way they implement MTD and other intrusion detection capabilities. The amount of time required to create a stealthy vector may also be increased further with additional MTD configurations effectively requiring whole new topology calculations.

3.5 Summary

In this chapter, we consider how attackers might utilise state-of-the-art topology learning techniques in order to validate the assumption of the full knowledge FDI attack. We utilise an established method

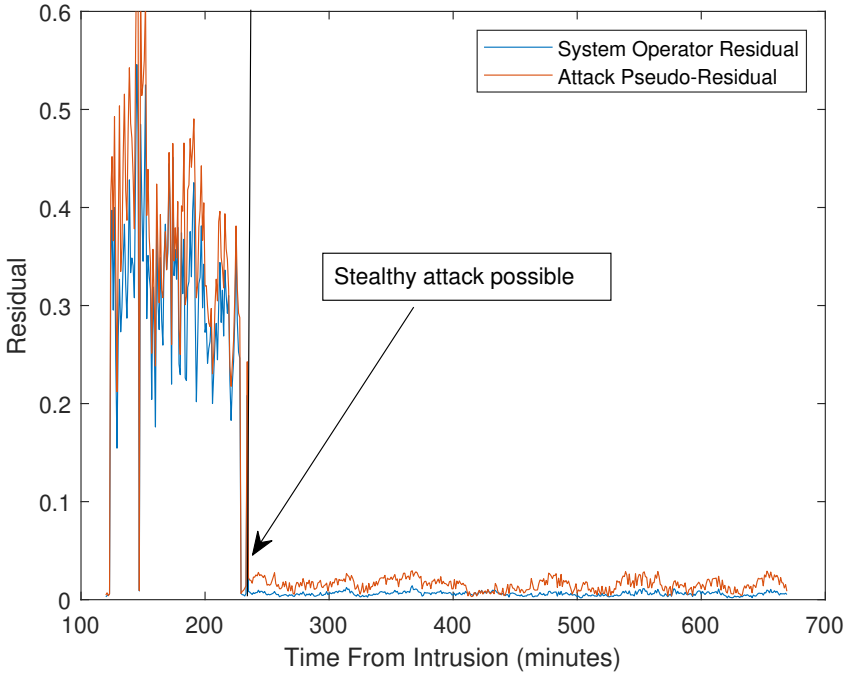

Figure 3.6: Residual value measured by the system operator in the presence of TD-FDI attack. The X-axis shows the number of minutes since intrusion.

for topology discovery which was previously seen as applied to distribution systems in [70]. This technique requires only available power flow information and does not require PMUs for topology discovery. From this we compose the Topology Learning FDIA (TL-FDIA). We show that this TL-FDIA can indeed attack power systems stealthily but will occasionally be discovered (around 20% of the time). As a result, we suggest combining this TL-FDIA with an attacker side attack readiness criteria via calculation of a pseudo-residual. This combined approach allows for high levels of success when we attack. We also learn that it is possible for an attacker to use incoming system measurements to verify his attacking model and hence refine and improve his attack success rate. Previously, blind style attacks have usually been performed under the linear assumption model. This is due to the difficultly in recreating a workable attack model with the underlying non-linearity caused by AC transmission systems. However, we were able to show via simulations on a 14-bus system that such an attack allows the attacker to perform a full knowledge AC-FDI attack under the blind assumption model. We also learn that it is indeed possible for an attacker to gather the necessary data in only a short period of time, namely a few hours. This is crucial going forward as it validates many assumptions about attacker capabilities made in other works; namely that they can get sufficient system knowledge to attack within a reasonable intrusion period.

3.6 Lessons Learned

Crucially, we draw the following lessons from this chapter, which we use to inform our research and bring us to the next chapter in this thesis.

- It is possible, given state-of-the-art topology discovery capabilities to perform a a TL-FDIA capable of attacking power systems under a blind assumption model (no branch or network incidence information available). Given this, we can validate many of the previous assumptions made about FDI attacks, namely; that it is plausible for an attack to have full knowledge of the system prior to attack.
- Intelligent attackers can use their generated models to assess attack strength and model whether their attack is likely to succeed. By utilising a pseudo-residual style calculation an attacker can verify the likelihood of an attacks success before an attack has even been committed. System operators should be aware of this and should take steps to attempt to prevent attacker side attack assessment.
- These attacks can be committed quickly (in a matter of hours) post-intrusion to the network system and with limited prior information. In fact, we show specifically that for the 14-bus system only around 250 measurement data are needed in order to construct a workable model.
- Considerations of defence against FDI attacks are important and need to be factored into both the attacking and the defensive models and we consider this in the next chapter.

In the next chapter, we start to explore MTD as a method of defence against FDI attacks to protect against these topology-learning attacks and hence invalidate the assumption of an attacker having full system knowledge.

Chapter 4

Stealthy MTD against Unsupervised Learning-Based Blind FDIAs in Power Systems

This chapter examines how moving target defence (MTD) implemented in power systems can be countered by unsupervised learning-based false data injection (FDI) attack and how MTD can be combined with physical watermarking to enhance the system resilience. A novel intelligent attack that incorporates dimensionality reduction and density-based spatial clustering is developed and shown to be effective in maintaining stealth in the presence of traditional MTD strategies. In resisting this new type of attack, a novel implementation of MTD combining with physical watermarking is proposed by adding Gaussian watermark into physical plant parameters to drive detection of traditional and intelligent FDI attacks while remaining hidden to the attackers and limiting their impact on system operation and stability.

The work outlined in our published journal paper [38] made up the basis of this chapter.

4.1 Introduction

As we showed in the previous chapter, we can not take the 'hiddenness' of the network system topology for granted. Therefore, it is important to explore how MTD can be used to replicate hiddenness and invalidate the attackers system knowledge assumption. It is worth considering why one might focus on using a physical-type defence such as network switching or watermarking over cyber defences such as encryption of telecommunications. Some notable advantages of applying protection at the physical over the cyber layer are as follows:

1. The physical layer comes first in the data collection process. Therefore, an attacker can bypass any need to attack the cyber layer by attacking the continuous measurements at the physical layer.

2. There can be significant processing time costs associated with encrypting large numbers of independent sources of data through techniques such as RSA. These could introduce delays in the state estimation process and slow the ability of the network operator to act.

3. Encryption can also be broken in polynomial time with Shor's algorithm. New techniques such as quantum computing appear set to reduce the time to break such codes further.

In an ideal scenario, a network operator would undertake both physical and cyber security to maximize defence against cyber-physical threats. However, as we show in this chapter simply applying MTD within the power system does not guarantee system security.

4.1.1 Novel Contributions

Where previously blind attacks have been based on static systems, we show dynamic systems can indeed also be attacked providing certain system conditions are met. An intelligent FDI attack with limited system knowledge assumptions is proposed to counter MTD. We introduce an extension to the existing blind attacks and replay-style attacks under AC, which create resilience to MTD defences. In addition, A new type of MTD in the form of a hidden key Gaussian watermark is proposed. This

method mimics the underlying noise of the system in profile and magnitude to stay hidden. The watermark is combined with cumulative error monitoring to spot minor but sustained changes in the system. We show that although most existing MTD techniques can be identified in the profiles of power flows, our proposed technique can indeed remain hidden. The closest work in the literature (to which this author did not contribute) to this chapter is [48] which also offered a camouflage form of MTD. However, we differ from this work in a number of ways.

- On the attacking front, this work introduces a novel counter-MTD technique. Where previous FDI attacks have been designed against static systems, we seek to offer new attacking considerations in the presence of dynamic systems with MTD. The proposed intelligent attack under zero system knowledge assumption combines dimensionality reduction and unsupervised learning to identify underlying clusters associated with network topology and design the corresponding attack vector. The method is shown to be effective and stealthy against traditional MTD.
- From the defensive perspective, we introduce a new implementation of MTD to drive detection against traditional and intelligent FDI attacks. The proposed defence strategy combines MTD and the physical watermarking concept [73] for the first time to add a Gaussian watermark into the physical plant parameters. Because the added watermark mimics the underlying noise of the system, the physical changes driven by MTD stay hidden. The physical watermarking is combined with cumulative error monitoring to spot minor but sustained changes in the system to trigger alarm. We demonstrate that the proposed MTD techniques are more resilient against intelligent FDI while inducing less impact on the system operation.
- Where previously blind attacks have been based on static systems, we show dynamic systems can also be attacked providing certain system conditions are met.

4.1.2 Constructing Attack Vectors

As we discussed previously, in the case of an infinitely resourced and knowledgeable attack, the attacker can gain full access to the metering infrastructure and change measured power flows in any

desired manner. In this case, it is trivial to design the attack to maintain a residual at a given value. The attacker can choose any linear combination of **Hc** where $\mathbf{c} \in \mathbb{R}^{n\times 1}$. The vector **c** can be selected to have the desired impact on the state vector **x**:

$$\mathbf{z}_a = \mathbf{z} + \mathbf{a} = \mathbf{z} + \mathbf{Hc}. \tag{4.1}$$

In a more realistic scenario, where the attacker has full access to the metering infrastructure but no understanding of how the network components interconnect or the branch admittance, the attacker has to commit a 'blind' form of attack by estimating plausible attack vector models based on historical measurements. One way of achieving this is to utilize blind source separation (BSS) techniques. This scenario has been outlined previously in [15]. The relationship between the state variables in a power system and latent independent variables **y** under a fixed topology **H** can be described by

$$\mathbf{x} = f(\mathbf{H}, \mathbf{y}). \tag{4.2}$$

In practice, **y** represents the loads of power system that vary independently while the topology is fixed, but other underlying latent variables may exist for some systems. The state vector x can be approximated as the first-order coefficient of the Taylor expansion **A** around y.

$$\mathbf{x} \approx \mathbf{Ay}. \tag{4.3}$$

Returning to the state estimation problem, the system states can then be expressed in terms of load such that

$$\mathbf{z} \approx \mathbf{HAy} + \mathbf{e}. \tag{4.4}$$

If the attacker can acquire **HA**, an attack vector can be constructed with a value selected for a change in power flows $\delta\mathbf{y}$ shown by

$$\mathbf{z}_b = \mathbf{z} + \mathbf{HA}\delta\mathbf{y}. \tag{4.5}$$

$$\mathbf{u} = \mathbf{Gv}. \tag{4.6}$$

A generalized form of blind source separation $\mathbf{u} = \mathbf{Gv}$ can be used, where $\mathbf{u}$ is the vector that can be directly observed, $\mathbf{G}$ is the fixed vector known as the mixing matrix and $\mathbf{v}$ is the underlying vector of signals. The state estimation can be constructed in an equivalent manner such that:

$$\mathbf{z} = \mathbf{HAy} = \mathbf{Gy}. \tag{4.7}$$

Provided the errors follow a Gaussian distribution and do not contain gross errors, $\mathbf{HA}$ can be extracted using independent component analysis as shown previously in [15] [71].

4.1.3 AC Extension of Blind Attack

Similar to the DC attack, AC-FDI attacks must satisfy the system model to remain hidden such that

$$\mathbf{z}_a = \mathbf{z} + \mathbf{a} = \mathbf{h}(\mathbf{x} + c). \tag{4.8}$$

This can be done without topology information, either by using the geometric approach [74] or using a historical measurement-based replay approach. *Chin et al* showed that where the vector angle between the normal power flows and attacking vector was defined by

$$\mathbf{z}^T\mathbf{a} = cos(\psi) \tag{4.9}$$

the attack can bypass AC detection, provided the vector space angle between the attacking vector and measurement vector was close to zero such that

$$\mathbf{z}^T \mathbf{z}_a = 1. \tag{4.10}$$

Under these considerations, a regression model can be extracted to attack the system. Alternatively, in the case of limited information, the attacker can implement a replay-style attack that reuses a previous vector from historical measurements such that

$$\mathbf{z}_t^a = \mathbf{z}_{t-q} \tag{4.11}$$

where q is used to denote a vector from a previous time period. Our AC simulations were built with this replay case in mind, but it should be noted that both methods are susceptible to conventional MTD.

4.2 Clustering to Circumvent MTD

The implementation of MTD will cause a change in $\mathbf{z}$, which can be used by an attacker to identify the existence of MTD and predict a change in the network topology. If the attacker uses data sourced from multiple MTD configurations to resolve a blind static model the attack model will contain gross errors. This section investigates how the data-driven approach can be applied to explore the vulnerability of existing MTD. In particular, an efficient method is proposed to identify changes in the network caused by the implementation of D-FACTS or switching through analysing the resultant power flow profiles. By doing so, the attacker can ensure only data points corresponding to the current configuration are used to create the blind attack. The proposed attack flow is as follows:

1. Observations of historical power flows are clustered into groups.
2. The clustering algorithm identifies the current power flow set to find corresponding measurements for the attack model.
3. The blind attack model is developed using only the data corresponding to the current power flow profile cluster.

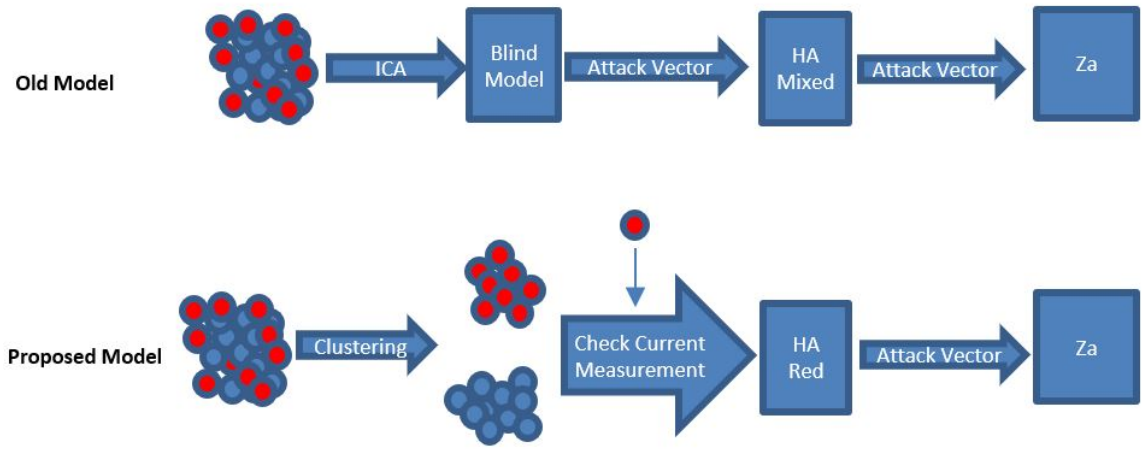

Figure 4.1: Proposed algorithm process to circumvent MTD. Red and blue points corresponded to observed power flows from 2 different network configurations.

A simple example of this process is illustrated and compared with the normal blind attack in Figure 4.1. To achieve this, we propose a combination of data prepossessing via T-distributed stochastic neighbour embedding (T-SNE) for dimensionality reduction followed by density-based spatial clustering of application with noise (DBSCAN) to classify the data sets. We outline the justification for our chosen methods below.

4.2.1 Attack Design Considerations

Power transmission systems are by their very nature large. To design such data-driven attacks, one of the key considerations is to maintain the feasibility of implementation in the real-time operation of large-scale systems. Therefore, it is essential to circumvent the curse of dimensionality (CoD) within the context of this attack. We hence explore the use of T-SNE to reduce the dimensionality of data sets before applying the clustering algorithm. In addition, due to the blind nature of the attack, no prior knowledge of the number of underlying topologies can be assumed, and therefore ,an unsupervised learning method, DBSCAN in this case, is developed.

T-SNE for Dimensionality Reduction

T-SNE is a form of dimensionality reduction that works by constructing probability distributions over pairs of objects containing high dimensionality [75]. T-SNE considers a set of N high-dimension objects. d_i and d_j are two points within this set. σ_i is the variance of the Gaussian function centred on data point d_i. The closeness of these data points is defined by the conditional probability $p_{j|i}$ that point d_j would select d_i as a neighbour given that the neighbours are picked proportionately to a Gaussian function centred around d_j. This is given by

$$p_{j|i} = \frac{\exp(-||d_i - d_j||^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-||d_i - d_j||^2/2\sigma_i^2)}. \tag{4.12}$$

The aim of T-SNE is to reduce these points into their low-dimensional counterparts g_j and g_i. These have an equivalent conditional probability $q_{j|i}$ defined by

$$q_{j|i} = \frac{\exp\left(-||g_i - g_j||^2\right)}{\sum_{k\neq i}\exp((-||g_i - g_j||^2))}. \tag{4.13}$$

If the map points g_j and g_i correctly model the similarity between the high-dimensional sets, the conditional probabilities $p_{j|i}$ and $q_{j|i}$ will be equal. The positions of g_i and g_j are determined via gradient descent between the distributions p and q, and this is used to minimise the Kullback-Leiber (KL) divergence via cost function C [76] shown by

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}, \tag{4.14}$$

where P_i is the conditional probability distribution over all data points, given that data point d_i and Q_i is the conditional probability distribution over every other map point, given map point g_i.

Native T-SNE itself has a time complexity of $O(n^2)$, but this can be reduced to $O(n)$ by using optimization techniques as discussed in [77]. The brunt of the computational load is therefore taken by T-SNE, which reduces the measurements of the network power flows into 2-dimensional space.

There are other possible unsupervised approaches for dimensionality reduction. Linear reduction algorithms, such as principle component analysis (PCA), are one such example. PCA performs linear mapping to lower dimensional spaces, and unlike T-SNE, it is deterministic rather than probabilistic. PCA being a linear algorithm means that it does have some computational benefits. However, PCA cannot represent complex polynomial relationships in the same way that T-SNE can. Also, the KL divergence minimisation that T-SNE employs means that much of the local structure of data is preserved in T-SNE, whereas it is not preserved to the same degree in PCA. We also consider that with the stated purpose of identifying like groupings of points, T-SNE is also the most appropriate choice. The probabilistic neighbour assessment approach of T-SNE seeks to identify neighbours specifically, which makes the output trend towards close and distinct cluster groups emerge (as shown in Figure 4.3). This makes it easy to identify groups for the next section of the attack algorithm (clustering and model building) to operate over.

DBSCAN for Unsupervised Learning

When the dimension-reduced data set is received, a cluster algorithm will be applied to identify the underlining clusters of the data. Due to the blind nature of the attack, we propose using DBSCAN for the unsupervised clustering portion of this attack. The DBSCAN algorithm works as follows:

1. An initial starting point is randomly selected. This point is then marked as visited.
2. The points adjacent to this point, defined by ε, are counted and added to a set.
3. If the number of points exceeds the defined min point value, the initial point is defined as a new cluster. This process is continued for all points in the neighbourhood.
4. If the number of points is less than the min, the point is defined as noise.
5. These steps are repeated until the whole set has been clustered.

DBSCAN shows good benchmark performance against other forms of unsupervised learning [78] and also offers several relevant advantages to this form of attack. DBSCAN has a time complexity

of $O(n^2)$, but this can be reduced to $O(n \log n)$ with parameter optimisation [79], unlike hierarchical clustering, which has a time complexity of $O(n^3)$ and is highly computationally intensive for large systems by comparison. DBSCAN also does not require pre-specification of the number of clusters (making it more appropriate for a blind-style attack) and is robust against outlying data points and noise. Density-based local outlier factor (LOF) was also considered and has previously been seen for FDI attack detection in [80]. In many ways, LOF is similar to DBSCAN but is more specialised for anomaly detection as opposed to direct clustering.

4.2.2 Intelligent Blind FDI Attack

This subsection details the proposed intelligent blind FDI, as outlined in Algorithm 1. Once the attacker obtained an adequate amount of measurement data, the attacker can initiate the attack algorithm. When the latest measurement data arrive, initially, T-SNE is applied for the dimensionality reduction of the sets of power flow observations into a two-dimensional space. The reduced form of the data set is then clustered via the DBSCAN algorithm into distinct subgroups of like measurements, and the one corresponding to the current system topology is identified. The mixing matrix is subsequently derived based on this subgroup of data by using independent component analysis as per the normal blind attack. A vector of false data $\mathbf{z}_a$ containing the desired attack bias will then be calculated based on the mixing matrix.

Ultimately, the attacker does not know which model corresponds to the base case or the case with MTD implemented. The attacker simply knows that there are multiple distinct underlying models and creates a series of models equal to the number of clusters. The attacker may be able to guess based on how the topologies represent in terms of timing, which is the base (no MTD case), but this is largely irrelevant for the attack. A minimum cluster size check will also be implemented to ensure that the attack has sufficient data to create the blind model.

The proposed process on applying this clustering technique is outlined in figure 4.2.

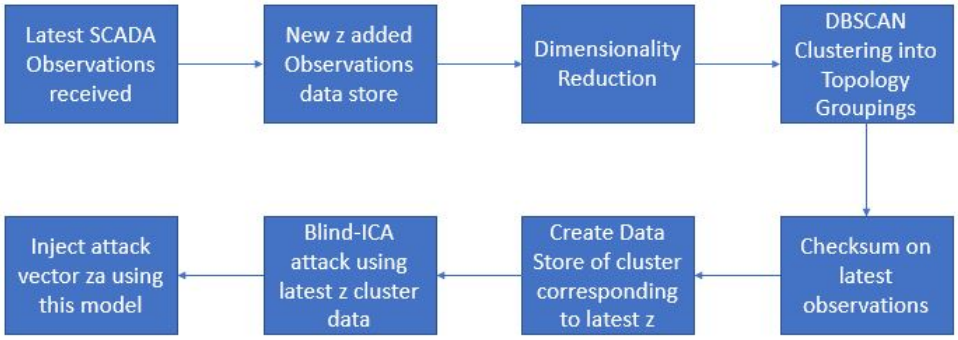

Figure 4.2: Proposed algorithm process flow chart.

Algorithm 1 DBSCAN Blind-ICA attack

Input: A set of power flow observations $\mathbf{z}_{obvs}$

Y = tsne (**z**˙obvs) %Dimensionality reduction

idx = dbscan(Y,mpts,ε) % Cluster power flows

For i = 1:length(unique(idx)) % Assign pf to cell

j = [j,idx] % Assign cluster to obvs

A{i} = j(j(:,l)==i,:)% Assign obvs to cell

c = j(end) %check what profile current **z** is

z = A(c) % Select only corresponding **z** measurements

HA = FastICA(**z**) % Run fastica for **HA**

$\mathbf{z}_a = \mathbf{z} + \mathbf{HA}\delta\mathbf{y}$ % Apply attack vector

Output: false data $\mathbf{z}_a$

MTD Cluster Distinctness

In order for MTD to be effective the system change needs to be sufficient so as to be captured in the CSE check. Therefore, in the existing MTD implementation, the change itself cannot be completely trivial and must be large enough to be visible. However, changes from transmission switching are very large and easily identifiable with any data driven methods. Previous works with D-FACTS devices for MTD apply between 10-20% changes to line branch values in order to evaluate FDI attacks. For example, considering the typical 14-bus system impact at the power flow level at 10% inductance change across branch 1-5 will cause a 6.1% change in the absolute power flow at this level (not trivial and likely identifiable above system noise). Additionally, on top of the absolute elements, T-SNE will capture elements of the relative values with the set (which will be highly dictated by the system topology). Which also makes distinct classification more likely. For example, in the MTD along branch 1-5 case we see the relative share of the power born by branch 1-5 fall from 0.48 to 0.44 with the implementation of MTD. T-SNE attempts to optimize the similarities in the high and

low dimensional space. The similarity calculation in T-SNE will reflect both the individual absolute values and relative factors values.

However, this clustering based intelligent attack relies on being able to distinguish between different topology sets from power flow measurements. If the changes in MTD are minor compared to the variation in power flow profile, it may become difficult to distinguish clusters, which will be explored in the next session to defence against such attack.

4.2.3 Performance Analysis

To demonstrate the performance of the proposed algorithm, a case study is carried out on a system with 14 lines equipped with D-FACTS for MTD. As shown in Figure 4.3, the proposed algorithm successfully identifies and clusters the potential topology sets, and only minor changes on topology (1% of base admittance) are applied. The computational performance of T-SNE and DBSCAN for different IEEE standard systems is shown in Figure 4.4 and compared with hierarchical clustering with embedded cluster evaluation. The case studies are performed for 1000 sets of observations. We note that for small scale systems (such as the 5-bus case), the computational performances are similar, but as the system becomes larger, the time for completion grows quickly for hierarchical clustering.

Real-Time Attacks in Large Systems

For real-time operation, the bottle neck for attacking with this technique comes in the identification and classification of the last meter measurement set within the wider pool, i.e. the ability to identify which model the attack should be based upon. The proposed technique can also be practical for large systems that may contain a high number of measurements. In Figure 4.5, we show the time to completion for the T-SNE and DBSCAN portions of the algorithm in the presence of very large random arrays (up to 10k points). It demonstrates the (expected) linear relationship from the T-SNE time complexity. Even when considering large data arrays with a large number of observations, the time to perform the T-SNE/DBSCAN flow is relatively short. For example, using 10,000 observations for a 10,000-point system, it took around 8.06 minutes using only an Intel Core i7-7820X CPU with

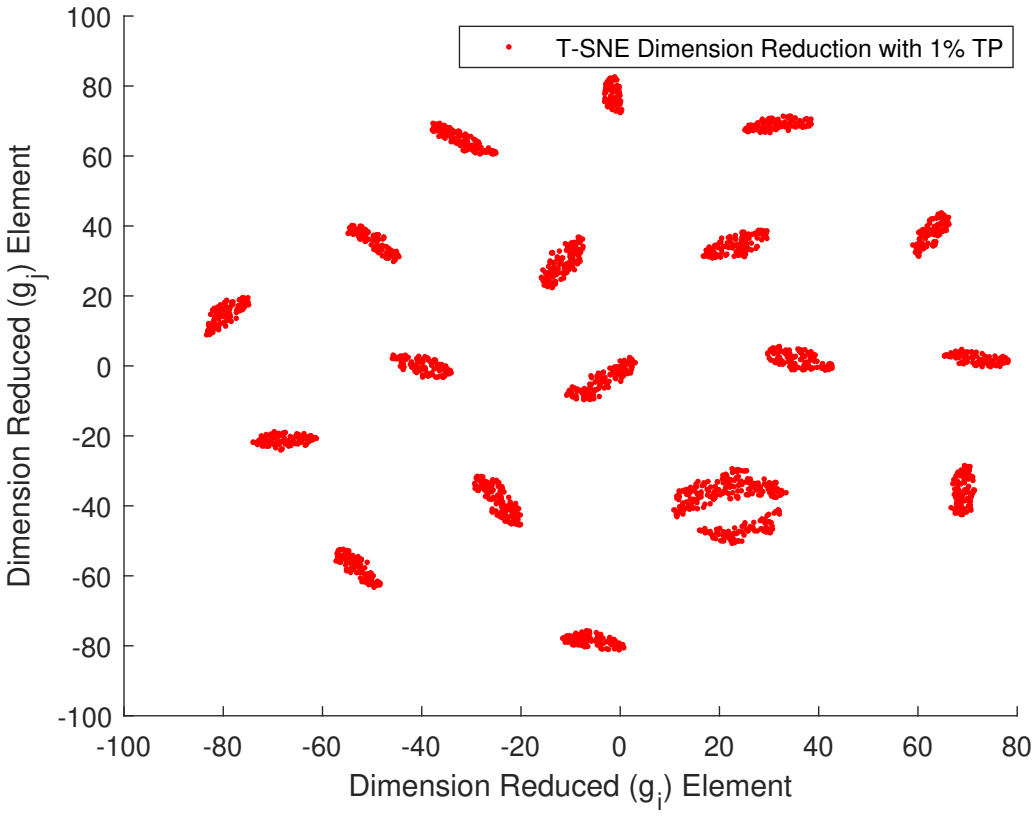

Figure 4.3: Power flow profile observations of 1% admittance perturbation MTD applied to 19 lines intermittently under T-SNE dimensionality reduction. The X and Y axis values are non-dimensional probabilistic reductions (g_i & g_j from equation 4.13). The system is reduced from a 34 dimension meter IEEE 14-bus system.

64 GB of RAM. We would expect that a highly motivated supranational attacker would have access to much more sophisticated hardware and would be able to execute such attacks even quicker.

4.2.4 Load Profile Bucketing

As seen in [48], large load variations can make distinguishing changes from MTD in the system hard. Fundamentally, load variation can be used itself to hide MTD. Therefore, we consider an extension to reduce a highly variate load system back to a steady load-style assumption under which the attacker can have more success. The attacker will use a combination of T-SNE and discrete bucketing to group load sets by their full system profile. The load variation within these individual buckets will be small and equivalent to a steady load-style assumption. The attacker can then run the attack over one of these buckets, and it work as if the steady load assumption were in place.

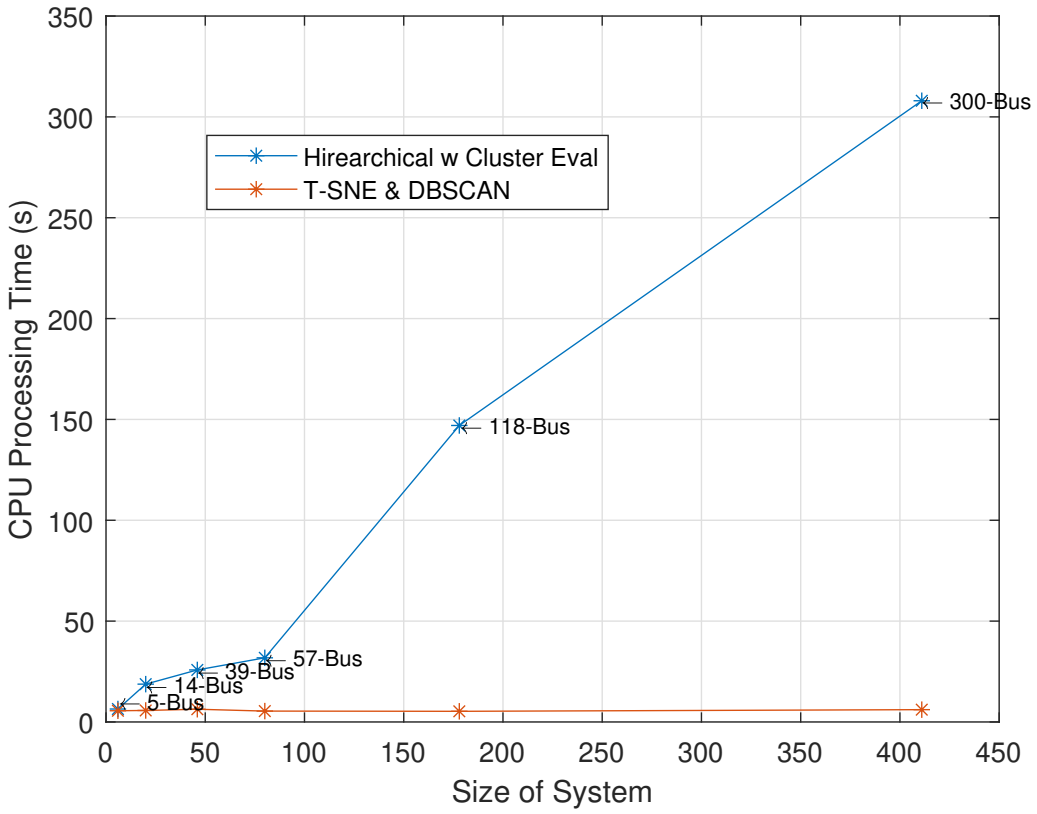

Figure 4.4: CPU processing time for the combined T-SNE/DBSCAN algorithm with an equivalent hierarchical method with embedded cluster selection performed on systems of increasing size. Performed for 1000 observations.

4.3 Physical Gaussian Watermarking with CUSUM

Although physical watermarking has not been applied in the power system space, the concept has been proposed in control systems such as in [73], where a watermark is added into LQG-based control signals to drive detection. However, the papers in these areas are not true 'physical' watermarks because they only change signal parameter dependencies and not the underlying physical plant itself. At the same time, it should be noted that although MTD in the form of D-FACTS control to change system topology has been explored, the use of watermarks in combination with MTD has not been investigated, and there is an opportunity to incorporate a true physical watermark into the system plant to enhance the system security.

Previously, topology perturbation and transmission switching have been proposed as methods to drive the detection of FDI attacks [37][48]. These methods implement significant changes to line admit-

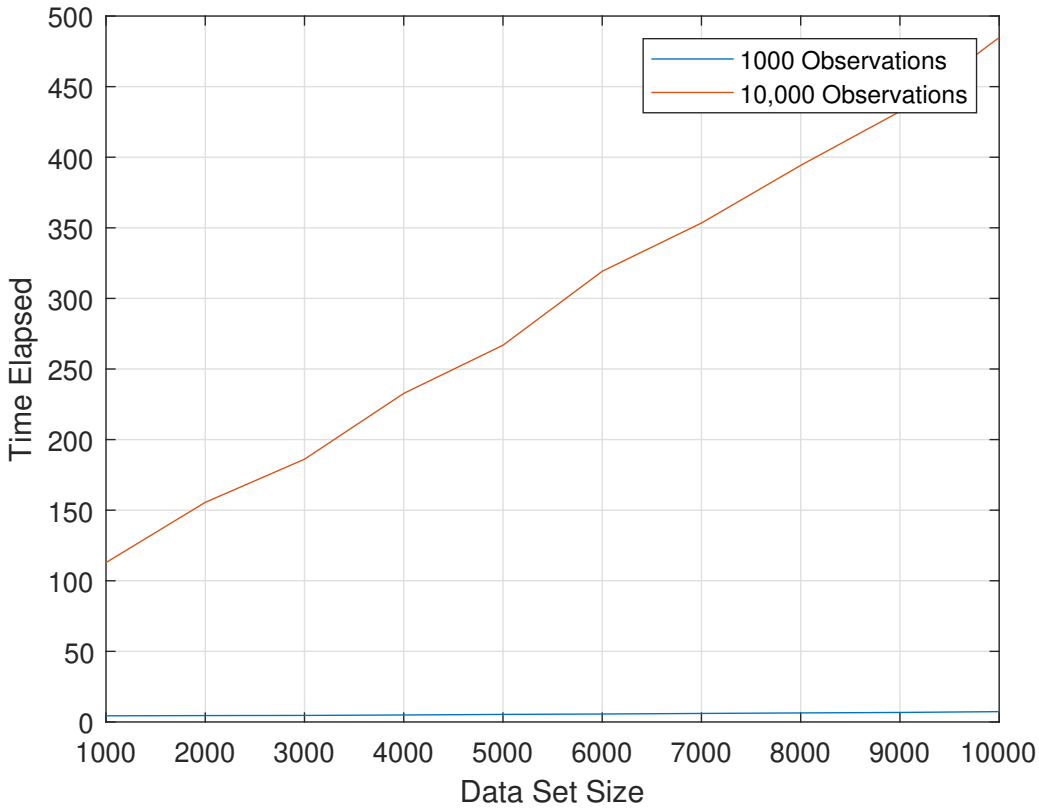

Figure 4.5: CPU processing time for the combined T-SNE and DBSCAN algorithm for increasing size of random data array upto 10,000 data points. Performed for 1000 and 10,0000 observations.

tance as required by the change needed in residual (typically around 10-20% for D-FACTS-based changes), which may not only lead to the interruption of system operation but also provide opportunities for the data-driven attack to spot the existence of MTD and counter it. It is hence crucial that the deployment of MTD can remain hidden to the attacker.

In this work, it is assumed that the SO will incorporate the capability of D-FACTS devices into the OPF model to optimize and select the lowest cost scenario, as shown in [81]. MTD will then be applied around this point. As outlined in [82], there is a non-trivial cost incurred when applying conventional MTD. This cost comes in the form of the non-optimal usage of power system assets. Whereas before D-FACTS are applied to minimise losses from reactive power, they are now being used for MTD purposes away from this optimal point. As a result, the defender will wish to reduce the overall application of MTD.

In this context, this section proposes a novel method to achieve this by combining MTD with physical

watermarking, which makes the MTD itself indistinguishable from the noise profile of the system, and monitoring sequential errors for long-run trends by using cumulative sum (CUSUM) monitoring. CUSUM monitoring is a sequential analysis technique that monitors for change detection over a number of measurements. Samples taken from the process are assigned a weighting and summed to monitor change detection. In this case, we will monitor the measured residual r under MTD defined by

$$CEM_t = \sum_{j=1}^{t} r_j - T. \tag{4.15}$$

where CEM is the decision statistic, T is the target value of residual dictated by monitoring the statistic under normal conditions and t is the number of periods in a measurement set, with upper and lower control limits CEM_t^+ and CEM_t^-. Because r is an absolute value, the lower bound CEM_t^- will be 0. CEM_t^+ can be selected based on engineering judgement from prior observations. Usually, the upper bound can be defined in terms of the residual variance and mean value under no attack:

$$CEM_t^+ = \bar{r} + B\sigma_r. \tag{4.16}$$

where B is defined by the user based on previous observations and minimising type II error.

The proposed defence strategy introduces these minor errors by using D-FACTS devices to alter the line admittance by a vector $\mathbf{w}$. The size of admittance changes applied to each line is based on the output from a pseudo-random number generator (PRNG), the seed value of which is only known by the network operator. This can be achieved with existing technology via a unified power flow controller (UPFC) [83] in combination with a processing unit. The watermark may be applied selectively such that

$$w_m \in \{0, \mathcal{N}(0, p)\}. \tag{4.17}$$

where p is the max change applied to the branch admittance.

The resulting power flow profile under physical watermarking will be equal to

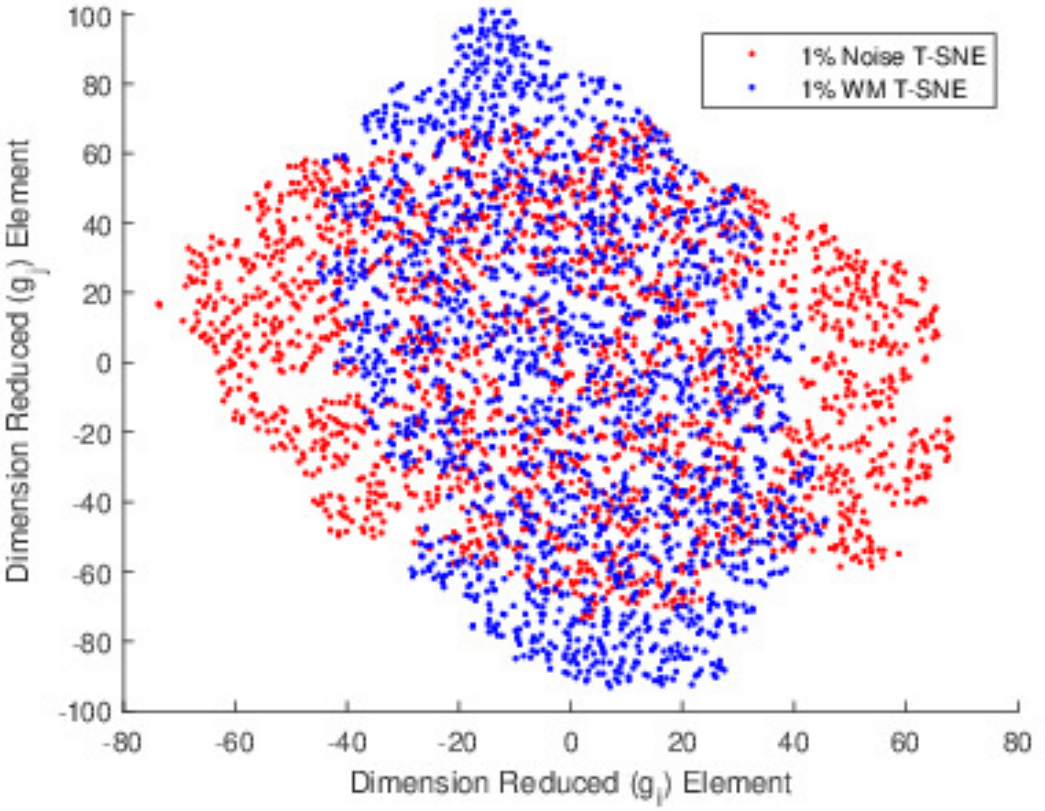

Figure 4.6: Power flow profile observations of 1% Gaussian watermark MTD applied to 14 lines intermittently under T-SNE dimensionality reduction. The X and Y axis values are non-dimensional probabilistic reductions (g_i & g_j from equation 4.13). The system is reduced from a 34 dimension meter IEEE 14-bus system.

$$\mathbf{z}_w = (\mathbf{H} + \mathbf{w})\mathbf{x} + \mathbf{e}. \tag{4.18}$$

where $\mathbf{w}$ represents the vector of admittance changes applied to branches and is known to the SO.

The impact of applying a Gaussian-style watermark in physical system parameters is shown in Figure 4.6. Compared with direct binary perturbation, the proposed MTD shows a similar profile as underline noise and makes it extremely hard for a clustering algorithm to identify the existence of MTD or to counter it.

The key advantages of the proposed defence mechanism can be summarised as below, which will be validated in the next session:

1. As the proposed MTD is on magnitude with the noise levels, the change in power flow observa-

tions resulting from the MTD becomes difficult to be identified. Therefore, MTD stays stealthy to the attacker.

2. Due to the stealthiness of the proposed MTD, it significantly increases the chance of the detection of an FDI attack and is specifically resilient to intelligent attack types such as the proposed DBSCAN blind-ICA attack.

3. The significantly-reduced magnitude of topology changes leads to less interruption on the system stability and economic operation.

4.4 Results and Analysis

This section assesses the performance of the proposed intelligent blind FDI attack in the presence of different forms of MTD on the standard IEEE 14-bus and IEEE 118-bus test systems [84]. All simulations were implemented using the MATPOWER toolbox in MATLAB [72] and performed using Intel Core i7-7820X CPU with 64 GB of RAM running on a Windows 10 system. In the graph legends, TP refers to topology perturbation (MTD via D-FACTS perturbation), and RS refers to switching MTD via circuit breaker control.

4.4.1 Model Assumptions

The priority of this section is to capture the change in detection between the blind FDI technique and the proposed intelligent attack under different types of MTD. Some assumptions have been made across all simulations:

- Uncoloured Gaussian noise error of 1% noise-to-signal was added to meter values as error $\mathbf{e}$, as seen previously in [85].

- A steady load assumption is made with a load variation of around 0.1% for initial simulations, as seen in [48]. Additional case studies were performed with multiple load profiles.

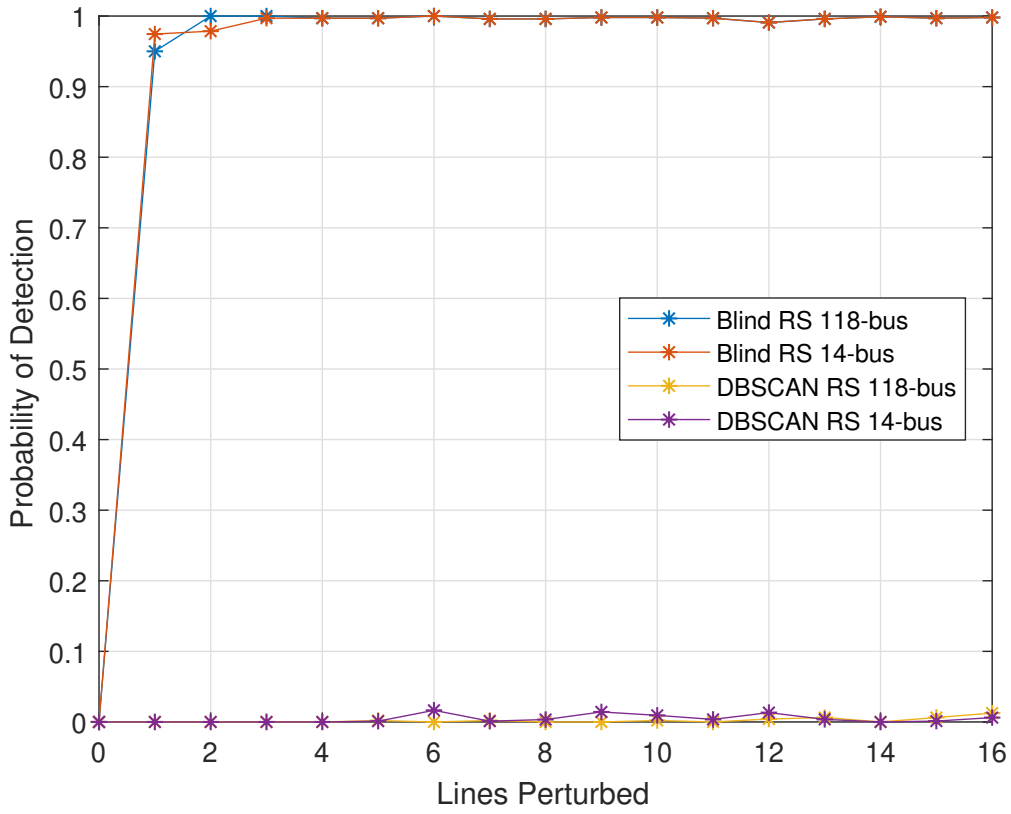

Figure 4.7: Probability of detection of blind FDI attack and the new attack under transmission switching for IEEE 14-bus and 118-bus systems under 99% confidence interval. Lines are not perturbed simultaneously.

- A minimum number of observations of 250 is assumed initially, which rises to 1000 sequentially over the course of the simulation.

4.4.2 Line Applications of MTD

In this paper, MTD is applied at the branch level in a fixed order shown in table 4.1 with inductance values, based on a % of the branch inductance. This order is consistent between MTD types to ensure a fair comparison between MTD performance. The number of lines perturbed (NLP) refers to the number of adjusted lines within a given scenario. Line adjustments are not applied simultaneously, and therefore, a simulation will have $NLP + 1$ potential underlying topologies that a successful attack will need to model. The NLP list is additive and the topologies within them randomly selected from within this list.

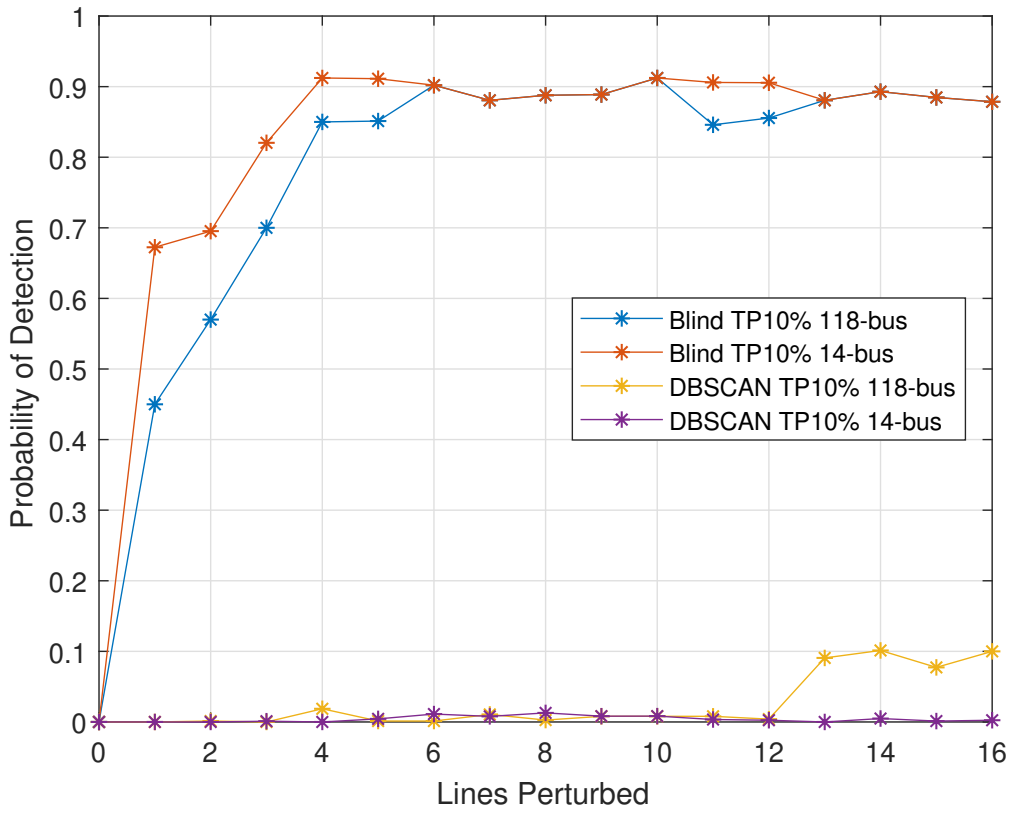

Figure 4.8: Probability of detection of blind FDI attack and the new attack under admittance perturbation for IEEE 14-bus and 118-bus systems under 99% confidence interval. Lines are not perturbed simultaneously.

4.4.3 Transmission Switching

The first form of MTD we trial is the direct use of system circuit breakers to create new topologies (transmission switching). In this case, lines are switched into and out of operation to change the underlying topology incidence matrix. This creates significant changes in the overall power measurement matrix. Figure 4.7 shows the impact of transmission line switching on the blind FDI attack and DBSCAN attack for the 14-bus and 118-bus cases. For the standard blind FDI attack, transmission switching is highly effective at introducing residual errors and driving alarms. With a single line, switching the detection is 100% for the standard blind FDI attack. However, these large changes in the system flows make it easy for an attacker to identify the MTD. Compared with the standard attack, the DBSCAN attack outperforms the standard blind FDI whenever MTD is used. Detection remained low (less than 1%) with up to 15 lines being switched in/out across the network at different times. Even with 16 possible topologies in use, the detection remained under 3%. Transmission switching is

Bus 1	Bus 2	R	X	NLP
1	2	0.01938	0.05917	1
1	5	0.05403	0.22304	2
2	3	0.04699	0.19797	3
2	4	0.05811	0.17632	4
2	5	0.05695	0.17388	5
3	4	0.06701	0.17103	6
4	5	0.01335	0.04211	7
4	7	0	0.20912	8
4	9	0	0.55618	9
5	6	0	0.25202	10
6	11	0.09498	0.1989	11
6	12	0.12291	0.25581	12
6	13	0.06615	0.13027	13
9	10	0.03181	0.0845	14
9	14	0.12711	0.27038	15
10	11	0.08205	0.19207	16

Table 4.1: Order of Number of lines perturbed (NLP) applied.

unlikely to be used for the sole purpose of attack detection due to the significant impact on the system operability.

4.4.4 Admittance Perturbation

Admittance perturbation is the most commonly proposed method of MTD for power systems in the current literature. This subsection implements an admittance perturbation defence against the typical blind FDI attack and the proposed DBSCAN version. A quantity equal to 10% branch admittance is injected into the lines given by the order in 4.1. As discussed, branch admittance are applied independently, and the number of underlying topologies will be equal to $NLP + 1$. When the inductance is injected, the system operator is expecting to see the change in admittance reflected in the resulting power flows. If the attacker is unaware and does not reflect the new admittance in their attacking vector, the residual will increase significantly, and BDD will be triggered. The results of admittance perturbation on detection of the standard and DBSCAN blind FDI attack are shown in Figure 4.8. System models with branch admittance perturbations of 10% were implemented. The standard blind FDI attack performs poorly against this form of MTD. For a single line at 10% perturbation, a detection level of over 95% is achieved. The detection rates for the DBSCAN-informed attack were

consistently low. This is due to the distinctive clusters of power flows emerging under the steady load assumption. There is a small spike that appears around 12 lines perturbed. This is likely due to the increasing number of lines perturbed in the system, which likely causes a mis-clustering in the underlying data set or depriving a cluster of enough data points for a decent model.

4.4.5 Physical Gaussian Watermarking with Cumulative Errors

In this session, we provide results for the novel implementation of the Gaussian watermark with combined CUSUM monitoring. In the same order and manner as in the transmission switching and admittance perturbation sections, we apply a Gaussian-style physical watermark as the defence. An inductance change of 1% to the system varied over a random distribution. Only one line change is applied at a time to keep it consistent with the other forms of MTD. The admittance profile is varied using a PRNG with a profile equivalent to the underlying noise of the system. Because we have used a 1% noise for our simulations, the p value is set as equal to 1% to ensure that this profile is not visible to the attacker. This is combined with CUSUM error monitoring, watching for sustained increased errors over 10 measurements with a cumulative limit based on 2 standard deviations above the average CUSUM measurement error summation under normal conditions.

Figures 4.9 & 4.10 illustrate the implementation of the Gaussian watermark with the assumption that an FDI attack starts from time instance 30. Figure 4.9 shows the traditional CSE residual error resulting from an FDI attack in the presence of the Gaussian Watermark. It is clear that the small system changes cannot directly drive the detection of an FDI attack in conventional residual-based BDD. Monitoring for the average of the last 10 measurements allows the system operator to identify long-term trends in the data, which in this case are caused by small but sustained gross errors introduced from the FDI attack. In Figure 4.10, the CUSUM method of detection is applied based on the last 10 measurements. Initially, we do not implement any attack for the first 30 runs of the system, and we note residual CUSUM averages in line with the normal value. At run 30, we introduce the attack vector. From inspection, it is much clearer that the system is under attack, and an alarm is raised after 4 consecutive measurements.

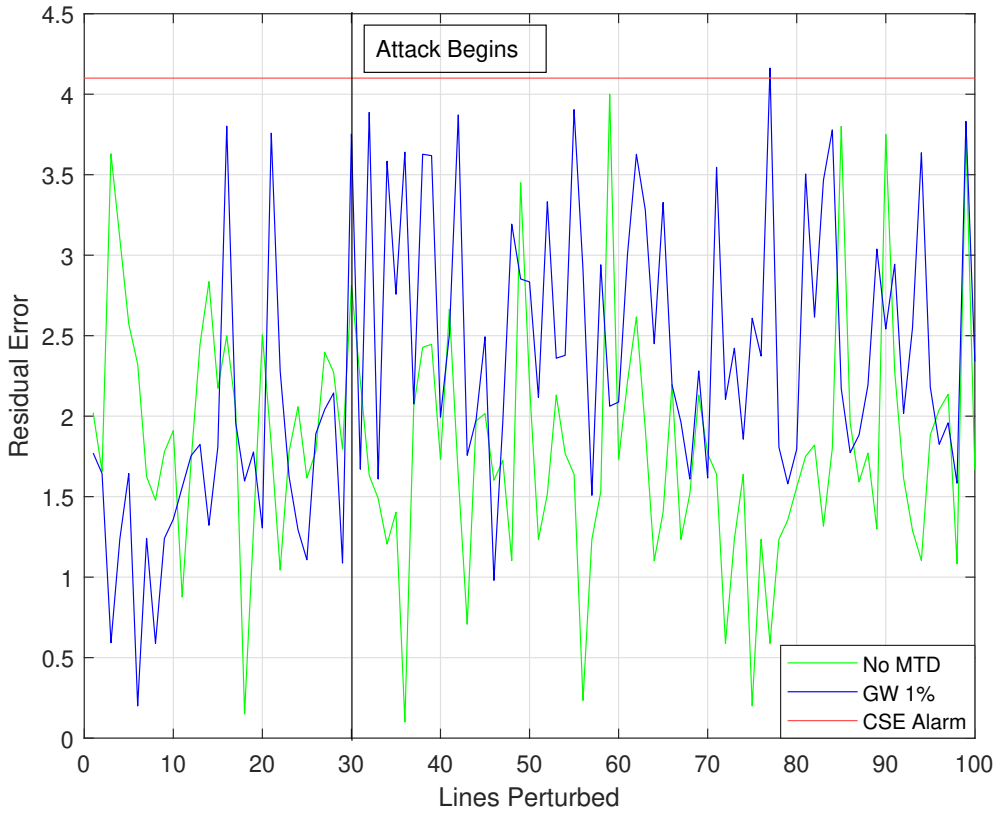

Figure 4.9: Conventional CSE Residual error for run numbers on 14-bus system with the Gaussian Watermark applied to 14 lines. A bus angle change of 20 degrees attempted across the system by the FDI attack.

As shown in Figure 4.11, under the DBSCAN blind FDI attack, the CUSUM Gaussian watermark shows significant improvements. As additional lines are added, these detection rates are close to 100%, compared with under 10% for standard admittance perturbation. The is due to the difficulty the DBSCAN algorithm has in identifying clusters for MTD of a magnitude identical to the noise profile of the system. Type-II error based on 2 standard deviation moves from the CUSUM average appears to give around 3% type-II error for this kind of measurement approach across 1000 measurements. As seen in 4.11, this cumulative approach also requires multiple measurements that potentially could lead to the attacker having additional time to attack before being caught. Therefore, there is a trade-off between the speed to spot attacks and the magnitude of the added watermark. Figure 4.12 illustrates this for the 118-bus and 14-bus systems, where for a lower level of added watermark, a larger number of measurement points are needed to break the threshold. This approach, of course, has other less quantitative benefits. The application of the Gaussian watermark makes the power profile even with

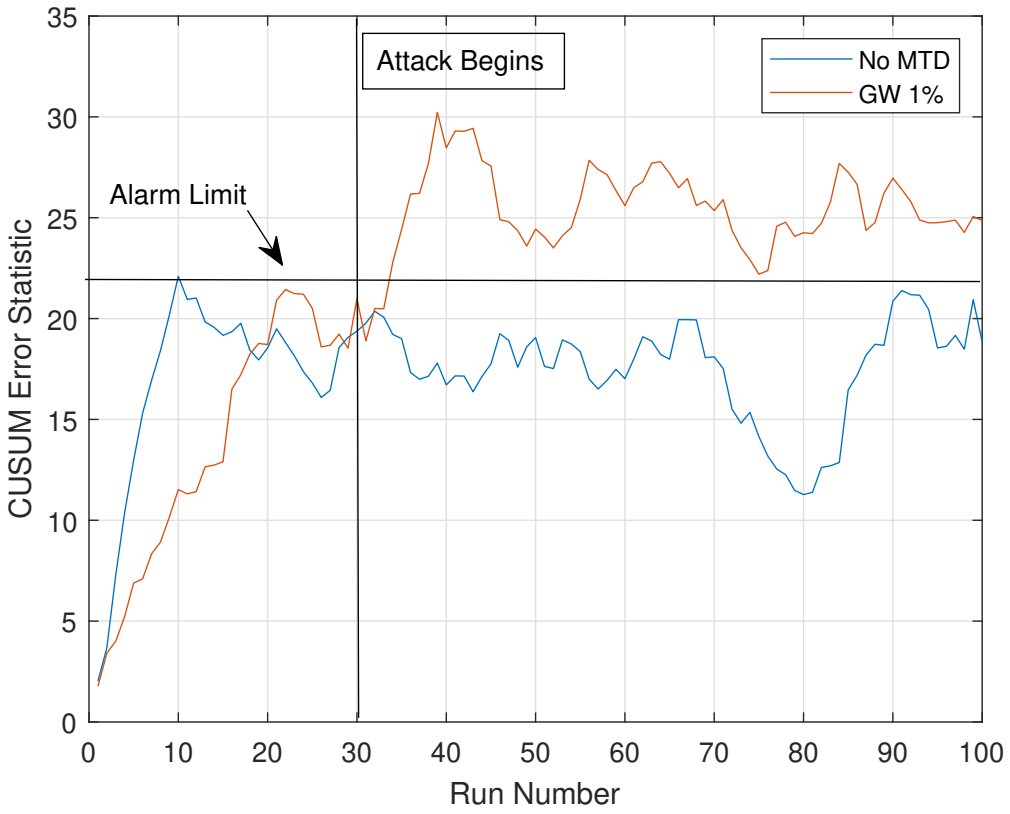

Figure 4.10: CUSUM rolling summations for run 14-bus system with the Gaussian Watermark applied to 14 lines. A bus angle change of 20 degrees attempted across the system by the FDI attacker.

MTD appear very close to random noise under the T-SNE dimensionality reduction. This is in contrast to the normal applications of MTD which can be see quite distinctly. In effect, even without the added benefit of difficulty of clustering it makes it unlikely that the attacker would even be aware that MTD is in operation. This can confer secondary benefits. For example, the presence MTD can often inform an attacker of where there might be high value target. In general, operators will not use MTD to protect low value targets, this is due to the implicit cost of infrastructure and operation. Therefore, by having an obscured profile we can prevent attackers from using MTD as a target selection criterion by which them inform themselves of which targets are high value based on the presence of MTD.

4.4.6 Load Variance Impact

In previous case studies, the simulations have been performed under steady load assumptions [48]. This session investigates the impact of large load variation on the performance of the DBSCAN

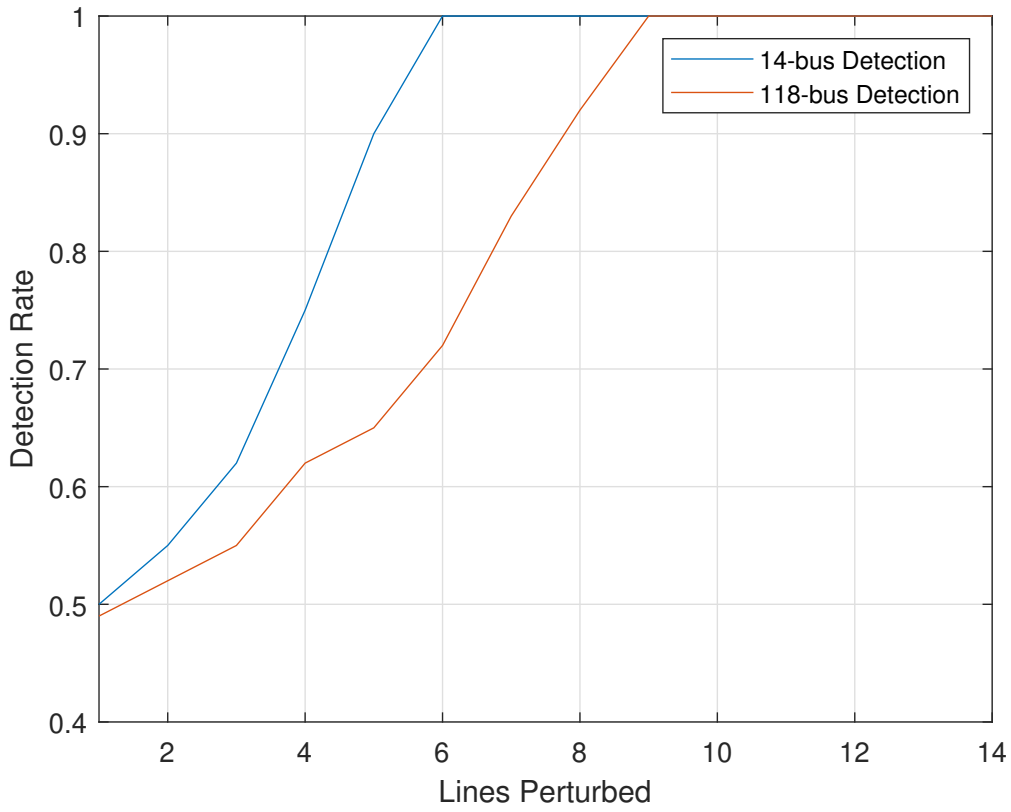

Figure 4.11: DBSCAN detection results under proposed Gaussian watermark with cumulative errors over 10 measurements. 14-bus and 118-bus systems simulated with baseline 10 measurement average as detection trigger.

attack. As shown in Figure 4.13, large load variations reduce the effectiveness of the DBSCAN under pressure from topology perturbation due to the increasing challenge to cluster the topology changes under a high varying load. To circumvent this challenge, we separate differing load values into buckets based on the system profiles reduced via T-SNE. Load values are observed directly, performing dimension reduction on them via T-SNE and assigning them to bins of similar values. In Figure 4.14, the profile of the loads themselves are observed, and the dimensions are reduced and bucketed. Within each load group, measurement observations can be used to obtain the clear MTD groups to develop the attack model. Load bucketing reduces the effective load variation back down to a steady load-style scenario. Figure 4.15 and Figure 4.16 show the results of high load variation on the system under MTD with and without load bucketing applied. It is clear that under load bucketing, the distinct groups of measurement observations becomes clearer as a result of MTD. Applying this bucketing reduces the effective variance of the power flow significantly from 10% to under 1%, with around 25 buckets for a 14-bus system. This effectively replicates the steady load assumption, even

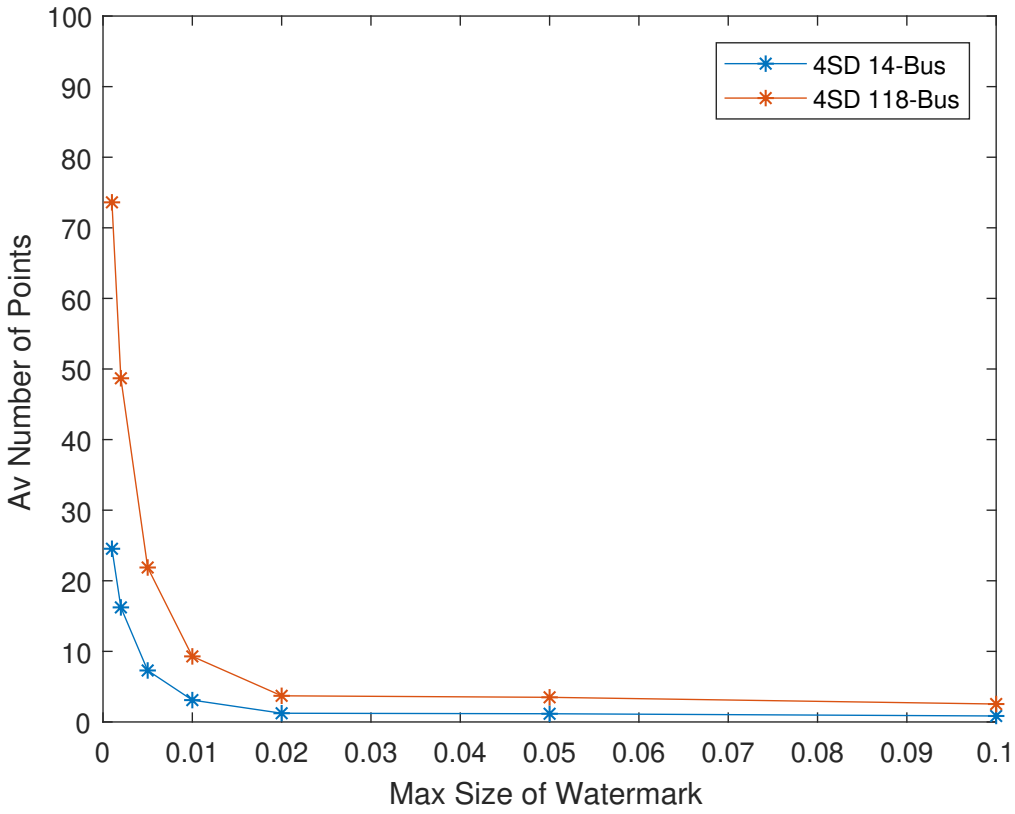

Figure 4.12: Average number of points required to break a 4 standard deviation upper limit for increasing size of watermark applied to a single line.

in the case of a more variate system. As the system variance becomes larger, additional buckets can be added to accommodate the larger variance of the system. The effect of this can be seen in Figure 4.13, with lowered detection for the DBSCAN attack when implemented against topology perturbation-style defence. Bucketing in this manner will require a large amount of data; however, based on the frequency of measurement at around 2-5 s [86] and the lengthy attack development phase [1], such a data requirement can be easily satisfied. Blind attacks themselves will always require a larger past data requirement than full knowledge attacks because they need to build a model, unlike full knowledge attacks in which they already possess the model.

4.4.7 Blind AC Replay Attack

We have implemented our clustering approach with a blind replay-style attack against an AC state estimation. Under this attack, the attacker attempts to inject a previously observed vector. The attacker

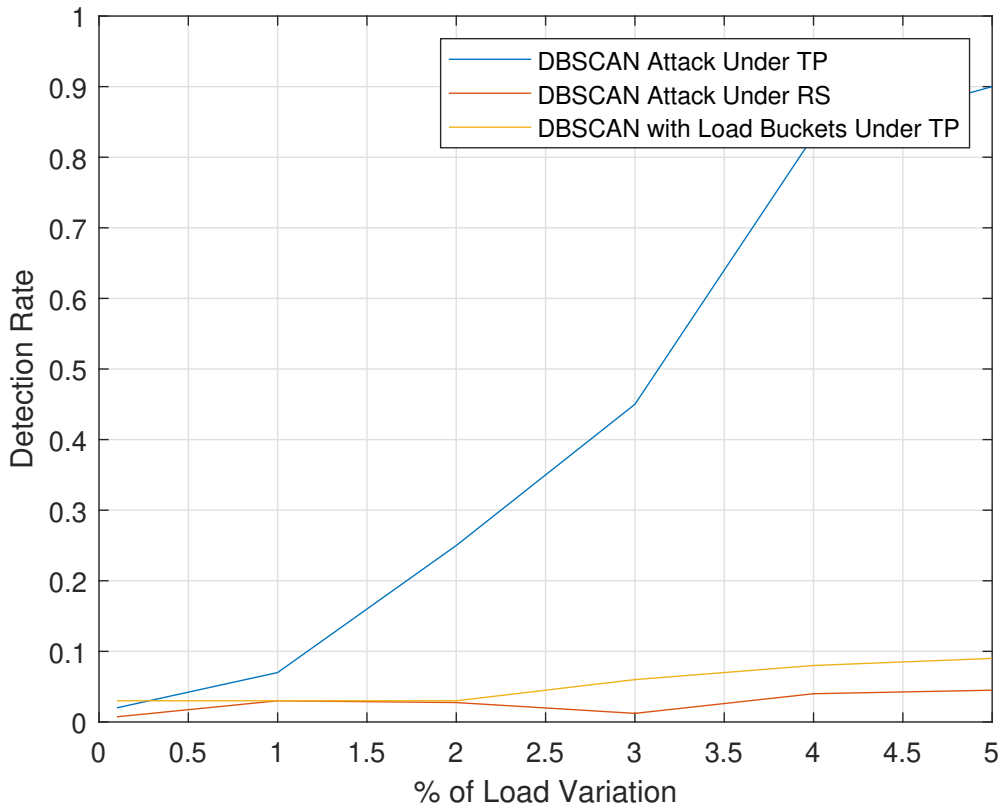

Figure 4.13: Detection of DBSCAN method with 10 lines perturbed with increasing load variance. Also featured is the DBSCAN with load profile reduction analysis with load variation effectively reduced using 10 load buckets.

is competing with MTD and wants to select the replay vector from a pool of values only containing those using the same topology configuration. In Figure 4.17, we can see that the distinctive cluster relationship exists within the AC model, as shown previously for DC. Figure 4.18 demonstrates that the proposed pre-clustering algorithm performs well in AC state estimation, provided a large number of samples have been received. The non-linearity in the AC model significantly reduces the correction rate of clustering, but increasing the number of observations allows good performance for the AC model. One of the advantages of opting for T-SNE as opposed to other dimensionality reduction methods (such as PCA) is that T-SNE is better capturing non-linear relationships. This will have a potential future-proofing effect for our intelligent, blind attacks as the clustering algorithm will be able to better capture and model for the non-linear impacts of AC power flow systems.

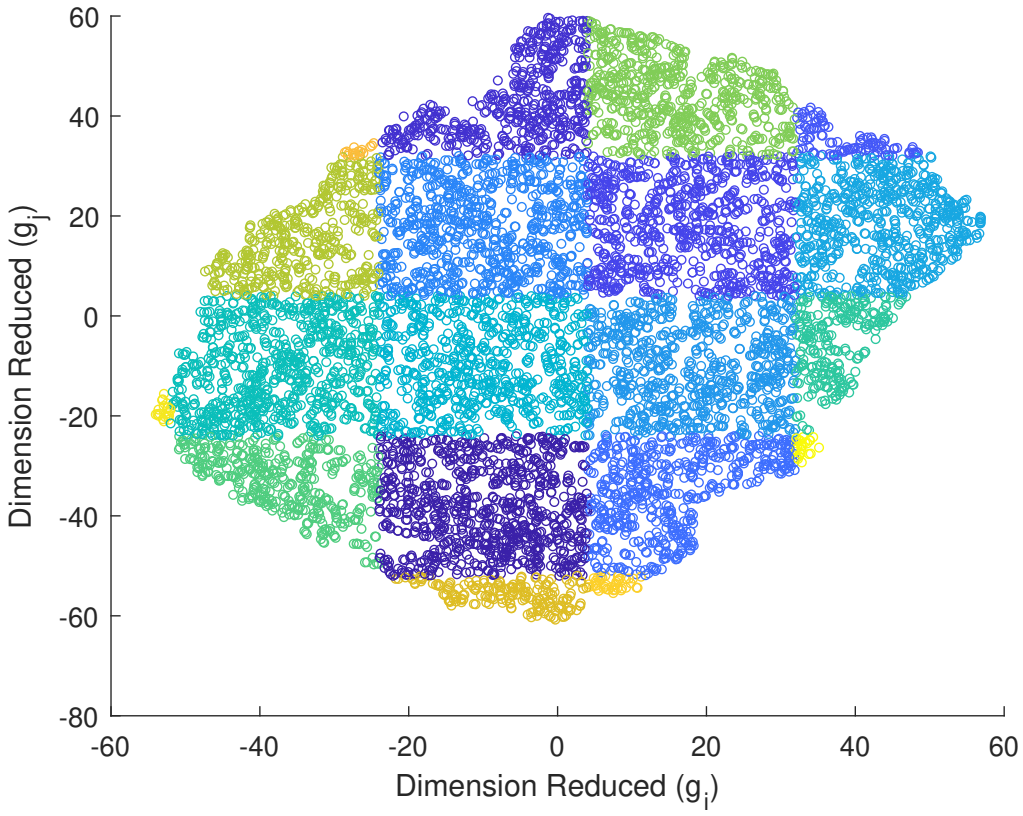

Figure 4.14: Real power load profile values reduced by T-SNE. Variation of 10% shown. Different colours represent different proposed load buckets. 10k measurements.

4.5 Summary

In this chapter, we investigate how MTD can be enhanced to withstand more intelligent FDI attacks which incorporate counter-MTD strategies. As we showed in the previous chapter, it is possible for an attacker to use data driven techniques to circumvent BDD and remain hidden. We also show that even physically driven forms of detection 'MTD' in some circumstances can be circumvented. To this end, we investigated how unsupervised learning and dimensionality reduction can be applied in blind FDI attacks to exploit the vulnerability of current forms of MTD. By incorporating a combination of T-SNE dimensionality reduction and the DBSCAN clustering algorithm, power flow observations can be clustered into their relative topology profiles. From here, the mixing matrix for the blind FDI attack can be calculated using only data under the same network topology. This technique is shown to be effective against admittance perturbation and transmission switching techniques. Indeed, this technique can be extended to a number of attack types to make them MTD resilient.

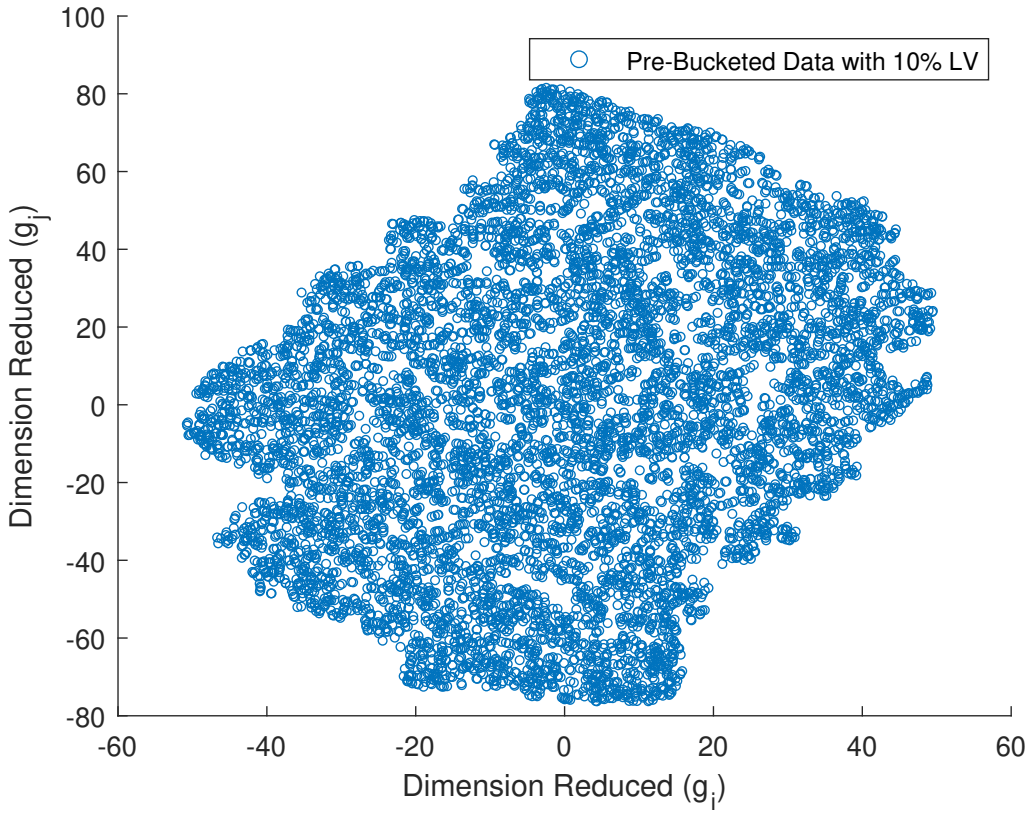

Figure 4.15: Power Observations for a 10 lines perturbed system using D-FACTS of 10% under load variance of 10% shown. This is prior to bucketing of data by load profile.

Given the lessons learned from the previous chapter, it becomes clear that system operators must not only consider that static systems can be attacked but also that systems with naive or simple applications of MTD are likely to be vulnerable to FDI. This is due to the proven attack proficiency of the intelligent counter-MTD FDI attack. At this point, we have now introduced new 2 models for attacking a power system: one that is effective against a static system and one that offers a MTD-resilient form of attack.

We therefore propose some innovations to defence. We consider that naive applications of MTD; namely those perturbed via predictable key spaces with large and predictable patterns are likely to be circumvented by an attacker (as they can be clearly seen). Therefore, an attacker must be able to hide the application of MTD within the noise profile of the system. We develop a novel defence strategy around this core idea to combat these new sophisticated attack types. We proposed that by combining MTD with physical watermarking to add an indistinguishable Gaussian-style physical watermark into the network topology and monitoring the sequential errors for long-run trends by using CUSUM

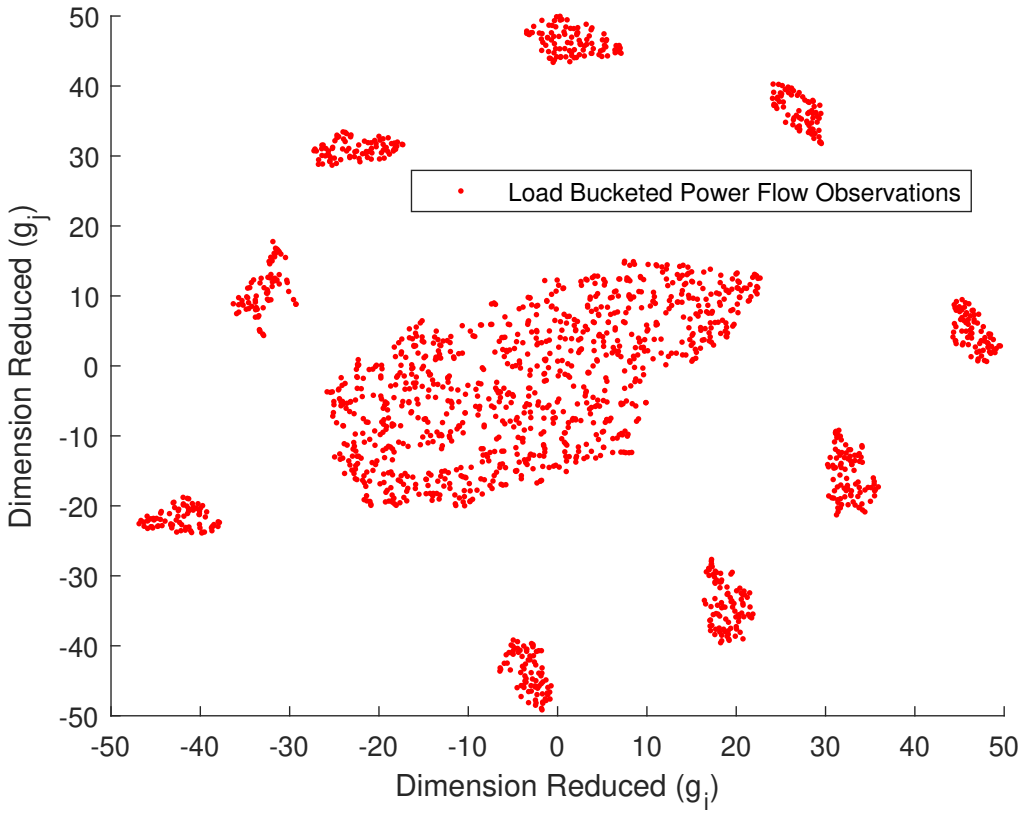

Figure 4.16: Post-bucketed power flow data with T-SNE applied for a 10 lines perturbed system using D-FACTS of 10%. Original load variance of 10% was used.

monitoring we could prevent these intelligent type of attacks from being stealthy. This technique is demonstrated to be effective at both inhibiting the attacker's ability to predict topological changes from visible power flows and reducing the overall impact on system operation by reducing the level of topology changes. The use of the camouflaged watermark gives less indications of the presence of MTD to the attacker. This is important, as attackers sometimes use these additional defences as markers for high quality targets. By obscuring the marker, the target looks just like any other point on the grid.

4.6 Lessons Learnt

Crucially, we draw the following lessons from this chapter that we use to inform our research and proceed to the next chapter in this work.

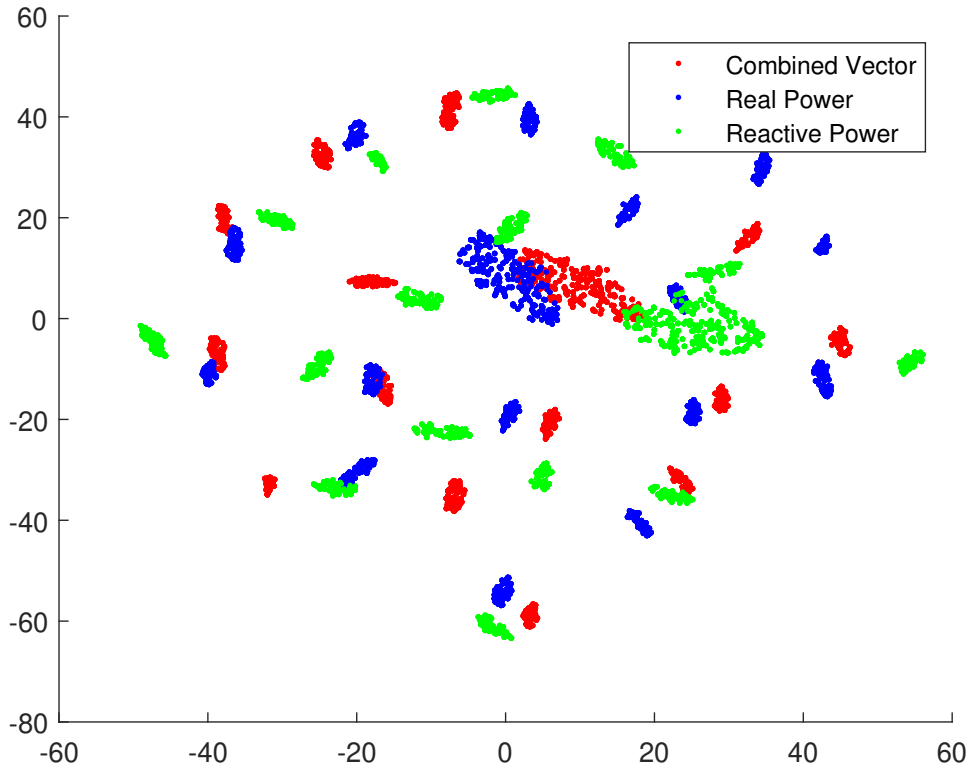

Figure 4.17: Observations of 1% MTD applied to AC system up to 16 lines intermittently. Data cuts of real power, reactive power and a combined vector incorporating both are compared. 1% Gaussian noise assumed.

- It is possible to circumvent naive applications of MTD, i.e. those that use a non-continuous waveform mode of direct perturbation. System operators should be considerate in how they apply MTD and should attempt to replicate natural wave forms (such as noise patterns) as a method of camouflaging MTD.

- We can easily provide additional support to system operator attack detection via the implementation of cumulative error-style monitoring. Small, sustained errors caused by small attack vectors can easily bypass BDD detection. Therefore, it makes sense to incorporate cumulative monitoring that will provide additional metrics for assessing the attacker's strength.

- Considering attacker psychology can also be important when structuring defences. In the past, naive applications of MTD such as "kicking the system" have been suggested. These kicks can often be seen very easily in the system profiles and an intelligent attacker can circumvent them. Camouflaging MTD can help prevent successful attacks as the attacker may not be even aware

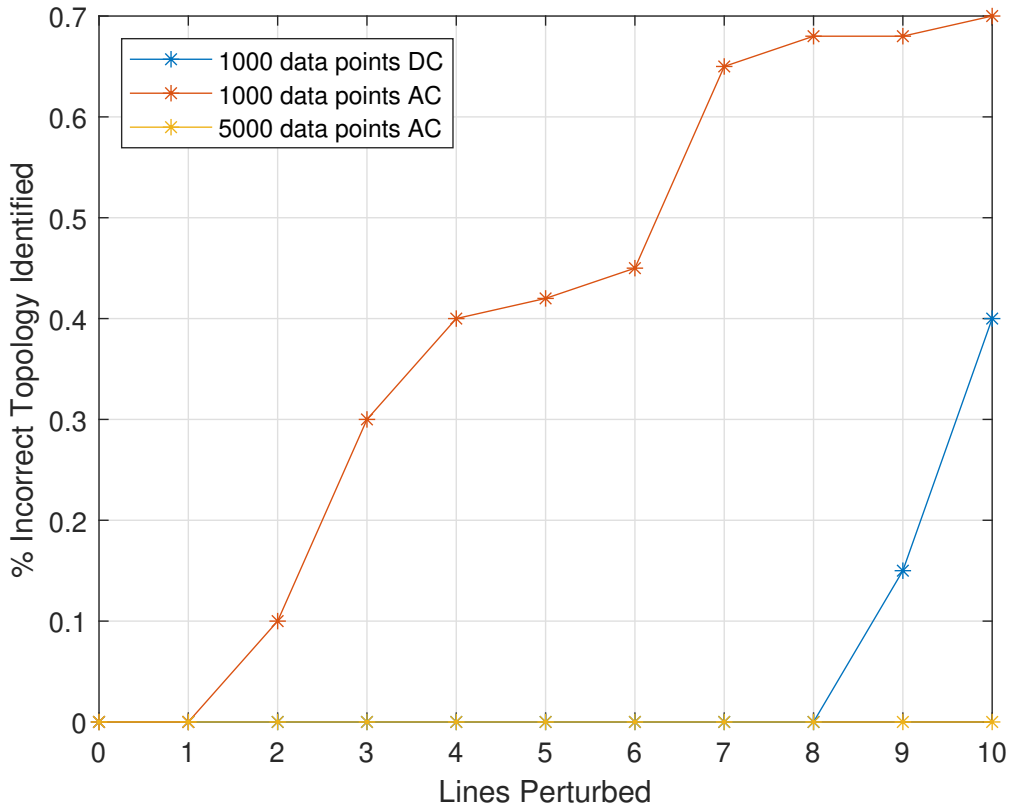

Figure 4.18: AC System probability of wrong cluster identified for in presence of D˙FACTs MTD with increasing lines perturbed to 14-bus system

MTD is in place.

In the next chapter, we start to consider how to minimise the overall application of MTD. We consider how event-based triggering of MTD can be used to limit the overall application of MTD. We also explore how distributed anomaly detection can be applied at a distributed level using Holt-Winters forecasting to decentralised BDD. Additionally we explore how MULTOS based-architecture can be used to provide additional enhanced and secure cyber-physical system security.

Chapter 5

Enhanced Cyber-Physical Security Using Attack-Resistant Cyber Nodes and Event-Triggered Moving Target Defence

This chapter outlines a cyber-physical authentication strategy to protect power system infrastructure against false data injection (FDI) attacks. We demonstrate that it is feasible to use small, low-cost, yet highly attack-resistant security chips as measurement nodes, enhanced with an event-triggered moving target defence (MTD), to offer effective cyber-physical security. A distributed event-triggered MTD protocol is implemented at the physical layer to complement cyber-side enhancement. The scheme is shown to be effective at preventing or detecting a wide range of attacks against power system measurement system.

The work outlined in our published journal paper [4] made up the basis of this chapter.

5.1 Introduction

In the past, the majority of works focused on either physical or cyber-security enhancements in isolation. However, more recently, works have been published that demonstrate the importance of con-

sidering overlapping cyber and physical solutions [87]. In this chapter, we propose and evaluate a practical and affordable system design that protects against attacks against both the physical and cyber layers and allows for distributed protection. The cyber domain of a power system is susceptible to various attacks. National guidelines suggest the use of cryptographic tools for authentication and encryption [88]. However, as outlined in [89], supervisory control and data acquisition (SCADA) networks (which are used for monitoring and controlling power systems), often lack secure modern encryption. This can be due to the age of assets (the encryption and security of networks being a more recent concern) or requirements for high speed in data processing with encryption adding too much time delay in measurement transmission [90]. The size and scale of power networks also makes the cost of refitting prohibitive with increased focus put on 'sweating' assets rather than incorporating new infrastructure. As a result, although a new cyber-layer protection solution is urgently required, it must be low cost (to ensure practical take up), secure by the Common Criteria security framework [91] and shown not to impede data transmission. Furthermore, the links between the cyber and physical systems are also vulnerable to attacks. As we have discussed, in the case of the power measurement system, a simple alternative to attacking the cyber-secured measurement node is to modify the analogue (or physical) measurement source, which cannot be protected by cyber-domain solutions. The attackers may conduct FDI attacks by manipulating analogue measurements before they even reach a cyber-secured point effectively bypass all encryption and authentication. We present our solution below, which addresses these security requirements in both the cyber and physical layers.

5.2 Novel Contributions

This chapter proposes and evaluates a practical and affordable system design that protects against both physical- and cyber-layer attacks and allows for distributed protection. We demonstrate how a cyber-physical authentication process can provide a foundation for system security at every level of measurement and transmission. This process is illustrated in Fig. 5.1.

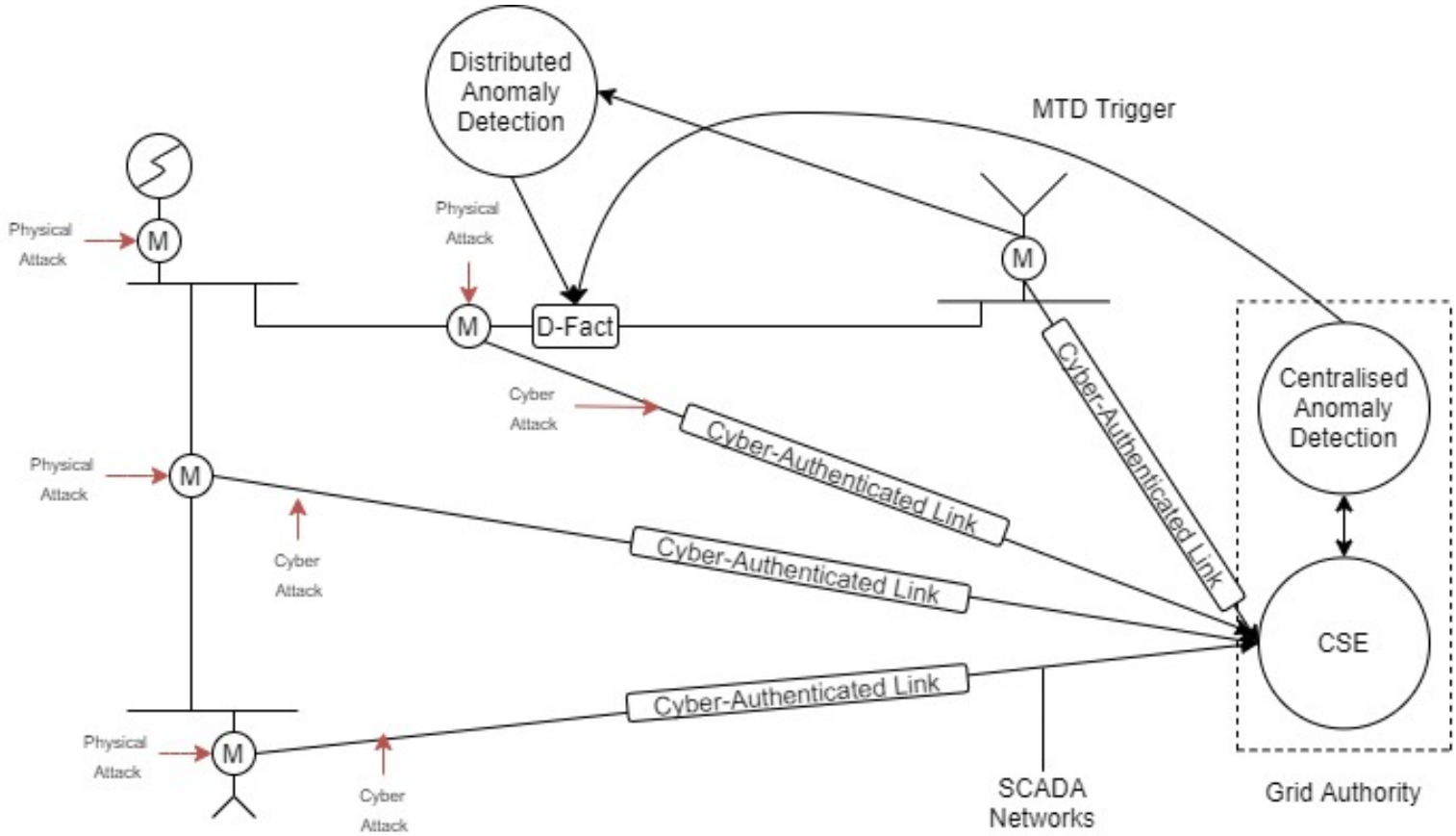

Figure 5.1: Outline of proposed cyber-physical defence model featuring the MTD trigger mechanism, metering and distributed anomaly detection.

Cyber Layer

In the cyber layer, we propose creating trustworthy measurement nodes by using a small encrypted payload protocol to secure the measurement reporting system while maintaining the required frequency of reporting. Each node is based on a highly tamper-resistant security chip that performs measurement capture, processing and reporting. The chip acts as a trustworthy platform, supporting the secure communication protocol with the grid authority (GA). The combination of the chip and protocol, offers effective protection against implementation attacks and communications attacks, such as man-in-the-middle FDI. The secured platform also allows confident delegation of some limited processing to the nodes, such as the proposed anomaly check and MTD triggering. The MULTOS Trust-Anchor for secure measurement transmission and MTD delegation has not been applied before. This may be because MULTOS platforms have been traditionally in the form of secured smart card and passport chips, which do not have general external I/O. Although AE has been standardised for some time, it has only relatively recently been studied on such security chips. The design secures the logical and implementation security of the system, but importantly, it provides a framework for the delegated/distributed implementation of trusted functions and self-defensive mechanisms. We have also shown that the secure nodes can sample, process and locally react much faster than would be feasible for a centralised system, offering the potential for new algorithms and research.

Physical Layer

In the physical layer, we propose a protection protocol based on event-triggered MTD. The protocol consists of an initial anomaly detection followed by MTD and traditional CSE. The anomaly detection uses Holt-Winters seasonal forecasting distributed to the individual measurement. The seasonality captures the intra-day demand differences to minimise the overall window for attack. If the anomaly detection is triggered, MTD is then implemented via inductance perturbation of D-FACTS devices. This changes the system model and drives detection of the FDI attack via increased residual errors. The event trigger comes prior to the application of MTD and is based on distributed measurements to target and minimise the overall use of MTD. This is important because while the FDI attacks are stealthy from a centralised residual detection perspective, their impacts can still be seen in the

distributed changes to power flows. By using D-FACTs devices in this manner, we are effectively promoting the use of inductance assets sub-optimally in order to evaluate FDI attacks.One of the key motivation drivers for this research is that MTD can be expensive to implement given the sub-optimal use of assets from an optimal power flow perspective, as shown in [82]. Therefore there is a requirement that we find ways of minimising the overall application of MTD only those times when there might be an attack present. Very few works have addressed the importance of distributed measurement with respect to FDI attacks, and none that we have seen combine a distributed and centralised approach as we have done here.

Our combined solution is a cyber-physical solution in that it combines secure cyber nodes with a physical mode of protection of the state estimation via MTD. This chapter proceeds as follows: first, Section 5.3 provides background to the proposed approaches. This includes outlining the state estimation environment for the power system. Section 5.4 proposes and describes a secure solution based on an authenticated encryption (AE) protocol and the use of distributed security chips (SCs). A novel event-triggered MTD is proposed in 5.5, and the Holt-Winters seasonal forecasting technique as triggering strategy is outlined. Experiments, performance results and the effectiveness of security measures are discussed in Section 5.6, with conclusions and suggestions for future work in Section 5.7.

5.3 Background

5.3.1 Cyber Vulnerability of the SCADA System in Power System

An attacker has several options when seeking to undermine the measurement and reporting system. The node itself can be attacked to either ensure that the node does not perform as intended or to reveal a credential (such as a cryptographic key) or proprietary functionality. It is also possible that the measurement source (e.g. the value to be sampled) could be modified so that incorrect values are captured but then securely delivered to the GA (upstream physical attack). However, the node's remote communications link presents the most attractive target, and attackers may seek to remove,

modify or replay valid transmissions, or generate fake transmissions that might be accepted by the node, and inject consequently false data into the state-estimator downstream of the physical system.

Node Implementation Attacks

Critical infrastructure may be attacked in a variety of ways. Adhering to information security best practice for the functional design of a measurement node (e.g. for algorithm, protocols, processes and keys) is not sufficient because attacks may target the implementation. Logical security attacks target weak design, typically in the algorithms, keys and poor software within the node platform, measurement application or loading and configuration functionality.

Direct attacks on hardware will generally require a higher level of expertise and equipment. They will also take longer to perform. However there are a few routes for entry via direct hardware attack. A chip could be decapsulated. It is also possible to reverse engineer the chip itself. Other alternatives include, probing buses and memories and modifying tracks. Fault attacks are less intrusive, and they disrupt normal operation of the chip but without damage. For example, faults can be generated by voltage glitches or radiation pulses. Under fault condition, an unprotected chip may reveal sensitive information, for example RSA keys [92]. Side-channel leakage is the leakage of sensitive information via an unintended channel. This can be key- or data-dependent timing, variations in power consumption or electromagnetic emissions. Simple analysis techniques (see [93] [94]) are powerful against naive implementations. In the case of a power measurement system, a simple alternative to attacking the measurement node is to modify the analogue measurement source (physical attack). A similar effect is achieved by manipulating any part of the path that digitises and communicates the source measurement value up until the point when it is within the node.

All these attacks can be successful against implementations that do not have specialist security protection. However, there is much industry expertise for the protection of secured micro-controller chips, which we exploit in our proposal, as discussed in Section 5.4.

Adversary Threat Model

We make the assumption of a well-resourced attacker, similar to those seen in [1], capable of launching attacks in both the cyber and physical realms. The attacker can attack either remotely via SCADA networks or directly at the physical sensor locations themselves. We also assume that the attacker has the capability to structure an FDI attack via the changing of meter measurements so that it is stealthy. Consistent with this is the assumption that the attacker has knowledge of the original system topology of the power system but is not aware of MTD configurations.

Node Communication Attacks

To undermine communications between two legitimate parties, an attacker may attempt to do the following:

1. Passively eavesdrop
2. Send a new message assuming a legitimate identity
3. Delay, replay or re-order legitimate messages
4. Modify a legitimate message
5. Block some legitimate messages
6. Send denial of Service (DoS) transmissions

The goal of eavesdropping is to capture unprotected data or information that can reveal credentials, such as keys or counters, which aid attacks against this protection. For grid measurement transmission, we are primarily concerned with the integrity of measurements rather than confidentiality. However, encrypting messages also protects associated control data and provides an extra barrier to attackers. Depending on the communications channel, an attacker might simply be able to source a fake message or may have to be positioned as a 'man in the middle'. In either case, the attack will be defeated by mutual authentication of communicating parties because the attacker will not have

the cryptographic credentials to authenticate, fake or modify a valid message. Most communication systems experience some variation in transmission delay, but excessive delay (or replay) can be detected by timestamps. Message re-ordering can also be detected by chaining encrypted transmission sequences. If a system occasionally loses messages in normal operation, then an attacker could block some without being detected. However, limited data loss could be overcome by redundant transmissions. DoS would prevent legitimate messages from getting through; however, it is a condition detectable by the legitimate communicating parties, which should be able to take some alarm and/or remedial action.

5.4 MULTOS Trust-Anchor

Our proposed solution requires trustworthy measurement nodes that are *strongly* attack resistant. Fortunately, there is a well-established process of assessment to determine this known as Common Criteria (CC) [91]. The CC evaluation is most commonly applied to security chips such as those used in bank cards, passports or hardware security modules (HSMs). High levels of evaluation confirm strong resistance to all known attacks, including logical, physical, fault and side-channel. The resistance is based on specialist hardware, supported by software defensive measures. A CC-evaluated SC should not be vulnerable to logic flow attacks, being compliant with security best practices for design and loading. For defending against physical attack, the CC chips will have numerous effective defences, including passive/active shields (to impede probing), hardware-encrypted buses and memories (to prevent data discovery), light and anomaly sensors (to detect decapsulation and faults) and scrambled circuitry (to impede reverse engineering). For fault-attack resistance, the hardware sensors detect fault insertion and prevent signals being released to the attacker from which he can gain system information. Software countermeasures provide secondary defences, e.g. verifying an algorithm result before providing an output and redundantly testing flags and loop counts. Side-channel leakage is well defended against, with countermeasures that impede the statistical averaging of signals or reduce the source generation of the leakage.

MULTOS [95] chip platforms are quite common in CC evaluations, and the MULTOS [96] Trust-

Anchor was selected for our proposal. MULTOS is a high-security multi-application smart card Operating System (OS) that is managed by the MULTOS Consortium. MULTOS chips are found in range of products including payment smart cards and electronic passports, achieving high-levels of CC certification. The initialisation, and personalisation of these products has strict security controls, meaning that only reviewed code and data from certified developers can be loaded. An introduction to MULTOS can be found in [97]. The Trust-Anchor is intended as a deployed HSM for the Internet of Things (IoT) and differs from a smart card chip in that it has added I/O capability and free-runs. However, there are some drawbacks. The chips do not represent the state-of-the-art with respect CPU performance. However, they do excel in cryptographic operations because they incorporate relatively high-speed crypto-co-processor (CCoP) hardware; related performance evaluations can be found in [98] [99] [100]. The choice of SC means we can reasonably assume that physical, fault, side-channel and malware attacks are *practically infeasible*, and so our focus turns towards a protocol for securing the interaction between the SC and GA. A previous study for EMVCo [101] measured the performance of AE [102] for future payment card processing with MULTOS as a test platform. A number of AE modes were implemented; however, for small message sizes (up to 32 bytes), encrypt-then-MAC (ETM) [102] was the most efficient due to the relative speeds of the CPU and CCoP. ETM is discussed next.

5.4.1 Authenticated Encryption

The ETM scheme (see Fig. 5.2) is *mechanism 5* in ISO/IEC 19772 [102], and offers a routine methodology featuring distinct encryption and MAC processes. The scheme has much to recommend it for securing grid communication. It provides authentication, confidentiality and integrity protection and has a counter for cryptographically chaining transmissions. If fixed-sized data are used (e.g. the optimum 32-bytes), then the counter may be predicted in the case of lost messages, avoiding the need for re-synchronism except when unavoidable.

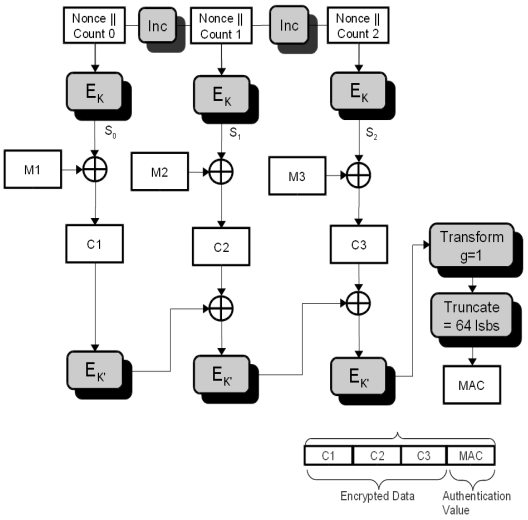

Figure 5.2: Encrypt then MAC Authenticated Encryption

5.4.2 Node Dynamic Initialisation

An SC would be initialised/personalised prior to use. This involves, for example, the setting (by a security authority) of IDs, long-term secret keys, access PINs, applications and data. Every node would be diversified, having different keys, nonces and counters, so attacking one node provides no advantage when attacking others. Because the long-term secret keys are of best-practice size and do not leave the Trust-Anchor, it is acceptable to use them in normal operations to create session keys, that may be refreshed as necessary. However, it is even better to use them sparingly. The session key establishment process also provides opportunity to dynamically change the nonce and counter values. The protocol steps for dynamic initialisation and session key generation are shown in Fig. 5.3, with symbols in Table 5.1.

The GA begins by generating AE session values: a random number, a random nonce and a starting counter value (random or preset). These values fill two AES blocks and are AE encrypted using the long-term secret keys and the *default* nonce and counter values for the particular Trust-Anchor. The resulting cipher blocks and MAC are then sent to the Trust-Anchor, and AE decrypts the message into two plain-text blocks and computes its own version of the MAC. If the local and received MACs do not match, the protocol ends with a failure. Otherwise, the plain text is copied into the local copies of the random number, nonce and count, and the random number and its increment are then AES encrypted under the long-term encryption key to generate the encryption and MAC session keys, respectively.

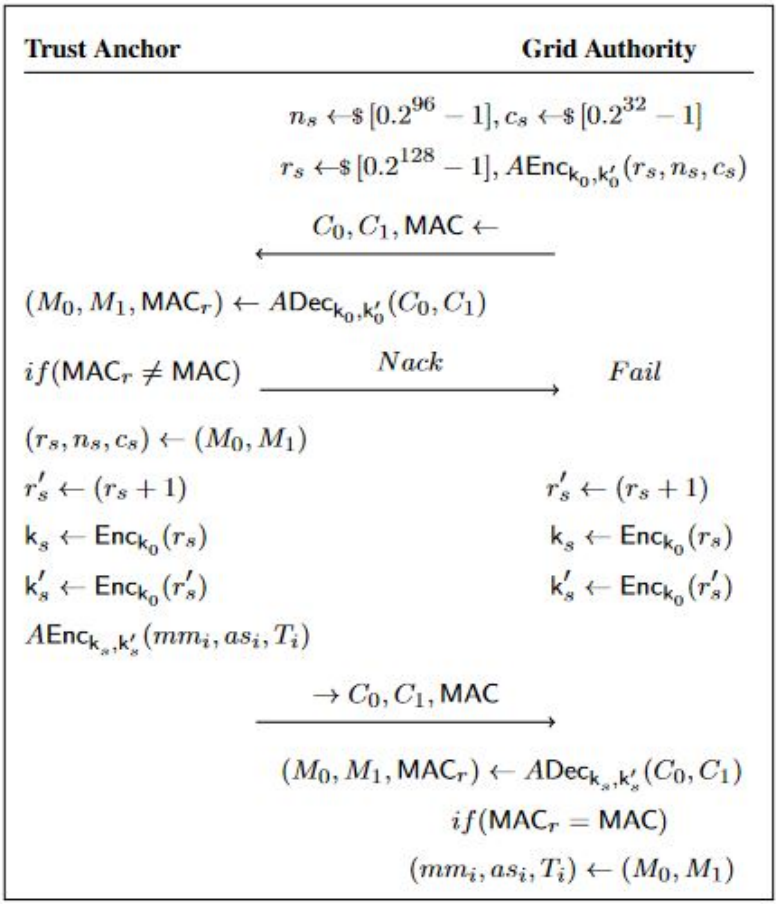

Figure 5.3: Dynamic Initialisation

The Trust-Anchor then AE encrypts a dummy measurement report under session keys and sends this to the GA (which also generates the keys). The GA AE decrypts the message using session keys, and if received and locally generated MACs match, the session key establishment has succeeded. The received data are copied locally and the session keys are operational. The initialisation may be repeated to change the AE session keys, nonce and counter values, either as security policy or due to lost synchronism. In the latter case, the AE count for initialisation is reset to the default for the particular Trust-Anchor.

5.4.3 Data Collection Processing and Reporting

Referring to Fig. 5.4 and Table 5.1, the Trust-Anchor is continually sampling the measurement source and storing results in a cyclic buffer. The GA begins a dialog by AE encrypting under the session keys, model control fields, model parameters and current time. The resulting cipher blocks and MAC are sent to the Trust-Anchor, which AE decrypts the message into two plain-text blocks and computes its own version of the MAC. If the local and received MACs do not match, the protocol ends with a failure. Otherwise, the plain text updates the local copies of model controls, parameters and real-time. The Trust-Anchor then runs the detection algorithm on the buffered samples, calculates the model measurements, updates the alarm status, creates a timestamp, generates a local alarm (if merited), subsequently AE encrypts the results under the session keys and sends them to the GA. If the MAC verification succeeds, the report is accepted into the modelling and state estimation application. If the Trust-Anchor does not respond or verification fails, the GA may attempt lost message recovery.

5.4.4 Lost Message Recovery

A goal of the protocol is to tolerate the loss of at least two consecutive measurement reports without the loss of sample data. The reported model measurement field includes the latest sample plus up to 7 previous samples, allowing data recovery even if blocks are lost. However, the AE counter advances during encryptions/decryptions, and using the wrong value will prevent correct decryption and MAC verification. Because the message sizes are always the same, the counter value after a lost message can

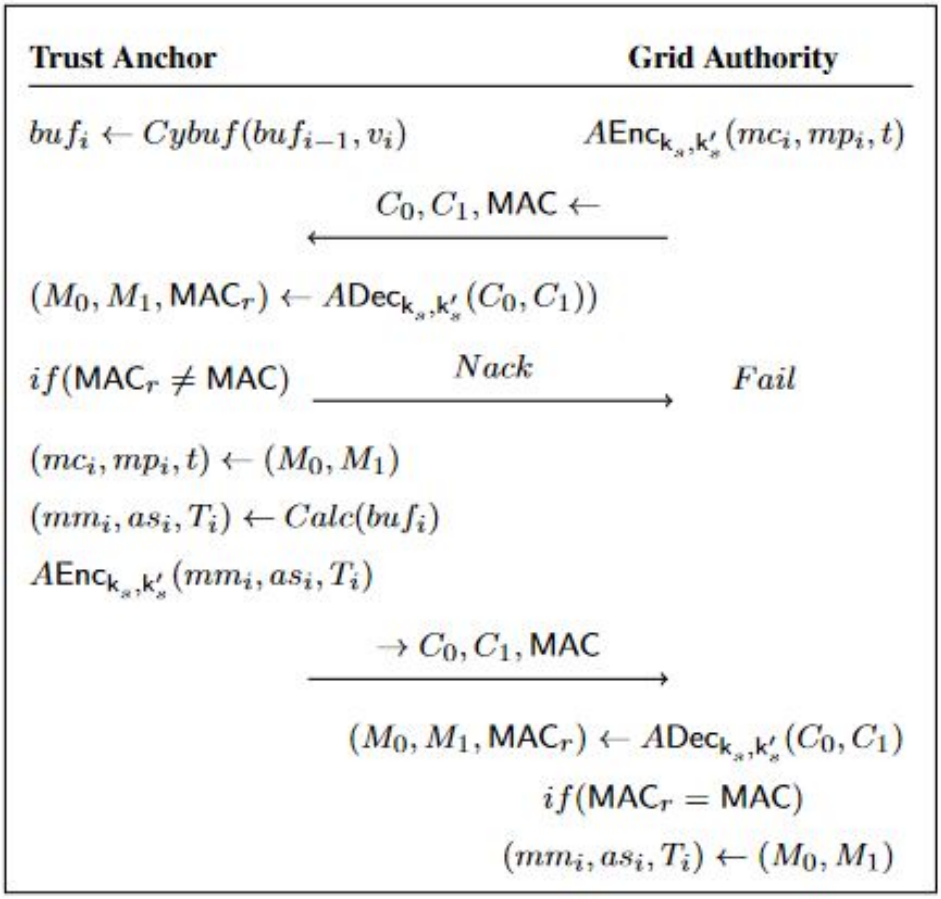

Figure 5.4: Data Collecting and Reporting

be estimated (within a small window), and the GA can attempt AE decryption with several alternative counter values. If successful, they will not lose any reported data. Similarly, at the Trust-Anchor, a MAC failure could indicate a missed or corrupted report request, and the next count values could be tried for the AE decryption. In the case of the irretrievable loss of synchronism, the GA will recover it by using the dynamic initialisation process.

5.5 Event-Triggered Moving Target Defence

5.5.1 Moving Target Defence

MTD involves using the system assets to change the underlying topology and expose FDI attacks. FDI attacks require the attack vector to be structured based on the network topology. For the real power residual, the error at the individual measurement level will be the difference between the measured

Table 5.1: Symbol Definitions

Symbol	Description
ID	Security Chip ID (128-bit, or 80-bit if ICCID))
$_0, _0'$	Personalised long-term encryption and keys (128-bit)
$_s, _s'$	The current session encryption and keys (128-bit)
r_s, r_s'	Random values for session key generation (128-bit)
n_s	The current AE session nonce (96-bit)
c_s, c_i	Starting and ith transmission, AE session counts (32-bit)
M_j, C_j	The jth message and cipher blocks (128-bit)
$_i, _r$	The ith AE token, and recomputed token (MAC: 64-bit)
t, T_i	The time and ith Timestamp (32-bit)
as_i	The alarm status information 24-bit)
dv_i, mm_i	The ith sample (15-bit) and model metrics (32-bit)
mp_i, mc_i	The model parameters (208-bit) and controls (16-bit)
$A_{y}, _{y}'$	Authenticated Encryption under key set y
$A_{y}, _{y}'$	Authenticated Decryption under key set y
$_y$	AES Encryption under key y

flows and estimated value from the system model, such that the real power residual can be expressed as a function of the individual meter measurements and system states by

$$\begin{aligned} r_{ij}^{P} = &-P_{ij}^{m} + V_i^2 g_{ij} - V_i V_j g_{ij} \cos\Delta\theta_{ij} \\ &- V_i V_j b_{ij} \sin\Delta\theta_{ij}. \end{aligned} \tag{5.1}$$

With a similar equation for resultant changes in the reactive power measured residual. If the attacker is aware of the system topology, he/she can structure the attack such that the residual is small. However, if changes to the physical system are introduced (via inductance modification in this case) and the attacker does not consider this in the updated attack vector $\mathbf{z}_a$, the residual will increase as differences between the expected and injected value emerge. Where ignoring the change in voltage angles and magnitudes for a simplified view, the impact to the residual will be approximated by

$$\Delta r_{ij}^{P} \approx V_i V_j \Delta b_{ij} \sin\Delta\theta_{ij} \tag{5.2}$$

MTD can also be enhanced to camouflage its existence to minimise the potential for attackers circumventing it [48] [38]. However, as shown in [40], the cost of application will mean that the system operator will want to minimise the overall use of MTD to only those times when the system is potentially under attack. As a result of this, we propose an MTD-triggering scheme based on individual meter measurement deviations.

The proposed defensive process has three main components: the anomaly detector, active defence protocol and final attack verification, which exploits the secured measurement and communications authentication provided by the low-cost MULTOS devices. The anomaly detection can be done either locally at the distributed level or via post-processing at the same place as the CSE. For deciding the error bounds to implement the trigger, we use Holt's exponential smoothing and forecasting and have outlined this process below.

5.5.2 Anomaly-Detection-Based Triggering Strategy for MTD

We propose a distributed trigger (located at the individual measurement level) based on Holt-Winters' exponential forecasting for anomaly detection. This distributed layer of detection confers several benefits to the grid. For one, distributed error checks can be preformed much more quickly than the CSE, which can take up to a few minutes and may not even converge. Distributed measurements devices can also offer some localised control options in the case of a wider network failure of the SCADA or communications network. Crucially with respect to FDI attacks, distributed measurement-based anomaly detection will not be susceptible to the same model-based FDI attacks outlined in [3]. Recent papers have shown that CSE can be attacked by structuring the attack vector in terms of the system topology, i.e. $\mathbf{z}_a = \mathbf{h}(\mathbf{x}+\mathbf{c})$ where $\mathbf{c}$ is a desired attack bias of length n. Therefore, to circumvent this type of stealthy bias injection, we consider a distributed error detection based on forecasted measurements to act as trigger for MTD. We consider the measured vector at a given bus system as a discrete, linear model and follow this form:

$$\hat{z}_{t+1} = F_t z_t + b_t + \nu \tag{5.3}$$

Where $\hat{z}_{t+1}$ is the estimated metered power-flow measurements at a given time t, ν represents the system noise that will follow a Gaussian distribution such that $\nu \hookrightarrow \mathcal{N}(0, Q)$. F_t represents the state transition factor, i.e. the expected change in the system state for a given time period and b_t represents the state trajectory vector used to capture long-run trends and will incorporate factors such as seasonality. F_t, b_t and ν are calculated by analyzing previous data trends using Holt-Winters' exponential forecasting [103] and contain seasonal elements so as to capture differences in the intra-day measurement resulting from load peaks. Anomaly detection in dynamic estimation will be the difference between the forecasted state value and the measured value for a given time and is given by

$$e_t = z_{t+1} - \hat{z}_{t+1}. \tag{5.4}$$

With the assumption that residuals are independent and follow a zero-mean Gaussian process. They

can either be defined by using static pre-set limits or, more commonly, by using an updated variance value σ^2 from observed residuals. The introduction of this additional criteria means that to attack stealthily (or at least to not trigger MTD), the attacker would have to now satisfy both the centralised and distributed criteria such that

$$\{\mathbf{z} \in \mathbb{R}^{m\times 1} \wedge ||\mathbf{z} - \mathbf{h}(\mathbf{x})||_2 < r_c \wedge z - \hat{z} < r_d\} \tag{5.5}$$

where r_d and r_c are the alarm limits for distributed and centralized triggering/detection. It is possible that an attacker may be able to structure an attack in this manager, but their flexibility in which loads can be overloaded will be greatly reduced.

5.5.3 Selection of Alarm Limits

Alarm limits can be imposed considering a number of factors.

Type-II Error

A type-II error is an important consideration for alarm limits. It might be tempting to set limits artificially low to capture more potential attacks. However, this could result in higher unnecessary costs to the system operator because MTD will be triggered frequently when no attack persists. Frequent false errors can also reduce reasonable responses because the system operator will grow used to seeing a high number of alarms and lack the bandwidth to verify them efficiently.

Criticality and Capacity

Individual measurement can also critically weigh into limit setting. A system operator may wish to lower limits for critical measurements. Conversely, areas with high additional transmission capacity operating well below their thermal limits will suffer fewer consequences from overloading FDI attacks. As a result, alarm limits could be relaxed to reflect their respective innate resilience of regions and focus on other weaker areas.

We prioritized type-II error considerations in our model and opted for limits based on the probability distribution function of a Gaussian distribution with 3 standard deviations (SDs) corresponding to a 99.7% confidence interval (or a type-II error of around 0.3%). This is in contrast to CSE, which usually operates with a 2-SD limit [3]. We believe it appropriate to have different alarm limits for these two approaches. For one, the frequency and number of distributed checks can be much higher for distributed decision-making. Unlike CSE, there are multiple potential alarm measurements, i.e. there are 34 alone in a 14-bus system, as opposed to CSE, which is only one centralized measurement. Also, as discussed, there is a cost in applying the MTD from an operational power-flow perspective, and it may not be desirable to use MTD triggering constantly when the branch power flow level offers no danger to the system. Ultimately, the selection of alarm limits comes down to engineering judgement. For highly critical regions, a system operator may tolerate increased false positives from a lower limit so as to increase sensitivity.

In our model, the upper limit is defined by $\lambda_{upper} = z_t + 3\sigma$, the lower limit is defined by $\lambda_{lower} = z_t - 3\sigma$ and they are bound such that

$$\lambda_{lower} < e_t < \lambda_{upper}. \tag{5.6}$$

For the power system models, we assume an intra-day profile similar to that of a consumer load flow. This means that our forecasting model needs to approximate the underlying seasonality in creating the prediction model. The Holt-Winters method is used to capture seasonality and trend when forecasting data sets. The method is made up of a forecasting equation and accompanied by three equations for the current level L_t

$$L_t = \alpha(\frac{z_t}{S_t}) + (1-\alpha)(L_{t-1} + T_{t-1}). \tag{5.7}$$

The overall trend of the data set T_t is described by

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}. \tag{5.8}$$

And the seasonality component of the data set S_t is described by

$$S_t = \gamma(\frac{z_t}{L_t}) + (1-\gamma)S_{t-p}. \tag{5.9}$$

Where p is the time period in a season. α, γ and β are the weighting values for level, trend and seasonality, respectively. These are selected for a minimised RMSE error based on observing differences between the model estimates and the actual values. The final predicted value using HW will be described by

$$\hat{z_t} = (L_{t-1} + T_{t-1})S_{t-p}. \tag{5.10}$$

5.5.4 Computational Considerations

It is assumed with respect to the CSE (which is already commonly used) that operators already have the capability to perform the state estimation for their system. With regard to the computation cost of applying the Holt-Winters distributed forecasting for the event trigger, the impact is small. As shown, the basic calculations are linear and can be performed quickly even with a large number of measurements. Also, as discussed, because the method is distributed to the individual meter level, the computational cost of this approach is not dependent on the size of the system, i.e. we would not expect real impact from the curse of dimensionality resultant from system size because the core calculations and regressions are still dependent on individual measurements only.

5.6 Experiments and Findings

This section assesses the performance of the proposed detection strategy on both the standard IEEE 14-bus test system Fig. 9.1) and the IEEE 118-bus test system Fig. 9.2. All grid simulations were implemented using the MATPOWER toolbox in MATLAB [72] and performed using Intel Core i7-7820X CPU with 64 GB of RAM running on a Windows 10 system. The node performance exper-

iments were performed in collaboration with Professor Keith Mayes at Royal Holloway University and were carried out using a MULTOS Trust-Anchor development kit.

5.6.1 MULTOS Trust-Anchor Performance

The use of the MULTOS Trust-Anchor satisfies the node attack-resistance requirements of the proposed system. However, it is a CPU speed- and memory-restricted device, so its ability to satisfy performance requirements needs to be determined by experiment. The GA is assumed to have access to powerful server capability, so only the measurement node performance was practically investigated. The Trust-Anchor chip used for the research had digital I/O but no analog-to-digital converter (ADC), so an external device was connected via its I2C bus. This did not change the attack assumptions because it was already accepted that the analogue source value might be modified, so this just extended to the ADC chip and the I2C bus. The ADC can sample orders of magnitude faster than needed, so the effective sample rate is determined by the Trust-Anchor reading from it. Normally, the Trust-Anchor would be required to free-run, but for precise experimental measurements, it was run in command/response mode, where actions were triggered by commands from the GA. The Trust-Anchor was used within a breakout board that allowed PC control via a USB port. GA messages were manually created and sent to the Trust-Anchor via the MULTOS MUTIL scripting utility. MUTIL logs include message timing, and a millisecond timer was also configured within the Trust-Anchor as a calibration check. Test software was in 'C' within a single application. For message timing precision, commands were run at least 64 times before response to compensate for residual measurement tolerance.

The initial performance results are presented in Table 5.2. Entry row eight is the worst case performance test. It assumes that a GA request arrives and must first be AE decrypted. The request requires the Trust-Anchor to acquire a new sample, add it to the cyclic buffer, compute the model means, test the sample and means against the model thresholds and finally AE encrypt the resulting report for transmission back to the GA. The total execution time is 82.83 ms, supporting a maximum repetition rate of approximately 12 Hz (12/s). This suggests that the Trust-Anchor could support more sophisticated, fine-grained local processing of measurement data as part of a distributed anomaly detection

Table 5.2: Trust-Anchor Results

	Function	Time	Blocks	Total	Max Rate
		(ms)		(ms)	(Hz)
1	ADC Read Loop	0.72		0.72	1391.30
2	As(1)+ update cyclic buf	1.70		1.70	587.16
3	Ad(2)+ calc means	12.88		12.88	77.67
4	As(3)+ check thresholds	33.52		33.52	29.84
5	ETM decrypt	12.20	2	24.41	
6	ETM encrypt	12.45	2	24.91	
7					
8	Sample Test and Report			82.83	12.07

system.

5.6.2 Anomaly Detection

As shown in [104], the UK transmission network typically operates at around 80% capacity. Given this, we opted to apply a 15% change in the system power flows at the attacker target. This would give the attacker enough flexibility to start to push these limits (at least at a regional level). In Fig. 5.6 and 6.2, we show the anomaly detection algorithm implemented for an FDI attack line overload of 15%. There is an initial training period for the HW seasonal forecasting for a system under normal operation. The focus is on individual line overloads, and Fig. 5.6 shows line 1-5 for the 14-bus system for 400 hours of operation with a dual peak profile occurring over 24 daily measurements. We show the measured data set in blue and the upper and lower bounds as predicted by the Holt-winters forecast in red. As we can see, the first few days of implementation require a data training period and have high volatility in the forecast model and a high corresponding type-II error. However, this declines as the model has sufficient data to train over. Once this training period is completed, the data sets fit closely, and we can see that no anomalies are detected over the next week of operation past the training period. We see a similar result for the IEEE 118-bus system in Fig. 6.2.

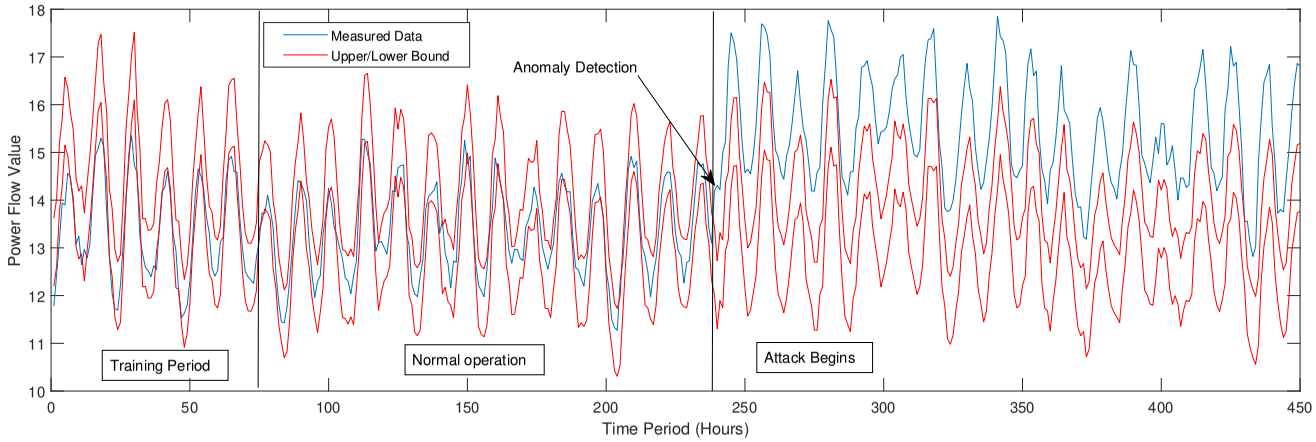

Figure 5.5: Power flow measurement across line 1-2 in the IEEE 118-bus system over a 24 hour period. System measurements shown in blue with upper/lower bounds designated via hole-winters seasonal forecasting in red. Three time periods shown: initial training period for the forecasting, followed by under normal operation and under the FDI attack initiated.

In Fig. 5.6 & Fig. 6.2, we initially continue basic operation for the 10 days before implementing the FDI attack. This attack is stealthy from a centralised perspective but is detected by our distributed anomaly detection. This illustrates the effectiveness of the anomaly detection, which can evaluate potential FDI attacks at a distributed level (we show that these attacks would otherwise bypass CSE). It should be noted that there is potential for false positives with this kind of anomaly testing. Because the forecasting is based on prior measurements, if new loads are added, it may take a few cycles for the forecasting to reflect the new reality. However, these type-II error anomalies for the distributed detection should occur relatively rarely (0.3% of measurements in a 3-SD operating window or around 5% for a 2-SD configuration) and can be configured based on the acceptable ranges of the system operator. Furthermore, the occasional false triggering only leads to the activation of MTD to confirm the alarm, and hence, the overall impact on system operation or cost is limited.

5.6.3 Event-Triggered MTD

In this subsection, we demonstrate the effectiveness of the proposed event-triggered MTD. The study is carried out between the hour 150 and the hour 300, as equivalent to Fig. 5.6 and Fig. 6.2. We first present the base case and assume no FDI attack is implemented. As shown in Fig. 5.7 & 5.11, the residual value of BDD in CSE under normal operation is well below the alarm limit. It should be noted that although there is some initial type-II error, this is not due to the FDI attack and is, in fact, consistent with a 95% chi-squared test with one anomaly point over the measurement period. In the cases where the FDI attacks are applied, we assume a stealthy FDI attack is launched from hour 240. In the case with only traditional BDD in CSE, Fig. 5.8 & 5.12 show that the residual value does not increase tangibly over the no-attack case (3 type-II errors still consistent with 2-SD confidence), which means this FDI attack is not detected. However, once the proposed protocol is applied, as shown in 5.9 & 5.13, a marked increase in the overall system residual is visible, which is due to the triggered MTD, as shown in Fig. 5.6, and the attacker's reliance on an outdated model. It should also be noted that the proposed protection protocol does not affect the centralized alarms when no attack is present, as shown between hour 240 and hour 300. In the case of false triggering of MTD, 5.10 & 5.14 show that even though MTD is applied, there is no tangible increase of residual in the

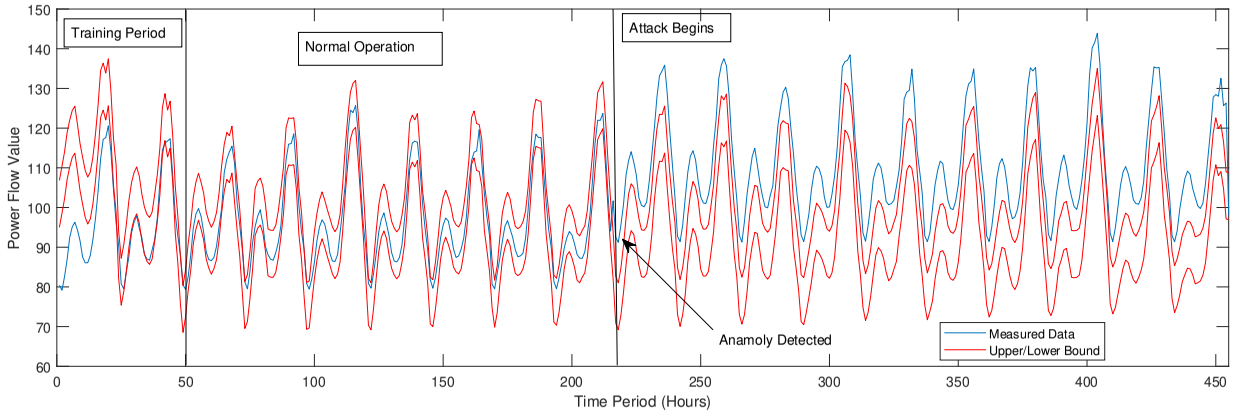

Figure 5.6: Power flow measurement across line 1-5 in the IEEE 14-bus system over a 24 hour period. System measurements shown in blue with upper/lower bounds designated via hole-winters seasonal forecasting in red. Three time periods shown: initial training period for the forecasting, followed by under normal operation and under the FDI attack initiated.

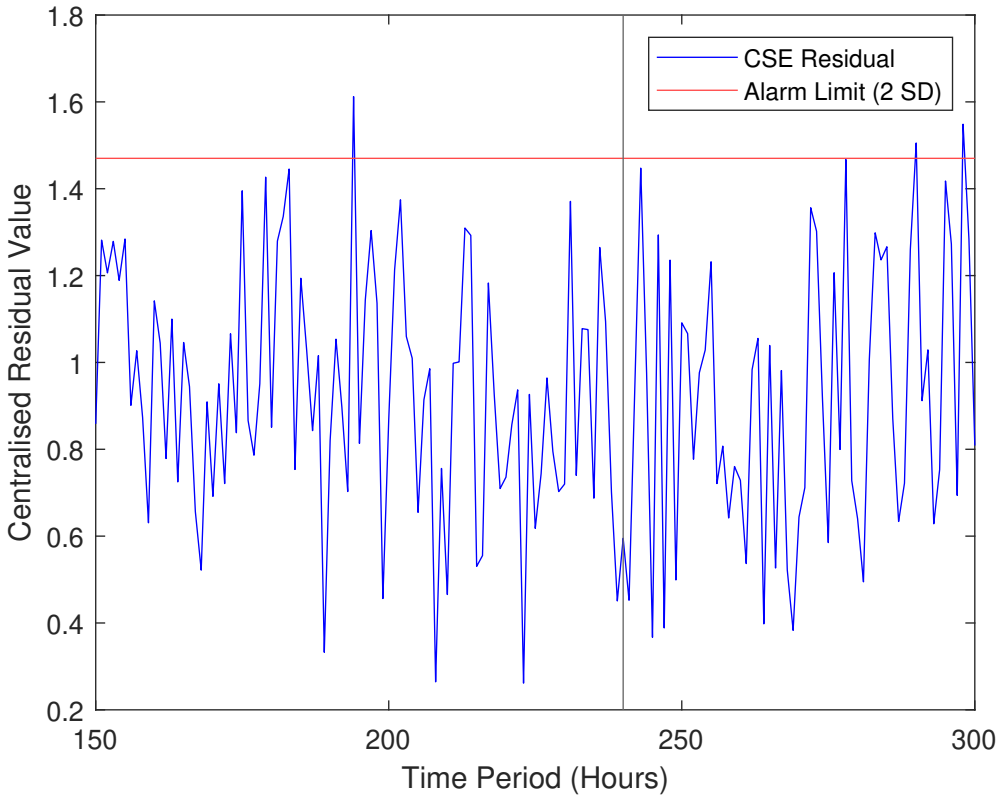

Figure 5.7: Residual value for CSE under no FDI attack for the IEEE 14-bus system.

centralized alarms because the new set of measurement reflects the updated system physical topology. In this case, no false alarm for cyberattack will be raised.

While it is clear that this type of event-triggered approach can help to greatly reduce the overall application of MTD, it should be noted that this kind of of method would not really be effective for a replay-style FDI attack. In fact, we would expect that for a replay-style attack, the residual profile would look like those seen in Fig. 5.6 and Fig. 6.2 and likely be indistinguishable from normal operation. The reason for this should be self-evident. Replay-style attacks (in the vast majority of cases) will be non-anomalous because they merely select previously used load profiles. In fact, to insure against this style of attack, it would be advisable to include some kind of time-based MTD protocol on top of the event-based protocol. This would ensure that replay-style vectors are eventually evaluated even if they bypass in-cycle anomaly-driven detection. Even with a time based protocol for MTD implementation a semi-sophisticated replay style attack could be problematic. Consider the in-cycle attacks proposed in [105]. If attacks can be performed between MTD implementations, using replay style vectors (which will look plausible) there is a high probability of success. On the other

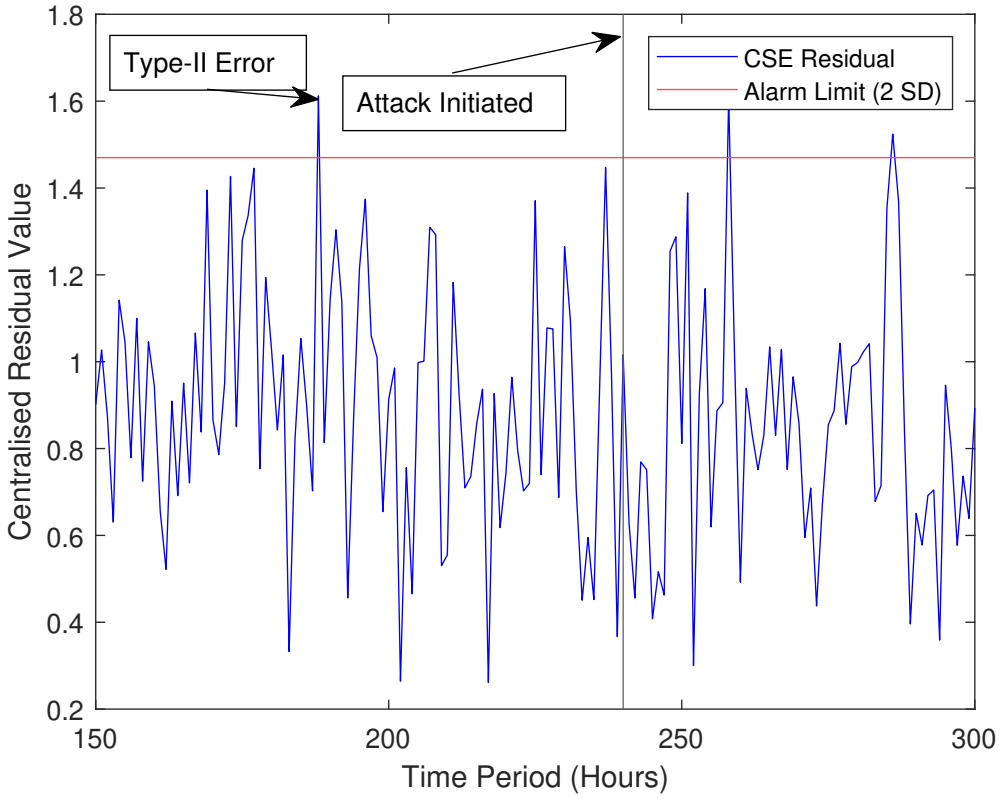

Figure 5.8: Residual Value for CSE under Stealthy-FDI Attack Applied from 240 hours without MTD for the IEEE 14-bus system.

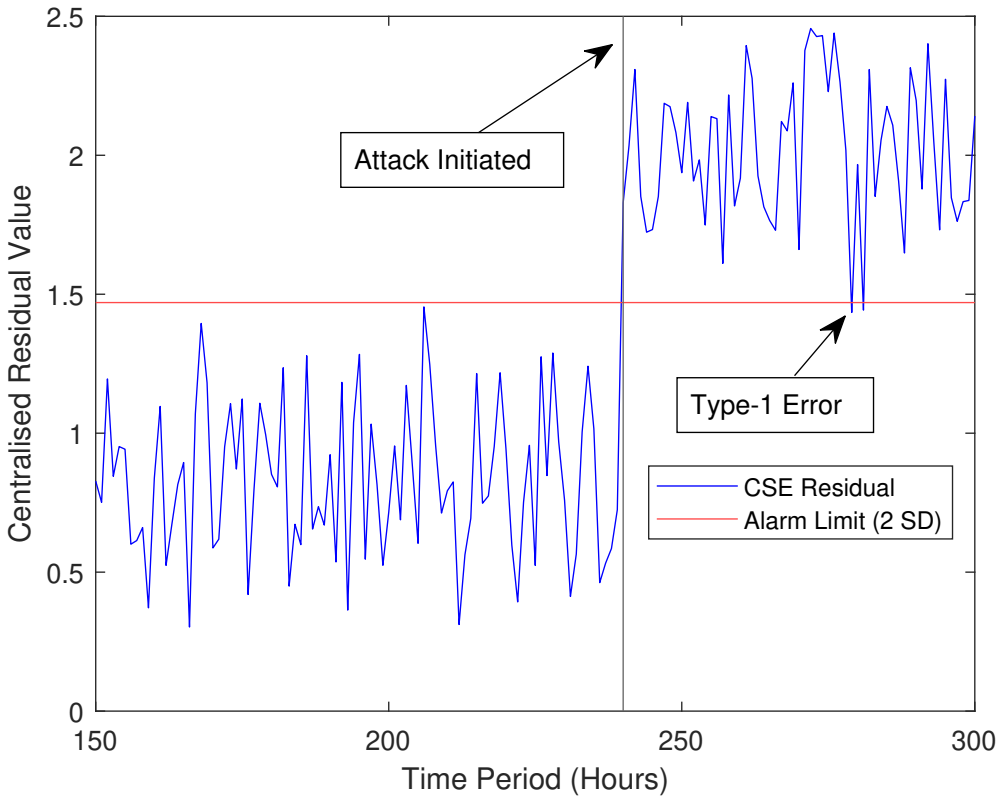

Figure 5.9: Residual Value for CSE under Stealthy-FDI Attack Applied from 240 hours with event triggered MTD for the IEEE 14-bus system.

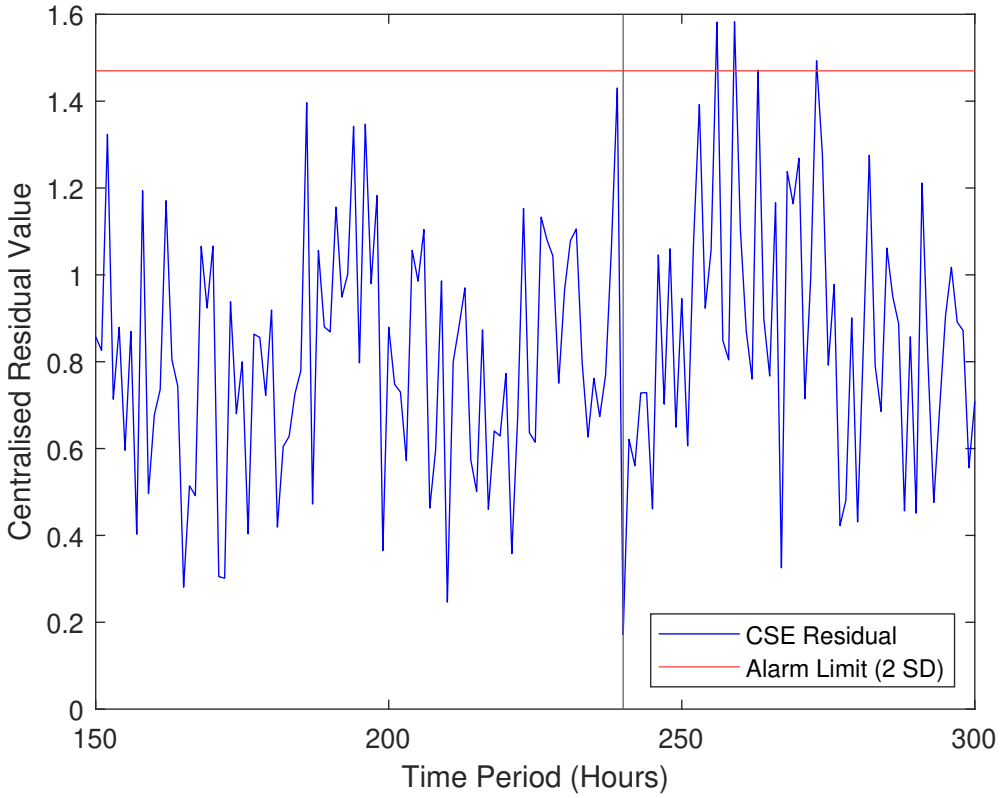

Figure 5.10: Residual Value for CSE under no stealthy-FDI attack with event triggered MTD triggered at 240 hours for the IEEE 14-bus system.

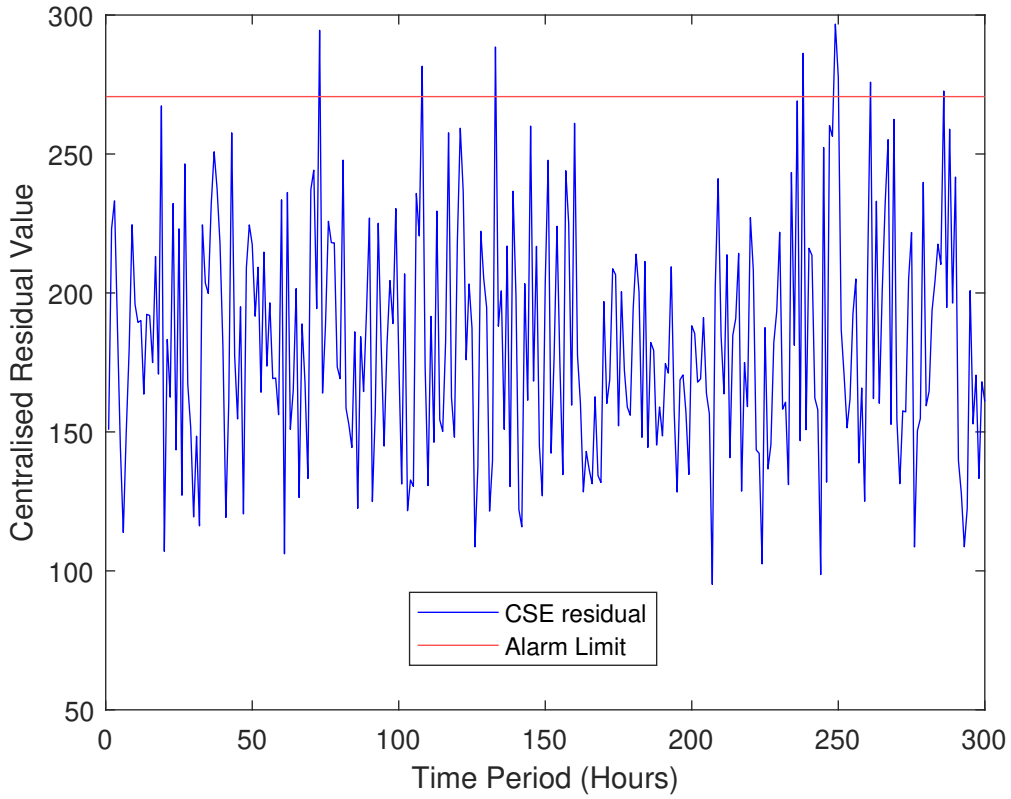

Figure 5.11: Residual Value for CSE under no FDI Attack for the IEEE 118-bus system.

hand, we might expect replay style attacks to have a lower capacity for damage as they are simply playing back old messages.

5.6.4 Security Attack Coverage

Having experimentally proven that the performance of the proposed solution is well within our design requirements, we now recap on the security attack coverage. Choosing the MULTOS Trust-Anchor provides a highly attack-resistant hardware security end-point, supporting secure the loading/configuration processes and making node implementation attacks *practically infeasible*. The use of open/standardised algorithms and diversified credentials also removes the motivation for sophisticated node implementation attacks. The use of AE prevents eavesdropping through encryption and message faking or modification via MACs. The use of a counter within the AE mode, and chaining of transmissions, prevents malicious re-ordering, and a timestamp detects message delay. The protocol design is resilient to some lost messages (whether natural or malicious) without the loss of sample

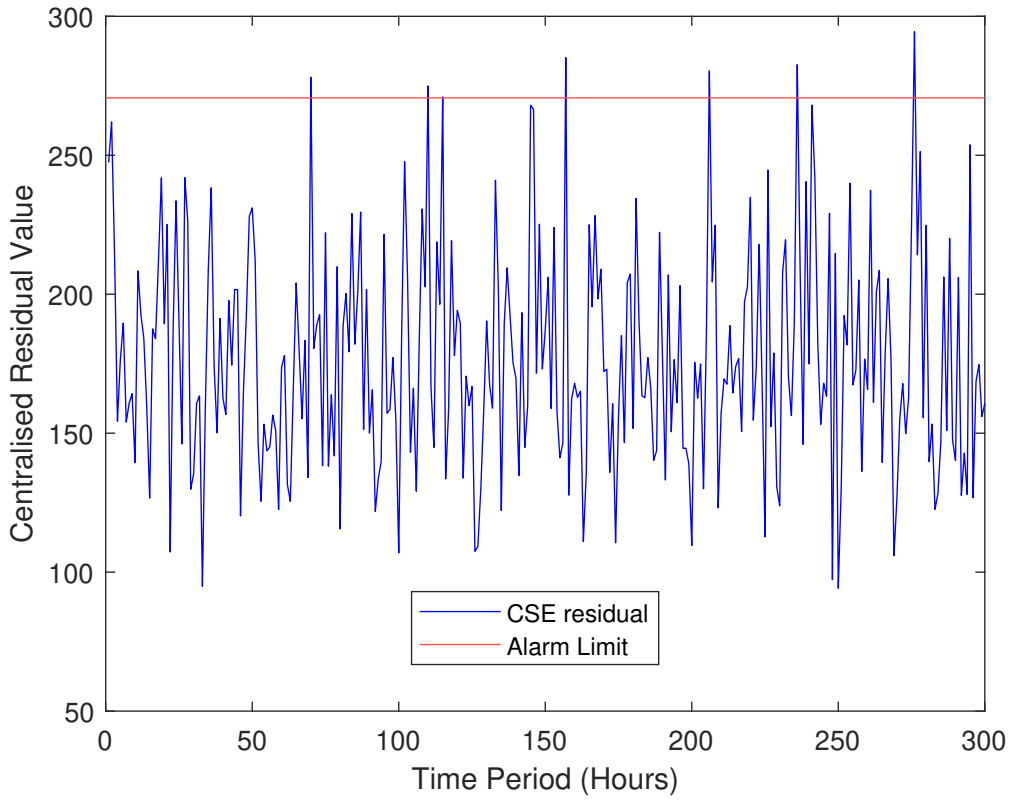

Figure 5.12: Residual Value for CSE under Stealthy-FDI Attack Applied from 240 hours without MTD for the IEEE 118-bus system.

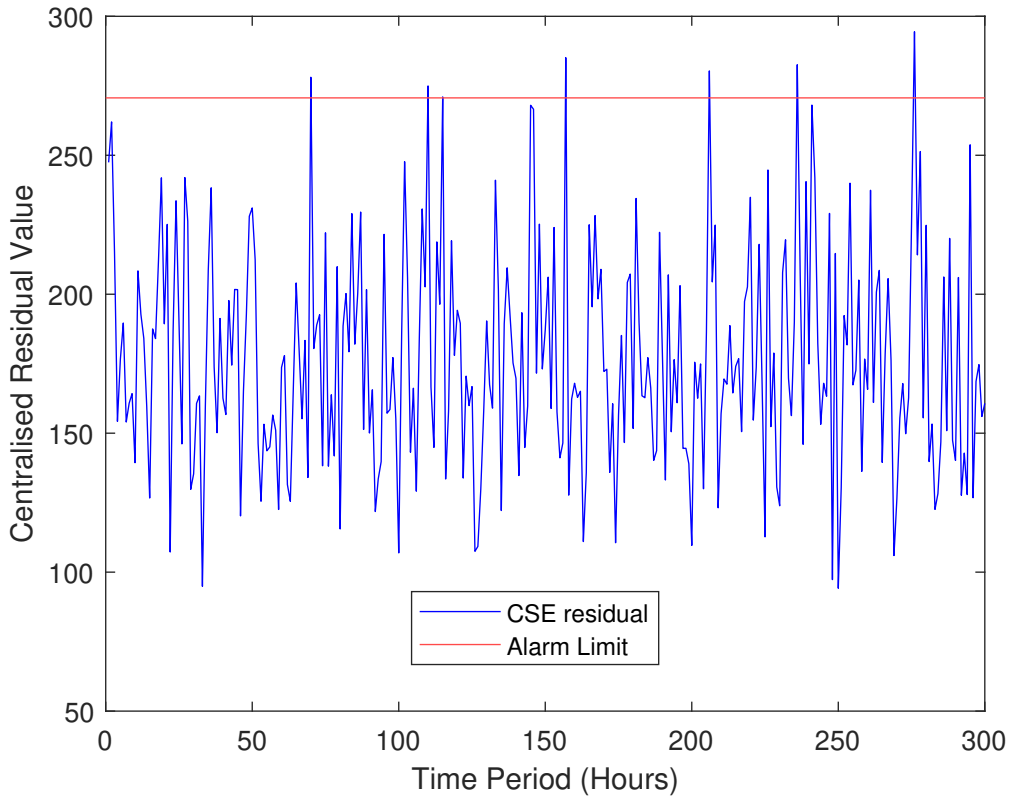

Figure 5.13: Residual Value for CSE under no stealthy-FDI attack with event triggered MTD implemented at 240 hours for the IEEE 118-bus system.

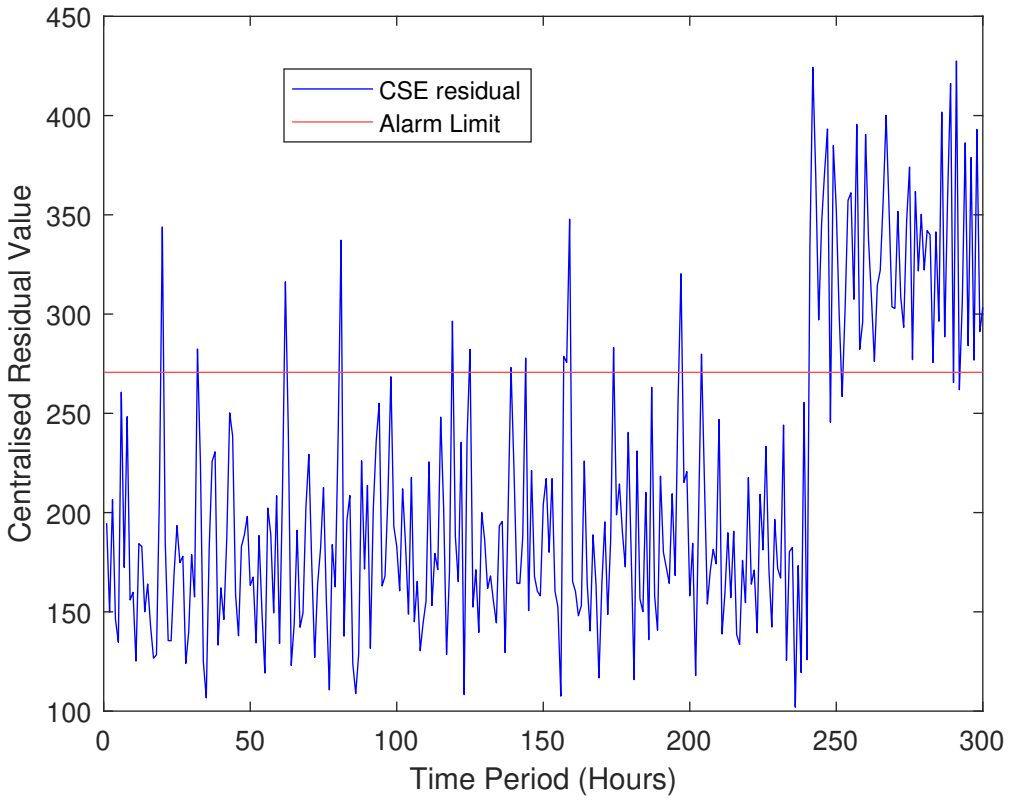

Figure 5.14: Residual Value for CSE under Stealthy-FDI Attack Applied from 240 hours with event triggered MTD for the IEEE 118-bus system.

data. The protocol also has resilience to the loss of synchronism in counter values, and the ability to re-synchronise when necessary, with refresh of session keys. Normal operation cannot continue under DoS attack. However, the condition is detectable by the GA and by nodes, and nodes are capable of raising a local alarm when denied communication to the GA. The nodes also have the capability to continue stand-alone passive data monitoring to detect extreme cases of source modification. Stealthy source modification is detected by event-triggered MTD. Malicious disablement of node local alarms can be detected by similar means.

5.7 Summary

This paper proposes a combined cyber-physical authentication protocol for secure and reliable state estimation of power grids in the presence of malicious actors. The solution combines distributed measurement nodes based on small low-power, low-cost, SCs via the MULTOS Trust-Anchor with a physical system event-triggered MTD protocol featuring distributed and centralised anomaly detection. These low cost, attack resistant measurement nodes can be used to dictate distributed anomaly detection and trigger MTD to help enhance the security of the power system.

On the MTD front we introduce the concept of an event triggered MTD. As we have seen previously in [82] MTD is costly and requires the use of assets in a non-efficient manner. We can reduce this by using anomaly driven MTD, namely that instead of applying MTD continuously (costly) or on a time driven protocol (risk of in-cycle attacks between MTD implementations) [105]. As we had a distributed, attack secure measurement solution in the form of the MULTOS devices we can reliably implement distributed solutions for anomaly detection for MTD. This makes further sense as the MTD effective of individual branches is limited to the local subgraph so therefore it makes sense to target your MTD implementation locally.

Simulations were performed on both the IEEE 14-bus and IEEE 118-bus systems to show the effectiveness of the proposed MTD and anomaly detection protocol in deriving FDI attacks. We used the holt-winters forecasting of branch power flows and looked for deviations from the norms in order to drive the event trigger. We used a 3SD difference from the expected mean operating point (adjusted

with a seasonal component using the holt-winters forecast) to act as the alarm point for the event trigger. Because the alarms are distributed, we effectively have an alarm measure for each possible branch flow in the network. There are a few drawbacks of this approach. It is possible in a system prone to large changes (i.e. loads being swapped out, bi-directional loads, storage implementations ect...) this distributed approach may become less effective. Large variations, could result in large numbers of alarms being sent. If the model is adjusted to these large variations it could also result in attacks being able to be hidden from the forecasting.

Practical experiments demonstrated that the Trust-Anchor could satisfy the most demanding GA request at a rate of 12/s; this is more than fast enough for envisaged operational scenarios. The communication protocol is resilient to missing measurement reports without loss of data and without the need for retries. The analysis of potential attacks showed that most were countered by the choice of Trust-Anchor and the protocol design. Passive monitoring by the node (using the delegated model) detects extreme measurement source modification. DoS will disrupt operation but is a detectable alarm event at the nodes and the GA, and a node's passive measurement monitoring will continue under DoS attack. On top of these communications protections, the physical protection protocol looks for distributed anomalies and probes for stealthy FDI attacks using grid system assets.

5.8 Lessons Learnt

Crucially, we draw the following lessons from this chapter, which we use to inform our research and proceed to the next chapter in this work.

- Event triggers are a reliable method of MTD and can crucially reduce the overall application of MTD in the power system. They provide adequate and effective alternative to time-based or continuous application approaches.
- It is also possible via Holt-Winters-style forecasting to provide a distributed trigger that does not need to be centrally checked and can provide low level protection without increasing communications bandwidth.

In the next chapter, we start the risk assessment of FDI attacks as a whole-system affair. We consider and discuss how reinforcement should be considered and how the system topology will inherently incentivize certain forms of reinforcement for differing targets.

Chapter 6

Multi-Layered Risk Assessment for FDI Attacks in the Presence of Moving Target Defence

The factors that influence the success of a false data injection (FDI) attack are multifaceted. Many works consider the FDI attack in the context of the ability to change a measurement in a static system only. However, successful attacks will first require intrusion into a system, followed by the construction of an attack vector that can bypass bad data detection (BDD). In this work, we develop a full-service framework for FDI risk assessment. The framework combines both the costs of system intrusion via a weighted graph assessment in combination with a physical, line overload-based vulnerability assessment. In this way, our cyber model considers meter intrusion, RTU intrusion and combined-style attacks, whereas our physical model analyses requires a level of topology divergence to protect against a branch overload from an optimised attack vector.

6.1 Introduction

The contemporary power system is a cyber-physical system with high levels of system inter-dependency and a near ubiquitous use of communications throughout. The move towards a cyber-physical system

has resulted in new vulnerabilities that have not been fully covered by the existing defence frameworks. Such vulnerabilities were exposed during the 2015 cyber attack against distribution companies in Ukraine [1]. Attacks like these have increased focus on the area of power system cyber security, and although many papers have focused on designing new attacks and novel defences, relatively few have focused on risk assessment of specific cyber attack types within the context of a cyber-physical system. In this paper, we develop a cyber-physical model of risk assessment that considers the inherent risk of a system topology, the interconnection between RTU and telemetered measurements and the consideration of attack plausibility in light of the active defence technique Moving Target Defence (MTD).

6.1.1 Proposed Risk Assessment Framework

Although prior works have offered a variety of methods for the risk assessment of critical infrastructure systems, they fail to cover the modern paradigm of FDI attacks and their defence mechanisms. Our cyber-physical risk assessment framework builds on these cases, particularly [58], [53] [51] & [52], to create overarching risk assessment criteria that consider both the intrusion (cyber) component of risk and stealthiness in the presence of MTD-portion (physical) risks. Our work considers both the cost of intrusion to a given attack point and the ability to remain stealthy in the presence of MTD system capabilities. To this end, we combine weighted min cost of intrusion modelling and the level of MTD required to protect a measurement when assessing risk to create a cyber-physical risk assessment approach. The closest paper in terms of contribution to this work is [51]. However, our risk assessment framework improves upon this work in the following ways:

- First, our model provides a weighted graph assessment of the FDI attack intrusion risk of the cyber components of the grid. In addition to FDI vulnerabilities or RTUs and meters, our model includes (for the first time) overlapping-style attack opportunities, i.e. not simply the choice of the RTU or the meter combinations for a given state but also some combinations of the two.

- We also introduce an MTD (post-intrusion) effectiveness criteria that considers system capacity constraints in the context of an FDI attack and the required level of MTD to expose an attack

for an overload-style attack. We model the level of divergence required to protect each bus and branch combination in the context of a min possible attack vector.

We next outline the threat model for the risk assessment framework.

6.2 Cyber-Physical Threat Model

6.2.1 Attacker Assumptions

We make some assumptions about the prospective attacker that help define our risk assessment model. These assumptions outline the modus operandi with respect to both the intrusion and system change elements of the attack.

- **System Intrusion**: The attacker is attempting an FDI-style attack and will capture meter measurements in the required sub-graph. His intrusion cost will be the cost of compromising the meter set required to change a given bus measurement, and they will seek to minimise this cost. The attacker can choose either to compromise the meter set, RTU or some combination to replicate the underlying attacking sub-graph.
- **System Change**: Once intruded, the attacker is attempting to create a simulated overload attack via FDI to the power flow profile. The attacker will attempt to optimise his attack vector to this effect, using the smallest possible attack vector needed to overload the given line.
- **Statistical Peak**: The attacker is conscious of the additional advantage peak loading can grant and will wait for such an instant before initiating his attack. We reflect this by performing simulations on the statistical peak.

Given these attacker aims, we first consider the intrusion risk in terms of (weighted) sub-graph capture cost. We then consider the ability of the SO to defend the system with MTD under the 'peak load'-style conditions.

6.2.2 Min Cost Point Capture Strategy

Previously, most works have envisioned an infinitely resourced attacker. In practice, attackers will be constrained in what elements they can compromise. They will likely choose targets based on ease to compromise and will prioritise low-cost targets. In this article, we outline formulations for weighted and unweighted bus-capture strategies. The unweighted number of meters to compromise MuC_n where k_m denotes whether a meter is present at the busbar or branch. This is represented by the number of non-zero terms in the column vector of $\mathbf{H}$ for a given node n and represents the number of meters (needed to compromise) to attack stealthily. This represents a simply unweighted cost that is shown by

$$k_m = \begin{cases} 1, & \text{if } \mathrm{col}_n(H_{n,m}) > 0 \\ 0, & \text{else} \end{cases}$$

$$MuC_n = \sum_{1}^{m}(k_m). \tag{6.1}$$

We can also add considerations of the difficulty to capture a given edge by adding a graph weighting. This makes sense given the co-existence of new and legacy measurement equipment in the power system. This can be represented with weightings for edges. For each edge E of graph G, there is an associated weight $w(m)$. This can represent either line redundancy or more meters, which are more resilient to attack. The weighted cost of meters MwC is shown by

$$p_m = \begin{cases} w(m), & \text{if } \mathrm{col}_n(H_{n,m}) > 0 \\ 0, & \text{else} \end{cases}$$

$$MwC_n = \sum_{1}^{m}(p_m). \tag{6.2}$$

Unlike in previous models, in our proposed intrusion model, the attacker can compromise either

the metering components (similar to previous interpretations of FDI intrusion), the RTUs or some combination ito replicate the needed sub-graph. We outline these approaches in the next subsection.

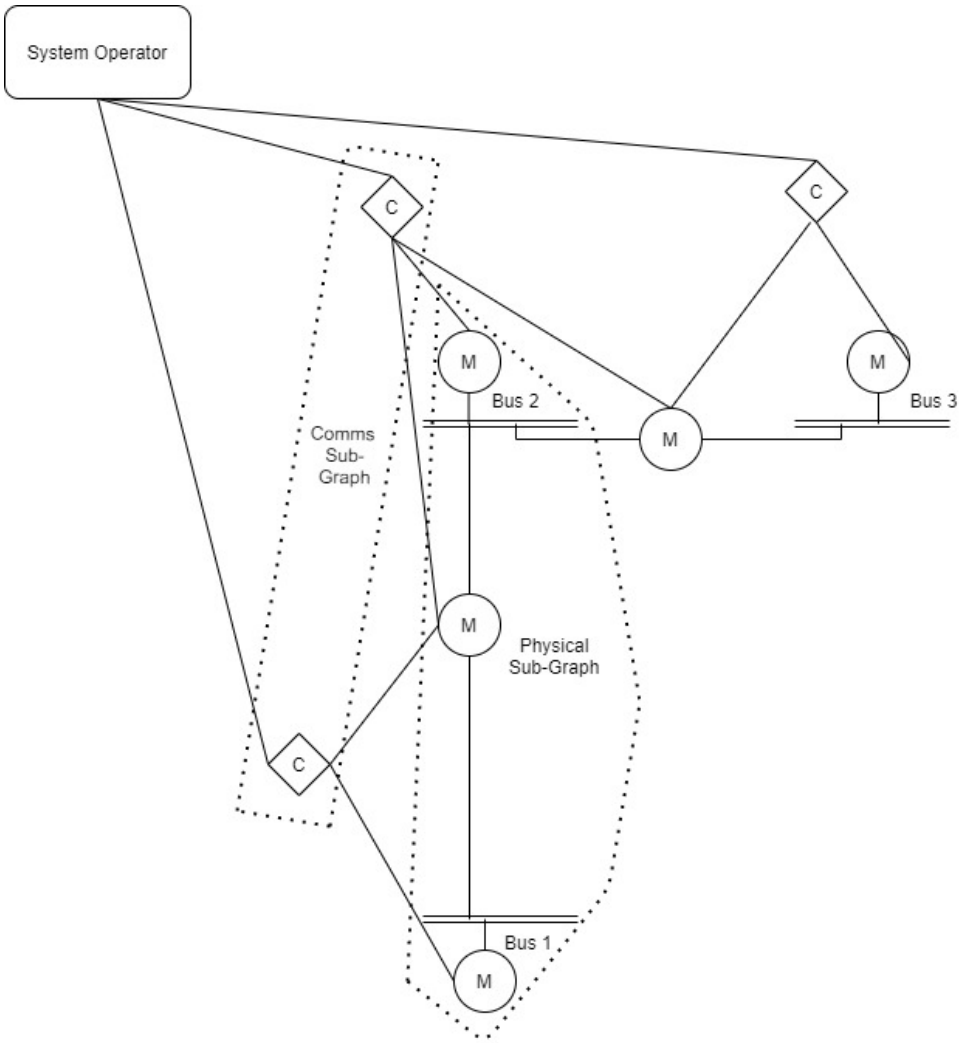

Figure 6.1: Cyber-physical network for 3-bus system with alternative cyber and physical capture attack strategies.

The weighted cost functionality also allows us to incorporate effects like the DoS-style attacks, as seen previously in [54].

6.2.3 State Capture Strategies

Meter Attack Criteria

For a given node/state wishing to be attacked under the physical system, the following is required to remain hidden:

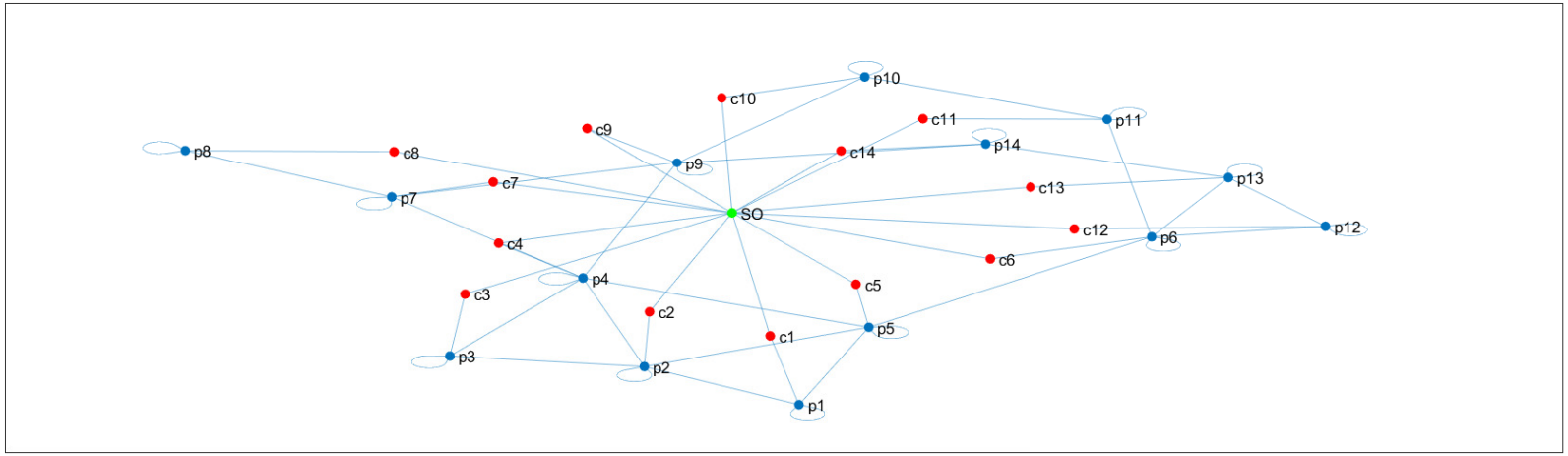

Figure 6.2: IEEE 14-bus cyber-physical graph representation with centralised system operator and communications nodes.

- The self-edge for the target node is captured, and the measurements are changeable.
- All edges emerging from the target node are captured.
- The neighbour nodes of the target node self-edges are also captured.

RTU Attack Criteria

For a given node/state wishing to be attacked under the communications network, the following is required to remain hidden:

- The attacker would need to capture the RTUs (equivalent to capturing the network bus and adjacent power flow meters) associated with the physical attack.
- The same physical sub-graph comprising of all the local measurements needs to be satisfied but through the capture of the upstream RTU, which (usually) holds multiple meter measurements.

Combined Attack Criteria

For a given node/state wishing to be attacked under the communication network, the following is required to remain hidden:

- The attacker would need to capture some combination of RTUs and individual meters to satisfy the original 'meter attack' criterion.

These attack options are shown in Figure 6.1 for a 3-bus system. The physical attack sub-graph requires capturing a number of system-level branches, whereas the communication strategy allows the capturing of just 2 upstream nodes. Alternatively, the attacker can opt for the combination attack, capturing one of the upstream RTU nodes and the leftover meter. 14-bus system representation is shown in Figure 9.1.

6.2.4 Physical Attack Risk

In the past, the risk of a successful FDI attack has been assessed in terms of the cost intrusion. However, this method of assessment fails to consider MTD and the current system state. To assess the ability to attack a system, a model should also consider the impact that active detection will have on the system residual in the presence of an attack. As we have shown previously in [4], deviations from expected values increase the chances of triggering anomaly detectors. We consider that the required level of MTD to protect a system is an important consideration, and this level of MTD will be dependent on the size of the attack vector **a**. Meanwhile, we must also consider that (usually) **c** will not be known ahead of the attack, so a true max-min optimisation based on the attack vector will not be possible. Therefore, from the defender perspective, it is better to base risk calculations on known quantities. We consider that to perform a branch overloading attack, the power flow profile of the system will have to be adjusted so that the power flow **z** exceeds the capacity overhead *co* . Therefore, the higher the **co**, the more tolerance for FDI attack an attacker has with respect to line overloading.

$$\mathbf{co} = \|\mathbf{z}_{cap}\| - \|\mathbf{z}\|. \tag{6.3}$$

Where S is an $m \times m$ diagonal containing the power flow sign of **z**. Using this, we can get a set of **c** values with the required branch change for overload. Setting all but the target branch in **co** to 0, we can use this capacity overhead to get the set of voltage angles required to overload the branch such that

$$\mathbf{z}_{ol} = \mathbf{z} + \mathbf{Hc}. \tag{6.4}$$

We constrain each case so that only a single bus is being attacked (consistent with the minimal possible meter selection problem) and thereby setting $\mathbf{c}^n = 0$ for all except the current target bus. The attack vector is then given by

$$\mathbf{a}_{ol}^{n,m} = \mathbf{Hc}. \tag{6.5}$$

Once the list of possible attack vectors has been evaluated for each bus, we use increasing magnitudes of available MTD in combination with a multi-variable optimisation of the topology $\mathbf{H}$ to return the maximum residual for the level of installed MTD capacity.

$$\begin{aligned} \max_{\mathbf{H}} \quad & \|(\mathbf{z}+\mathbf{a}_{ol}^{n,m}) - \mathbf{H}(\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}(\mathbf{z}+\mathbf{a}_{ol}^{n,m})\|_2 \\ \text{s.t.} \quad & \mathbf{H} < \mathbf{H}_{limit} \end{aligned} \tag{6.6}$$

In practice, this optimisation can be performed quickly with limited processing power. This is because the optimisation is highly constrained. For one, the optimisation of topology only occurs over the relatively small range of the D-FACTS system limits. In addition, only a few branches will contribute to the residual calculation (namely those affected by the attacking sub-graph). Because power systems are sparse, this means that even in large systems, only a small fraction of devices will need to be optimised for a given attacking sub-graph. As a result of these factors, the overall boundary of optimisation in practice is very small and can be done quickly. The level of MTD in % terms required for each attack vector will be used as an assessment factor, with regions requiring larger applications indicating an easier attack opportunity. We can use the WLS multiplier to find the relative level of divergence DIV such that

$$DIV = \|\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{H}_{mtd}\|. \tag{6.7}$$

We use this required level of divergence to denote regions of higher risk with respect to FDI. Areas with high levels of **CO** can be defended with almost no MTD applied, whereas regions operating at capacity can be overloaded with almost no FDI change and are therefore more difficult to protect using MTD.

6.3 Cyber-Physical Assessment Algorithm

6.3.1 Weighted Min Cost Communications

We use various tools from the MATLAB grTheory package [106] to establish the respective communications, physical and possible combination sub-graphs and then evaluate the respective weighting for all of them. Initially, the algorithm takes the power network and communications graphs as inputs. It identifies the underlying attacking sub-graphs. It then uses the MATLAB 'NCHOOSEK' function to outline the different possible combinations of RTUs or meters that can achieve this sub-graph, subsequently weighing each possible combination.

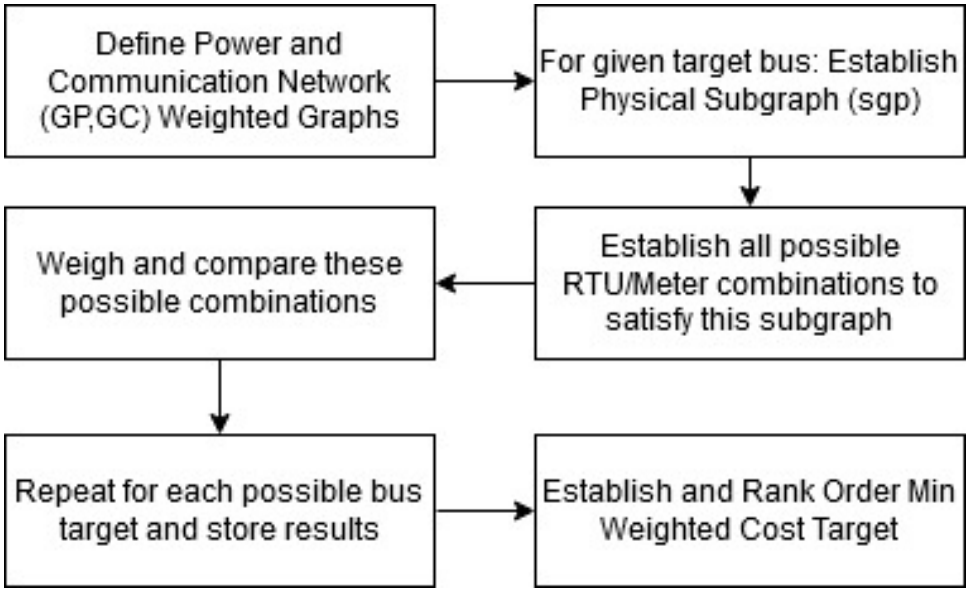

Figure 6.3: Outline of algorithm process for assessment of the weighted min meter cost.

This algorithmic flow of this process is illustrated in Figure 6.3. This algorithm can be applied quickly and simply. The time for varying system sizes is shown in 6.4 with (as expected) linear time scaling due to the relative simplicity of the operation. Because power systems are sparse, even large systems would have broadly linear scaling. The exception to this would be systems where the level of interconnection grows with the system size, such as in the case of 'complete' graphs.

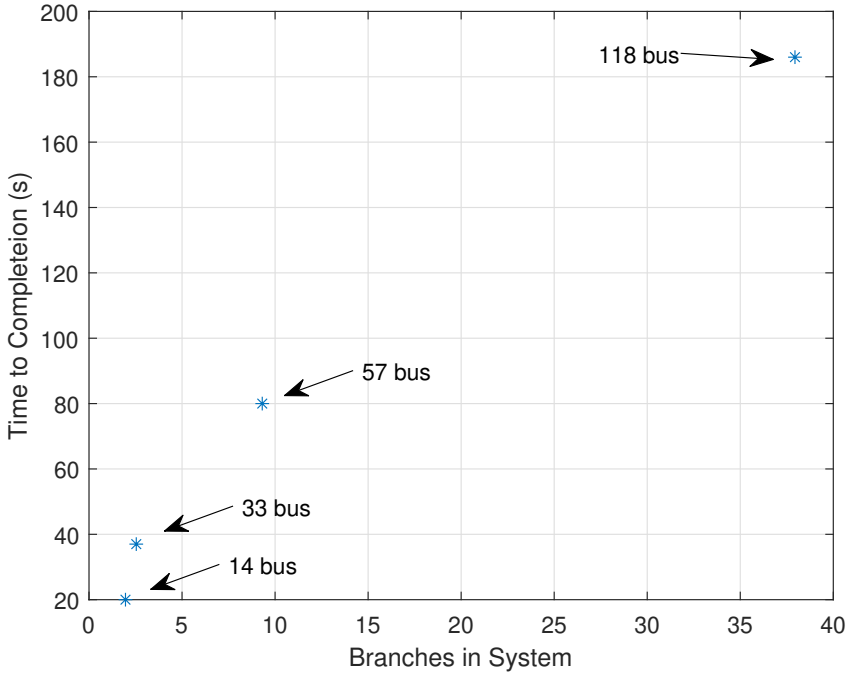

Figure 6.4: Time to completion of cyber-assessment algorithm process for systems of differing sizes.

6.3.2 MTD-Based Physical Vulnerability Algorithm

We use the MATLAB multi-variable optimisation package FMINCON to establish the maximum residual value for a given level of MTD capacity (ranging from 1-50% of base branch inductance). Initially, the algorithm takes the topology, MTD limits, power flows and power limits as inputs. Based on the power flow limits, min potential attack vectors are constructed from voltage angle adjustments for each branch. We then use the optimal MTD algorithm, with increasing capacity, to identify those regions that require the most overall MTD (as a % of the base) to be protected.

Computationally, this optimisation can run quickly with relatively low computational power. Further, even with large systems we don't anticipate non-linearity in computational cost. This is due to a few reasons

1. The optimisation is based on the linear model for power flow residual, meaning most of the internal (non-optimising) computation are simple matrix operations which have low lose and linear complexities.

2. For networks characterised by spare interconnection matrices (such as those seen in commonly

power systems) each potential attacking subgraph only accounts for measurements only within 1 vertices of difference. This means that in practice, even for very large systems the local MTD optimisations are only based around a few branches (typically less than 5).

3. The optimisation ranges are also highly contained and defined. Usually, we would consider the limits of MTD from D-FACTs style devices to be constrained well within 1 PU measurement (usually less). This means in practice the range of possible values are low and good solutions can be found quickly.

We outline the algorithmic flow of this process is illustrated in Figure 6.5. The algorithm takes a combination of topology and power flow data and produces optimal divergence profiles from which we take summed absolute values to produce some easy to interpret data sets.

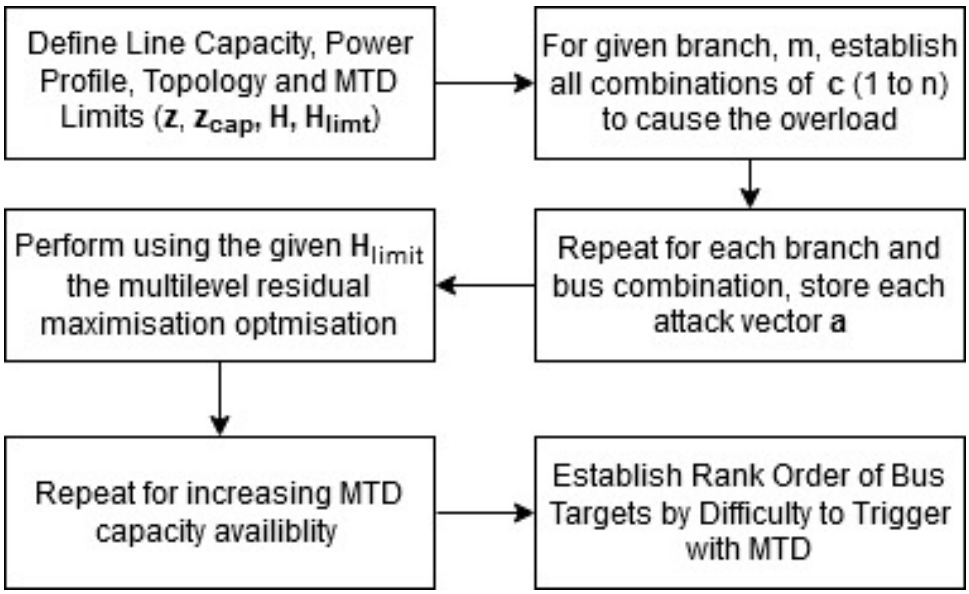

Figure 6.5: Outline of algorithm process for physical risk assessment using MTD divergence and line capacities.

6.3.3 Statistical Load Peak

We also consider the peak attack point for a respective attacker. In the past, previous papers have performed risk assessment assumptions with respect to FDI attacks under the assumption of normal or average operating conditions. In fact, from an attackers perspective this assumption is quite unlikely. More probably is that an attacker will attempt to capitalise on the interim system state to maximise

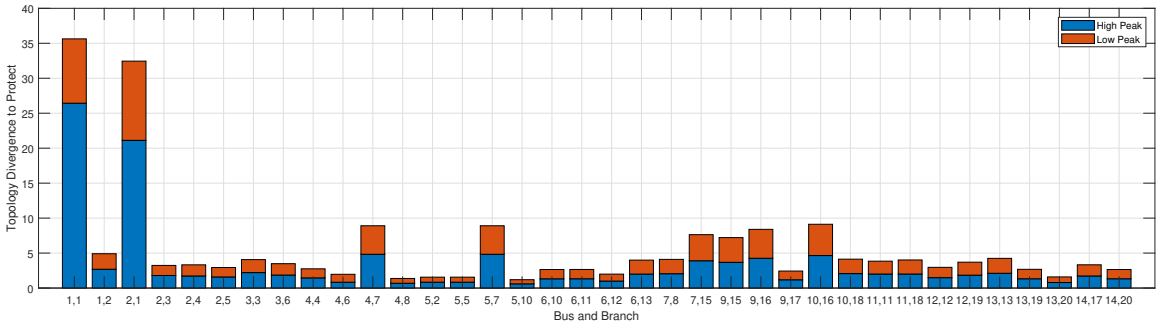

Figure 6.6: Absolute topology divergence to evaluate an attack for each bus and corresponding branch overload under a statistical peak of 3 standard deviations operating conditions high and the same boundary lower.

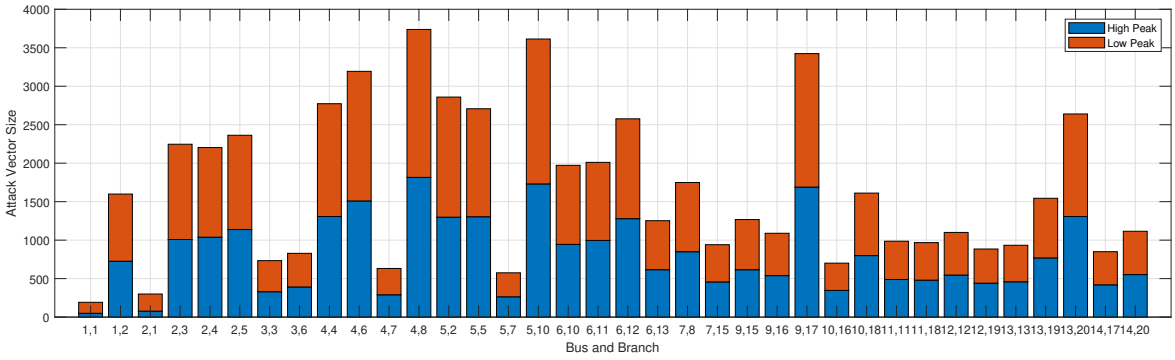

Figure 6.7: Absolute required relative attack vector size to overload a line an attack for each bus and corresponding branch overload under a statistical peak of 3 standard deviations operating conditions high and the same boundary lower.

the damage to the system or the chance of his remaining undetected. An analogous for circuit breaker control based attacks would be to wait for an extreme weather event before performing disconnections in order to mask the underlying attack. These issues become increasingly problematic for the FDI attack when one considers that often, attackers can remain hidden for many months when intruding a system. During the 2017 Ukraine attacks the attackers had an intrusion time of about 3 months before they begun to actually launch their attack [1]. It makes sense, to attempt to incorporate this kind of worse case system modelling into our risk model. Therefore, we introduce assumptions on the system under attack. We consider that the attacker will wait for some kind of (relatively) opportune moment before launching his attack. Given the constraints excess network capacity has on the attackers overloading style FDI attack. We represent this opportune moment as a statistically significant load of 3 standard deviations from the mean such that

$$\mathbf{z}_s = \bar{\mathbf{z}} + 3\mathbf{SD}_z. \tag{6.8}$$

Where **SD** is a vector of standard deviations of z branch and bus values. This allows us to model an attacker under a high but plausible level of load pressure on the system. Larger system capacity, relative to the branch flow means the attacker requires a larger absolute attack vector in order to successfully overload the line. Consequently, these vectors become easier to evaluate by the system operator using MTD. Therefore, by framing our simulations around this point we represent the mostly likely temporal attacking point.

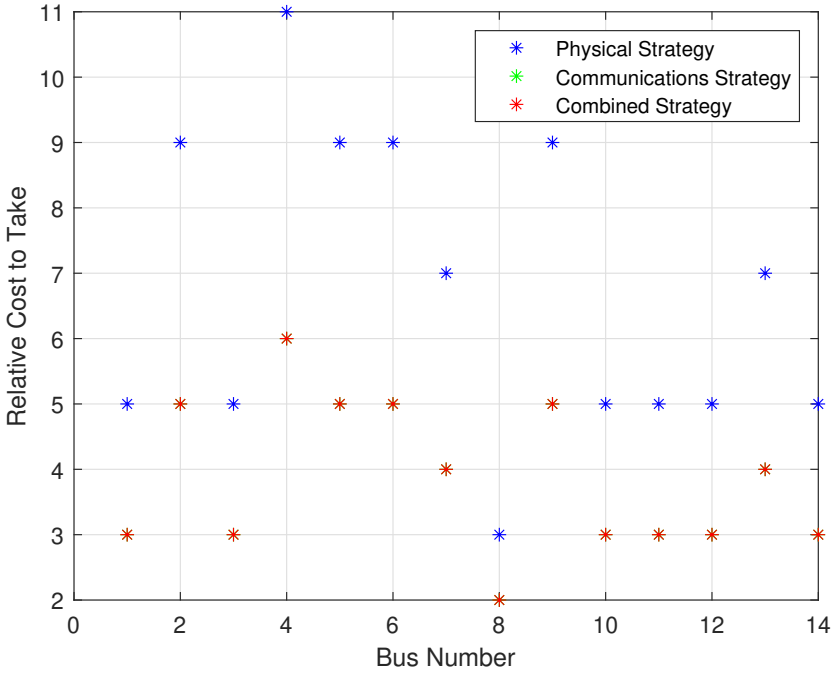

Figure 6.8: Weighted cost of each strategy with Node meters, Branch meters and RTUs equal to a weighted cost of 1

6.4 Results & Analysis

This section shows the results of the proposed risk assessment strategies on both the standard IEEE 14-bus test system [84]. All grid simulations were implemented using the MATPOWER toolbox in MATLAB [72] and performed using Intel Core i7-7820X CPU with 64 GB of RAM running on a

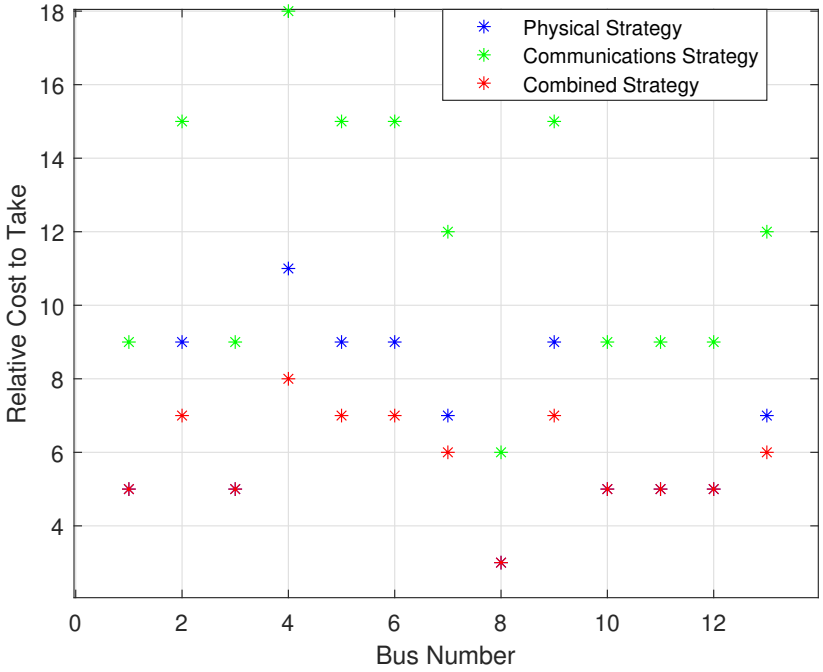

Figure 6.9: Weighted cost of each strategy with Node meters, Branch meters equal to 1 and RTUs equal to a weighted cost of 3

Windows 10 system.

6.4.1 IEEE-14 Bus System Cyber

In Figure 6.8, we show the min target costs under the assumption of a flat cost of 1 for both meters or RTUs. From this graph, it is clear that the communications-only strategy is always the most efficient under the assumption that the devices are of equal difficulty to capture. The reason that this effect occurs is because the RTUs sit upstream of meter measurements and thus have control of downstream meter measurements, i.e. each RTU effectively has equivalence in capture to multiple meter measurements. Therefore, an attacker can replicate attacking sub-graphs by attacking fewer of these upstream nodes rather than the meters directly. This is because this first example is an unweighted, non-reinforced model. In reality, most SOs would be aware that the RTUs would usually represent a better target to the attacker due to this relationship (even prior to risk modelling), so we accept that this specific scenario is unlikely in a practical system. RTUs will probably have embedded defences against intrusions in place, which will likely mean that the RTUs will have higher levels of intrusion

protection than base meter measurements. We illustrate this in Figure 6.9. In this graph, we introduce a weighted cost for RTU capture of factor 3 times the direct meter compromise. In practice, weighted reinforcement at the RTU level is a more realistic assumption and makes sense from a system operation perspective. RTUs control multiple functions and have downstream capabilities. The value in their protection will be more crucial than simple meter measurements, which only provide telemetered measurements. Crucially, the weighting of RTUs in this manner shows the emergence of combined strategies emerging as the most efficient use of resourcing. This makes sense because these combined approaches allow an attacker to utilise well-connected RTUs and isolated meters to complete underlying attack sub-graphs.

6.4.2 IEEE-14 Bus System Physical

In Figure 6.6, we explore the impact of increasing MTD divergence on system residual and detection. As discussed previously, we use the optimised min attack vector required to overload a line within each attacking sub-graph. We note that in branches where the natural power flow profile is close to the network capacity, only smaller attacking vectors are needed to achieve the simulated overload. This makes sense from a system operation perspective because areas close to limits require only small changes to overload a branch in excess of capacity limits. We should expect to see that branches with higher peak flows have higher MTD divergence requirements (because the attack vectors are smaller and harder to evaluate with MTD). Indeed, we see that for regions with high overload capacity (low branch power flow relative to capacity), only a minor application of MTD-based divergence is required to evaluate the optimal branch overloading attack. However, as we can see in Figure 6.7, there is an inverse relationship between the size of the attacking vector and the level of MTD required to protect against the attack. This is particularly clear when the high vs. low peak results are observed. For example, bus 1-2 operates at the closest point to the branch capacity, and we see that large levels of MTD divergence are required to protect this bus sufficiently from cyberattacks. Significantly higher levels of overall divergence are required to defend the system, which means that these points are comparatively susceptible to FDI-based changes. As the average size of attack vector to breach the system becomes lower, the level of divergence required to evaluate a FDI attack increases. This makes

buses with these close to overloaded branches relatively better targets than other regions where the attack vector has to be large (and hence more easily evaluated).

6.4.3 IEEE-14 Bus System Cyber-Physical

We now consider attack targets in terms of their whole system risk. Bus 8 is by far the most vulnerable target in every non-reinforced model. Its relatively low interconnectivity means it has the lowest capture cost of all the available buses, with just 2 components needed to compromise and gain a stealthy intrusion. Also, the lack of interconnection also means that MTD protocols are ineffective because MTD requires at least 2 interconnections within an attacking sub-graph to drive changes to the residual under attack. This means that despite the low line power flows, MTD is ineffective. In light of this consideration, bus 8 should be the priority busbar for intrusion-based reinforcement, and physical reinforcement should be ignored and will provide no benefits to this busbar. This particular case has crucial implications for systems such as the IEEE 33-bus distribution-style network due to their lower levels of interconnectivity resulting from the tree-style topology. The vast majority of attack points in these types of systems would gain no defensive advantage with MTD, so intrusion-based defences should be prioritized. Similarly, in the case of busbar points 1 & 2, there are large defensive MTD requirements to protect the branch 1-2 measurement. Although this branch is defensible with MTD from an absolute cost perspective, it is likely better to consider enhancing intrusion protection. Predictably, highly connected have some innate protection when it comes to FDI-style attacks. Busbar 6, for example, has 4 interconnections, which means an innately higher system protection from an intrusion perspective, with 9 underlying meters needing to be captured to commit stealthy changes.

6.5 Summary

In this chapter, we developed a cyber-physical risk assessment framework for FDI attacks. Our assessment criteria combines weighted graph assessment of the cyber vulnerabilities in combination with a residual-based assessment of the physical system with relation to MTD. This framework provides, for the first time, an intrusion and change introduction model for risk assessment. This model

first considers the weighted min cost of intrusion into the network sub-graph by both RTU, meter and combined means. This initial model provides an effective measure for the safeness of a specific state of a given measurement to stealthily change via an FDI attack. It can also be weighted to apply reasonable reinforcement reflections for different measurement types, such as RTUs. The model also evaluates overlapping-style attacks i.e. utilising different measurement layers to attack. This allows for a better representation of the true risk of compromising a state.

This work is also a cyber-physical analysis of the vulnerabilities to FDI attack. Our model goes onto to look at the risk of change from a FDI attack given the underlying system state. We consider how likely an attack is to succeed, specifically in the realm of overloading style attacks against systems with MTD in place. Initially, the risk assessment model takes what we consider the statistical load profile. This is to tacitly incorporate the attackers willingness to wait for an opportune moment to attack. This statistical load profile is a 3SD change from the system average load. These higher loads represent lower challenge thresholds for the attacker. The optimal, min change, overloading attack vector for a given line is then calculated. This is used to elucidate the most stealthy vector, for a given line, which can cause branch overloading. Around these optimal overloading attack vectors we structure optimised MTD defence. These allow us to model for which busbars require the highest level of divergence in order to successfully defend them from FDI change. This is done by modelling the residuals under the min overloading attack in the presence of MTD.

This work principally evaluates how defensible certain targets are with certain approaches and when to use certain types of defences. The model is all encompassing in the sense that it incorporates both intrusion and prospect for change into the system model.

6.6 Lessons Learnt

Crucially, we draw the following lessons from this chapter, which we use to inform our research and proceed to the next chapter in this work.

- The application of defensive reinforcement to the power system must be made in terms of a

whole system or (at least) a cyber-physical approach, and reinforcement must make considerations of the system topology.

- If we consider the potential for damage done from line overloading, the physical (post-intrusion) risk of an FDI attack is related to the branch overloading capacity and how large an attack vector is needed to achieve an outcome. Large attacks are easier to expose with MTD.
- Isolated busbars (similar to those seen in a distribution network) represent the highest risk in terms of both the cyber and physical risk factors. Their lack of interconnection means they cannot be defended with MTD, and they represent an inherently lower cost of capture in unweighted models of power system capture risk.

In the next chapter, we provide conclusions and comments that cover the whole body work and consider future areas for research.

Chapter 7

Conclusion

This chapter concludes the thesis, outlines the achievements herein and provides suggestions for system operators based on the lessons learnt throughout.

7.1 Summary of Thesis Achievements

This thesis proposes novel methods for improving both the attack and defence against FDI attacks against power system infrastructure. Our achievements with respect to each chapter are listed below.

7.1.1 Topology-Learning-Aided False Data Injection Attack without Prior Topology Information

This chapter is based on the work done in [64], wherein we present a topology-learning-aided FDI attack capable of attacking power systems under a blind assumption model (no branch or network incidence information available). The attack is committed against the AC power system and uses the latest state-of-the-art topology discovery techniques to build a model for the network. We introduce an attacker-side criteria assessment via a pseudo-residual calculation to allow probabilistic assessment of attack success before any attack committed, allowing the attacker to ensure stealthiness. We also

show that regional pseudo residuals can be used to verify local attacks, even in the presence of global topology errors. We demonstrate how quickly the attacker can develop the full knowledge of system topology and parameters and effectively invalidate the full system knowledge assumptions in previous studies.

7.1.2 Stealthy MTD against Unsupervised Learning-Based FDIAs

This chapter is based on the work done in [38]. The work is a mixed front paper. We explore both attack and defence against FDI attacks. This is one of the first papers in the literature to consider the prospect of counter MTD with respect to FDI attacks. Where previous FDI attacks have been designed against static systems, we seek to offer new attacking considerations in the presence of dynamic systems with MTD. The proposed intelligent attack operates under zero system knowledge assumption. It combines dimensionality reduction and unsupervised learning to identify the underlying clusters associated with network topology and design the corresponding attack vector. The method is shown to be effective and stealthy against traditional MTD. This initial push raises obvious concerns about the vulnerability of current MTD applications. Consequently, we also introduce a new implementation of MTD to drive detection against traditional and intelligent FDI attacks. The proposed defence strategy combines MTD and the physical watermarking concept for the first time by adding a Gaussian watermark into physical plant parameters. Because the added watermark mimics the underlying noise of the system, the physical changes driven by MTD stay hidden. The physical watermarking is combined with cumulative error monitoring to spot minor but sustained changes in the system.

7.1.3 Enhanced Cyber-Physical Security Using Attack-Resistant Cyber Nodes and Event-Triggered Moving Target Defence

This chapter is based on the work outlined in [4]. The enhancements in this chapter come purely on the MTD and power system defence side. In this chapter, we implement an anomaly detection scheme at the distributed level using Holt-Winters seasonal forecasting. The seasonality element of the Holt-Winters forecast captures the intra-day demand differences to minimise the overall window

for attack. The importance of this technique comes in the cost and stability implications of applying MTD in the power system. This work has also been targeted for commercialisation in collaboration with Royal-Holloway University via the Cyber Security Academic Startup Accelerator Programme with a prototype currently being research and expected early 2022. Additionally, we attempt to consider a more distributed approach to the use of MTD and propose a protection protocol based on event-triggered MTD. The protocol consists of an initial anomaly detection followed by MTD and traditional CSE.

7.1.4 Multi-Layered Risk Assessment for FDI Attacks in the Presence of Moving Target Defence

In this final chapter, we developed a risk assessment framework for false data injection attacks. Our assessment criteria combines weighted graph assessment of the cyber vulnerabilities in combination with a residual-based assessment of the physical system with relation to MTD. This framework provides, for the first time, an intrusion and change introduction model for risk assessment. This model first considers the weighted min cost of intrusion into the network sub-graph by both RTU, meter and combined means. This initial model provides an effective measure for the risk of cyber-intrusion for a given busbar, defined by the surrounding meters and upstream RTUs. It can also be weighted to apply reasonable reinforcement reflections for different measurement types, such as RTUs. We also evaluate overlapping-style attacks. This is important from a cyber-physical perspective. In the past, we have see assessment performed in a physical-only-style approach. Meaning the capture of metering sub-graph for a given FDI attack. This work incorporates overlapping-style attacks, which considers upstream vulnerabilities. In addition to this intrusion risk assessment or cyber-style risk. Our model goes onto to look at the risk of change from a FDI attack given the underlying system state. We consider how likely an attack is to succeed, specifically in the realm of overloading style attacks against systems with MTD in place. Initially, the risk assessment model takes what we consider the statistical load profile. This is to tacitly incorporate the attackers willingness to wait for an opportune moment to attack. This statistical load profile is a 3SD change from the system average load. These higher loads represent lower challenge thresholds for the attacker. The optimal minimum change overloading

attack vector for a given line is then calculated. This is used to elucidate the most stealthy vector, for a given line, which can cause branch overloading. Around these optimal overloading attack vectors we structure optimised MTD defence. These allow us to model for which busbars require the highest level of divergence in order to successfully defend them from FDI change. This is done by modelling the residuals under the min overloading attack in the presence of MTD.

7.2 Suggestions for System Operators

Here we outline some suggestions for how system operators can improve their system resilience from the lessons learnt from this thesis herein.

7.2.1 Topology-Learning-Aided False Data Injection Attack without Prior Topology Information

It is possible to perform a a topology-learning-aided FDI attack capable of attacking power systems under a blind assumption model (i.e. no branch or network incidence information available). System operators should consider their system topology common knowledge and not expect hidden knowledge of the system topology to be a suitable defence against cyber-attack. Attackers can use their generated models to assess attack strength and model whether their attack is likely to succeed and whether it can be done relatively quickly. Attacks can be committed in a matter of hours post-intrusion to the network system and with limited prior information. By comparison, attacks like those in Ukraine were done over a period of months, so it is very likely that a motivated attacker will have sufficient time to collect the data needed for an attack such as these.

7.2.2 Stealthy MTD against Unsupervised Learning-Based FDIAs

It is probable that a motivated attacker will be able to detect and circumvent naive applications of MTD, i.e. those that use a non-continuous waveform mode of direct perturbation. System operators

should be considerate in how they apply MTD and should attempt to replicate natural wave forms (such as noise patterns) as a method of camouflaging MTD. System operators should consider providing additional support to CSE-style residual detection systems via the implementation of cumulative error-style monitoring. Small, sustained errors caused by small attack vectors can easily bypass BDD detection. Therefore, it makes sense to incorporate cumulative monitoring that will provide additional metrics for assessing the attacker's strength.

7.2.3 Enhanced Cyber-Physical Security Using Attack-Resistant Cyber Nodes and Event-Triggered Moving Target Defence

Because MTD is costly, as shown in [40], the operator will likely want to consider how he can minimise his overall application. We suggest an anomaly-driven MTD protocol known as 'event-triggered' MTD. This can reduce the overall application of MTD in the power system to only those times when an attacker may be present. They provide an adequate and effective alternative to time-based or continuous application approaches that have the potential to miss in-cycle FDI attacks. We have also shown that it is also possible, via Holt-Winters-style forecasting, to provide a distributed trigger that does not need to be centrally checked. This is advantageous because it can provide low-level protection without increasing communications bandwidth.

7.2.4 Multi-Layered Risk Assessment for FDI Attacks in the Presence of Moving Target Defence

When assessing risk, the application of defensive reinforcement to the power system must be made in terms of a whole system or (at least) a cyber-physical approach, and reinforcement must make considerations of the system topology. We should consider this in terms of the potential for damage done from line overloading. The physical (post-intrusion) risk of an FDI attack is related to the branch overloading capacity and how large an attack vector is needed to achieve an outcome. Large attacks are easier to expose with MTD, and MTD-based defences are most appropriate in areas with high capacity overhead (because the attack vectors will be larger). Isolated busbars (similar to those

seen in a distribution network) represent the highest risk in terms of both cyber and physical risk factors. Their lack of interconnection means they cannot be defended with MTD, and they represent an inherently lower cost of capture in unweighted models of power system capture risk. We advise a strategy of intrusion defence for these types of points to ensure they are adequately insured from FDI attacks.

7.3 Future Work

In this session, we summarise some potential future research avenues for the different chapters in this thesis.

7.3.1 Topology-Learning-Aided False Data Injection Attack without Prior Topology Information

In this chapter, we examined how state-of-the-art topology discovery could be used to attack static power systems. However, the method is applied under the assumption of a system with access to all system infrastructure and principally lacking network topology information. Future work should seek to find ways to reduce the overall measurement requirement. As we showed in later chapters, capturing these resources is costly, and therefore, it would serve an attack to reduce this cost by reducing the capture requirement needed for a successful topology inference.

7.3.2 Stealthy MTD against Unsupervised Learning-Based FDIAs

In this chapter, we suggest a clustering-based method of countering basic MTD implementations and a way of camouflaging MTD implementations. However, the solution for countering MTD is based on a fairly naive implementation of MTD. A solution is still needed to counter some of the more complex 'hidden' MTD-style solutions, such as those seen in [48]. Further, the suggested continuous application of MTD is likely to be costly and is likely unnecessary for some portions of the grid.

Research into targeted defence protocols that incorporate some of the work from our later chapter on risk would likely result in more cost-efficient applications.

7.3.3 Enhanced Cyber-Physical Security Using Attack-Resistant Cyber Nodes and Event-Triggered Moving Target Defence

In this chapter, we examined how event triggers can be used in combination with MTD to evaluate FDI while reducing the overall application. However, although the distributed anomaly detection is effective at highlighting manipulated measurements, it will not be effective in practice with plausible vectors within expected variance. We anticipate replay-style attacks to be particularly difficult to detect. Future work, could look at how to enhance these event triggers for more plausible replay vectors. It could also look at how to estimate the specific attack vector rather than simply the anomaly in place. This would allow for a more optimised MTD and reduce overall implementation.

7.3.4 Multi-Layered Risk Assessment for FDI Attacks in the Presence of Moving Target Defence

In this chapter, we examined how to assess cyber-physical risk within the context of FDI-style attacks. We believe that additional research on whole system risk is needed. Consideration of overlapping-style risks in power and whole system modelling has been limited, and we believe the current literature on risk assessment needs to be extended to cover these overlapping and whole-system risks.

7.3.5 Other Areas for Future Work

Whole-System/Combined FDI attack Styles

While previous works have focused on single system FDI attacks, combined style attacks which utilise both gas/power networks could provide additional challenges for system operators. The ability to attack and manipulate states in both domains could exacerbate outages and reduce the attacker cost

requirements for state capture. However, they could also provide additional opportunities for defence whereby a whole system residual could be used to enhance protection of both the power and gas networks. Further work in this area could look at the development of a whole system residual MTD approach to further enhance system security across networks. It should also be considered, that while these attacks are academically interesting, from a practical perspective they would be quite difficult to initiate. Further work on attacks that build on these idea but more practical application should be of interest.

Post-Attack Rectification

As we discussed earlier, one of the most under-served areas with respect FDI attacks is in the field of post attack discovery. It is also, arguably, the most crucial. In the case of the Ukraine attacks, re-energising the system as possible in a few hours but post attack visibility was still poor for months afterwards. Tools to ensure visibility post-attack either purely data driven or partially MTD assisted would greatly enhance system resilience post-successful cyber attack.

7.3.6 To Conclude

This dissertation has studied operational advancements to both FDI-style attacks and MTD-style defences. We have shown throughout how it is necessary to consider enhancements to both fields in enhancing the capabilities of system operators. Although the initial focus was on practical enhancements to FDI and operational applications of MTD, we have also extended this thesis to the risk assessment of FDI attacks within the context of MTD and the current system state.

Chapter 8

Bibliography

Bibliography

[1] Gaoqi Liang, Steven R. Weller, Junhua Zhao, Fengji Luo, and Zhao Yang Dong. The 2015 Ukraine Blackout: Implications for False Data Injection Attacks. *IEEE Transactions on Power Systems*, 2017.

[2] Laura McCamy. 7 things you can hire a hacker to do and how much it will (generally) cost, 6 2021.

[3] Yao Liu, Peng Ning, and Michael K. Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, 2011.

[4] Martin Higgins, Keith Mayes, and Fei Teng. Enhanced Cyber-Physical Security Using Attack-resistant Cyber Nodes and Event-triggered Moving Target Defence. 10 2020.

[5] A Monticelli. *State Estimation in Electric Power Systems: A Generalized approach*, volume 1. Springer, 1st edition, 5 1999.

[6] Felix F. Wu. Power system state estimation: a survey. *International Journal of Electrical Power & Energy Systems*, 12(2), 4 1990.

[7] Yi Tan, Yong Li, Yijia Cao, and Mohammad Shahidehpour. Cyber-attack on overloading multiple lines: A bilevel mixed-integer linear programming model. *IEEE Transactions on Smart Grid*, 9(2):1534–1536, 2018.

[8] Hwei Ming Chung, Wen Tai Li, Chau Yuen, Wei Ho Chung, Yan Zhang, and Chao Kai Wen. Local Cyber-Physical Attack for Masking Line Outage and Topology Attack in Smart Grid. *IEEE Transactions on Smart Grid*, 8 2018.

[9] Xuan Liu, Zhiyi Li, Xingdong Liu, and Zuyi Li. Masking Transmission Line Outages via False Data Injection Attacks. *IEEE Transactions on Information Forensics and Security*, 11(7):1592–1602, 7 2016.

[10] Javad Khazaei. Cyberattacks with limited network information leading to transmission line overflow in cyber–physical power systems. *Sustainable Energy, Grids and Networks*, 27, 9 2021.

[11] Javad Khazaei and M. Hadi Amini. Protection of large-scale smart grids against false data injection cyberattacks leading to blackouts. *International Journal of Critical Infrastructure Protection*, 35:100457, 12 2021.

[12] Sara Mohammadi, Frank Eliassen, Yan Zhang, and Hans Arno Jacobsen. Detecting false data injection attacks in peer to peer energy trading using machine learning. *IEEE Transactions on Dependable and Secure Computing*, 2021.

[13] Gaoqi Liang, Junhua Zhao, Fengji Luo, Steven R. Weller, and Zhao Yang Dong. A Review of False Data Injection Attacks Against Modern Power Systems. *IEEE Transactions on Smart Grid*, 2017.

[14] Md Ashfaqur Rahman and Hamed Mohsenian-Rad. False data injection attacks with incomplete information against smart power grids. In *GLOBECOM - IEEE Global Telecommunications Conference*, 2012.

[15] Mohammad Esmalifalak, Huy Nguyen, Rong Zheng, and Zhu Han. Stealth false data injection using independent component analysis in smart grid. In *2011 IEEE International Conference on Smart Grid Communications, SmartGridComm 2011*, 2011.

[16] Ruilong Deng and Hao Liang. False Data Injection Attacks with Limited Susceptance Information and New Countermeasures in Smart Grid. *IEEE Transactions on Industrial Informatics*, 15(3):1619–1628, 3 2019.

[17] Mehmet Necip Kurt, Yasin Yilmaz, and Xiaodong Wang. Real-Time Detection of Hybrid and Stealthy Cyber-Attacks in Smart Grid. *IEEE Transactions on Information Forensics and Security*, 14(2):498–513, 2 2018.

[18] Runhai Jiao, Gangyi Xun, Xuan Liu, and Guangwei Yan. A New AC False Data Injection Attack Method without Network Information. *IEEE Transactions on Smart Grid*, pages 1–1, 2021.

[19] Yi Wang, Mahmoud M. Amin, Jian Fu, and Heba B. Moussa. A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids. *IEEE Access*, 5:26022–26033, 11 2017.

[20] Saeed Ahmed, Youngdoo Lee, Seung Ho Hyun, and Insoo Koo. Unsupervised Machine Learning-Based Detection of Covert Data Integrity Assault in Smart Grid Networks Utilizing Isolation Forest. *IEEE Transactions on Information Forensics and Security*, 14(10):2765–2777, 10 2019.

[21] Keke Huang, Zili Xiang, Wenfeng Deng, Chunhua Yang, and Zhen Wang. False Data Injection Attacks Detection in Smart Grid: a Structural Sparse Matrix Separation Method. *IEEE Transactions on Network Science and Engineering*, pages 1–1, 7 2021.

[22] Xuefei Yin, Yanming Zhu, and Jiankun Hu. A Sub-grid-oriented Privacy-Preserving Microservice Framework based on Deep Neural Network for False Data Injection Attack Detection in Smart Grids. *IEEE Transactions on Industrial Informatics*, pages 1–1, 2021.

[23] Jorge Valenzuela, Jianhui Wang, and Nancy Bissinger. Real-Time Intrusion Detection in Power System Operations. *IEEE Transactions on Power Systems*, 28(2):1052–1062, 2013.

[24] Yuancheng Li, Yuanyuan Wang, and Shiyan Hu. Online Generative Adversary Network Based Measurement Recovery in False Data Injection Attacks: A Cyber-Physical Approach. *IEEE Transactions on Industrial Informatics*, 16(3):2031–2043, 3 2020.

[25] Fuxi Wen and Wei Liu. An Efficient Data-Driven False Data Injection Attack in Smart Grids. In *International Conference on Digital Signal Processing, DSP*, volume 2018-November. Institute of Electrical and Electronics Engineers Inc., 1 2019.

[26] Jinsub Kim, Lang Tong, and Robert J. Thomas. Subspace methods for data attack on state estimation: A data driven approach. *IEEE Transactions on Signal Processing*, 63(5):1102–1114, 3 2015.

[27] Jinping Hao, Robert J. Piechocki, Dritan Kaleshi, Woon Hau Chin, and Zhong Fan. Sparse Malicious False Data Injection Attacks and Defense Mechanisms in Smart Grids. *IEEE Transactions on Industrial Informatics*, 11(5):1198–1209, 10 2015.

[28] Jiazi Zhang, Zhigang Chu, Lalitha Sankar, and Oliver Kosut. Can attackers with limited information exploit historical data to mount successful false data injection attacks on power systems? *IEEE Transactions on Power Systems*, 2018.

[29] Meng Tian, Zhengcheng Dong, and Xianpei Wang. Analysis of false data injection attacks in power systems: A dynamic Bayesian game-theoretic approach. *ISA Transactions*, 115:108–123, 9 2021.

[30] T. S. Sreeram and S. Krishna. Protection against false data injection attacks considering degrees of freedom in attack vectors. *IEEE Transactions on Smart Grid*, pages 1–1, 6 2021.

[31] Qingyu Yang, Liguo Chang, and Wei Yu. On false data injection attacks against Kalman filtering in power system dynamic state estimation. *Security and Communication Networks*, 9(9):833–849, 6 2016.

[32] Hossam A. Gabbar and V Dinavahi. On False Data Injection Attack against DynamicState Estimationon Smart Power Grids. In *5th IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, Canada, 2017.

[33] Hadis Karimipour and Venkata Dinavahi. Robust Massively Parallel Dynamic State Estimation of Power Systems Against Cyber-Attack. *IEEE Access*, 6:2984–2995, 12 2017.

[34] Mohsen Jorjani, Hossein Seifi, Ali Yazdian Varjani, and Hamed Delkhosh. An Optimization-Based Approach to Recover the Detected Attacked Grid Variables after False Data Injection Attack. *IEEE Transactions on Smart Grid*, pages 1–1, 2021.

[35] Ruilong Deng, Gaoxi Xiao, Rongxing Lu, Hao Liang, and Athanasios V. Vasilakos. False data injection on state estimation in power systems-attacks, impacts, and defense: A survey. *IEEE Transactions on Industrial Informatics*, 2017.

[36] Shaocheng Wang, Wei Ren, and Ubaid M. Al-Saggaf. Effects of Switching Network Topologies on Stealthy False Data Injection Attacks Against State Estimation in Power Networks. *IEEE Systems Journal*, 11(4):2640–2651, 11 2015.

[37] Kate L Morrow, Erich Heine, Katherine M Rogers, Rakesh B Bobba, and Thomas J Overbye. Topology Perturbation for Detecting Malicious Data Injection. In *2012 45th Hawaii International Conference on System Sciences*, 2012.

[38] Martin Higgins, Fei Teng, and Thomas Parisini. Stealthy MTD Against Unsupervised Learning-based Blind FDI Attacks in Power Systems. *IEEE Transactions on Information Forensics and Security*, 4 2020.

[39] Chensheng Liu, Min Zhou, Jing Wu, Chengnian Long, Abdallah Farraj, Eman Hammad, and Deepa Kundur. Reactance Perturbation for Enhancing Detection of FDI Attacks in Power System State Estimation. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017.

[40] Subhash Lakshminarayana and David K.Y. Yau. Cost-Benefit analysis of Moving-Target defense in power grids. In *Proceedings - 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2018*, pages 139–150. Institute of Electrical and Electronics Engineers Inc., 7 2018.

[41] Subhash Lakshminarayana, E. Veronica Belmega, and H. Vincent Poor. Moving-Target Defense Against Cyber-Physical Attacks in Power Grids via Game Theory. *IEEE Transactions on Smart Grid*, pages 1–1, 2021.

[42] Zhenyong Zhang, Ruilong Deng, David Yau, Peng Cheng, and Jiming Chen. Analysis of Moving Target Defense Against False Data Injection Attacks on Power Grid. *IEEE Transactions on Information Forensics and Security*, 2019.

[43] Beibei Li, Gaoxi Xiao, Rongxing Lu, Ruilong Deng, and Haiyong Bao. On Feasibility and Limitations of Detecting False Data Injection Attacks on Power Grid State Estimation Using D-FACTS Devices. *IEEE Transactions on Industrial Informatics*, pages 1–1, 6 2019.

[44] Zhenyong Zhang, Ruilong Deng, Peng Cheng, and Mo-Yuen Chow. Strategic Protection against FDI Attacks with Moving Target Defense in Power Grids. *IEEE Transactions on Control of Network Systems*, pages 1–1, 7 2021.

[45] Bo Liu and Hongyu Wu. Systematic planning of moving target defence for maximising detection effectiveness against false data injection attacks in smart grid. *IET Cyber-Physical Systems: Theory and Applications*, 2021.

[46] Mengxiang Liu, Chengcheng Zhao, Zhenyong Zhang, Ruilong Deng, and Peng Cheng. Analysis of Moving Target Defense in Unbalanced and Multiphase Distribution Systems Considering Voltage Stability. Technical report.

[47] Tua A Tamba, Bin Hu, and Yul Y Nazaruddintld. *An Actuator Intrusion Detection Mechanism for Event-Triggered Moving Target Defense Control*. 2019.

[48] Jue Tian, Rui Tan, Xiaohong Guan, and Ting Liu. Enhanced hidden moving target defense in smart grids. *IEEE Transactions on Smart Grid*, 10(2):2208–2223, 3 2019.

[49] Omur Ozel, Sean Weerakkody, and Bruno Sinopoli. Physical watermarking for securing cyber physical systems via packet drop injections. In *2017 IEEE International Conference on Smart Grid Communications, SmartGridComm 2017*, 2018.

[50] Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 2014.

[51] Gabriela Hug and Joseph Andrew Giampapa. Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks. *IEEE Transactions on Smart Grid*, 2012.

[52] Henrik Sandberg, André Teixeira, and Karl H Johansson. On Security Indices for State Estimators in Power Networks. Technical report.

[53] Andre Teixeira, Kin Cheong Sou, Henrik Sandberg, and Karl Henrik Johansson. Secure control systems: A quantitative risk management approach. *IEEE Control Systems*, 35(1):24–45, 2 2015.

[54] Kaikai Pan, Andre Teixeira, Milos Cvetkovic, and Peter Palensky. Cyber Risk Analysis of Combined Data Attacks Against Power System State Estimation. *IEEE Transactions on Smart Grid*, 10(3):3044–3056, 5 2019.

[55] Martín Barrère, Chris Hankin, Nicolas Nicolau, Demetrios G. Eliades, and Thomas Parisini. Identifying Security-Critical Cyber-Physical Components in Industrial Control Systems. 5 2019.

[56] Wenbo Wu, Rui Kang, and Zi Li. Risk assessment method for cybersecurity of cyber-physical systems based on inter-dependency of vulnerabilities. In *IEEE International Conference on Industrial Engineering and Engineering Management*, volume 2016-January, pages 1618–1622. IEEE Computer Society, 1 2016.

[57] Pravin Chopade and Marwan Bikdash. Critical infrastructure interdependency modeling: Using graph models to assess the vulnerability of smart power grid and SCADA networks. In *2011 8th International Conference and Expo on Emerging Technologies for a Smarter World, CEWIT 2011*, 2011.

[58] Katherine R. Davis, Charles M. Davis, Saman A. Zonouz, Rakesh B. Bobba, Robin Berthier, Luis Garcia, and Peter W. Sauer. A Cyber-Physical Modeling and Assessment Framework for Power Grid Infrastructures. *IEEE Transactions on Smart Grid*, 6(5):2464–2475, 9 2015.

[59] A Bargiela, M R Irving, and M J H Sterling. OBSERVABILITY DETERMINATION IN POWER SYSTEM STATE ESTIMATION USING A NETWORK FLOW TECHNIQUE. Technical Report 2, 1986.

[60] Ashraf Tantawy, Abdelkarim Erradi, Sherif Abdelwahed, and Khaled Shaban. Model-Based Risk Assessment for Cyber Physical Systems Security. 5 2020.

[61] Chee Wooi Ten, Chen Ching Liu, and Govindarasu Manimaran. Vulnerability assessment of cybersecurity for SCADA systems. *IEEE Transactions on Power Systems*, 23(4):1836–1846, 2008.

[62] Yi nan Wang, Zhi yun Lin, Xiao Liang, Wen yuan Xu, Qiang Yang, and Gang feng Yan. On modeling of electrical cyber-physical systems considering cyber security. *Frontiers of Information Technology and Electronic Engineering*, 17(5):465–478, 5 2016.

[63] Allen J Wood, Bruce F Wollenberg, and Gerald B Sheblé. Power Generation, Operation, and Control. Technical report, Wiley, New Jersey, 1 2014.

[64] Martin Higgins, Jiawei Zhang, Ning Zhang, and Fei Teng. Topology Learning Aided False Data Injection Attack without Prior Topology Information. In *IEEE PES General Meeting (GM)*, pages 1–1, 7 2021.

[65] Zhiyi Li, Mohammad Shahidehpour, Ahmed Alabdulwahab, and Abdullah Abusorrah. Analyzing locally coordinated cyber-physical attacks for undetectable line outages. *IEEE Transactions on Smart Grid*, 9(1):35–47, 1 2018.

[66] Saverio Bolognani, Nicoletta Bof, Davide Michelotti, Riccardo Muraro, and Luca Schenato. Identification of power distribution network topology via voltage correlation analysis. In *52nd IEEE Annual Conference on Decision and Control (CDC)*, Florence, 2013. IEEE.

[67] Xiaozhe Wang and Konstantin Turitsyn. PMU-Based Estimation of Dynamic State Jacobian Matrix. In *ISCAS*. IEEE, 2017.

[68] Seyed Sina Mousavi-Seyedi, Farrokh Aminifar, and Saeed Afsharnia. Parameter estimation of multiterminal transmission lines using joint PMU and SCADA data. *IEEE Transactions on Power Delivery*, 30(3):1077–1085, 6 2015.

[69] Jiafan Yu, Yang Weng, and Ram Rajagopal. PaToPa: A Data-Driven Parameter and Topology Joint Estimation Framework in Distribution Grids. *IEEE Transactions on Power Systems*, 33(4):4335–4347, 7 2018.

[70] Jiawei Zhang, Yi Wang, Yang Weng, and Ning Zhang. Topology Identification and Line Parameter Estimation for non-PMU Distribution Network: A Numerical Method. *IEEE Transactions on Smart Grid*, pages 1–1, 3 2020.

[71] Zong Han Yu and Wen Long Chin. Blind False Data Injection Attack Using PCA Approximation Method in Smart Grid. *IEEE Transactions on Smart Grid*, 2015.

[72] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 2011.

[73] Yilin Mo, Sean Weerakkody, and Bruno Sinopoli. Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems*, 35(1):93–109, 2 2015.

[74] Wen Long Chin, Chun Hung Lee, and Tao Jiang. Blind false data attacks against ac state estimation based on geometric approach in smart grid communications. *IEEE Transactions on Smart Grid*, 9(6):6298–6306, 11 2018.

[75] Laurens Van Der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. Technical report, 2008.

[76] Daniel G. E Silva, Mario Jino, and Bruno T. De Abreu. Machine learning methods and asymmetric cost function to estimate execution effort of software testing. In *ICST 2010 - 3rd International Conference on Software Testing, Verification and Validation*, pages 275–284, 2010.

[77] Nicola Pezzotti, Boudewijn P.F. Lelieveldt, Laurens Van Der Maaten, Thomas Höllt, Elmar Eisemann, and Anna Vilanova. Approximated and user steerable tSNE for progressive visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(7):1739–1752, 7 2017.

[78] Leland McInnes, John Healy, and Steve Astels. Benchmarking Performance and Scaling of Python Clustering Algorithms.

[79] Martin Ester, Hans-Peter Kriegel, Jiirg Sander, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Technical report, 1996.

[80] Charalambos Konstantinou and Michail Maniatakos. A Data-Based Detection Method Against False Data Injection Attacks. *IEEE Design and Test*, 2019.

[81] J. Aghaei, M. Gitizadeh, and M. Kaji. Placement and operation strategy of FACTS devices using optimal continuous power flow. *Scientia Iranica*, 19(6):1683–1690, 12 2012.

[82] Subhash Lakshminarayana and David K. Y. Yau. Cost-Benefit Analysis of Moving-Target Defense in Power Grids. *IEEE Transactions on Power Systems*, pages 1–1, 2020.

[83] L T Gyqyi R Rietman A Edris C D Schauda L willirms. THE UNIFIED POWFBFLOW CONTROLLER: A NEW APPROACH TO POWER TRANSMISSION CONTROL. Technical Report 2, 1995.

[84] Richard Christie. Power Systems Test Case Archive, 1.

[85] Kush Khanna, Bijaya Panigrahi, and Anupam Joshi. AI based approach to identify compromised meters in data integrity attacks on smart grid. *IET Generation, Transmission & Distribution*, 12(5):1052 – 1066, 2018.

[86] Kenneth C Budka, Jayant G Deshpande, and Marina Thottan. Computer Communications and Networks Communication Networks for Smart Grids Making Smart Grid Real. Technical report.

[87] Haibo He and Jun Yan. Cyber-physical attacks and defences in the smart grid: a survey. *IET Cyber-Physical Systems: Theory & Applications*, 1(1):13–27, 12 2016.

[88] Victoria Pillitteri and Tanya Brewer. Guidelines for smart grid cyber security. Technical report, NIST, London, 2014.

[89] Sagarika Ghosh and Srinivas Sampalli. A Survey of Security in SCADA Networks: Current Issues and Future Challenges, 2019.

[90] Song Tan, Debraj De, Wen Zhan Song, Junjie Yang, and Sajal K. Das. Survey of Security Advances in Smart Grid: A Data Driven Approach, 1 2017.

[91] Common Criteria for Information Technology Security Evaluation Part 1: Introduction and general model. Technical report, ISO Security Standards, 2017.

[92] Dan Boneh, Richard A Demillo, and Richard Lipton. On the Importance of Checking Computations. Technical report, Princeton, New Jersey, 1 1997.

[93] Paul C Kocher. Timing Attacks on Implementations of Diie-Hellman, RSA, DSS, and Other Systems. Technical report, Cryptography Research Inc, San Francisco, 1 1996.

[94] Paul Kocher, Joshua Jaaee, and Benjamin Jun. Differential Power Analysis. Technical report, Crytograph Research Inc, San Fransico, 1 1999.

[95] Multos Inc. Multos Trust Anchor Website, 10 2020.

[96] Multos Inc. MULTOS Trust Anchor Development Board Website, 5 2020.

[97] Keith Mayes and Konstantinos Markantonakis. *Smart Cards, Tokens, Security and Applications, Chapter 17*, volume 1. Springer, London, 1st edition, 1 2017.

[98] Keith Mayes, Steve Babbage, and Alexander Maximov. *Performance Evaluation of the new TUAK Mobile Authentication Algorithm*.

[99] Keith Mayes. Performance of Authenticated Encryption for Payment Cards with Crypto Coprocessors. In *ICONS*, pages 10–15, London, 1 2017. Thinkmind.

[100] Keith Mayes. Performance Evaluation and Optimisation for Kyber on the MULTOS IoT TrustAnchor. In *Proceedings - 2020 IEEE International Conference on Smart Internet of Things, SmartIoT 2020*, pages 1–8. Institute of Electrical and Electronics Engineers Inc., 8 2020.

[101] EMVco LLC. EMVco Payment Processing Website, 10 1999.

[102] IEC. 19772 Information technology - Security techniques - Authenticated encryption. Technical report, International Electrotechnical Commission, London, 1 2009.

[103] James W. Taylor and Patrick E. McSharry. Short-term load forecasting methods: An evaluation based on European data. *IEEE Transactions on Power Systems*, 22(4):2213–2219, 11 2007.

[104] National Grid. Electricity Ten Year Statement. Technical report, National Grid, London, 10 2017.

[105] Ammara Gul and Stephen Wolthusen. In-cycle sequential topology faults and attacks: Effects on state estimation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11260 LNCS, pages 17–28. Springer Verlag, 2019.

[106] A. Ahmad, M. A.K. Rizvi, A. Al-Lawati, I. Mohammed, and A. S. Malik. Development of a MATLAB tool based on graph theory for evaluating reliability of complex mechatronic systems. In *2015 IEEE 8th GCC Conference and Exhibition, GCCCE 2015*. Institute of Electrical and Electronics Engineers Inc., 3 2015.

Chapter 9

Appendix

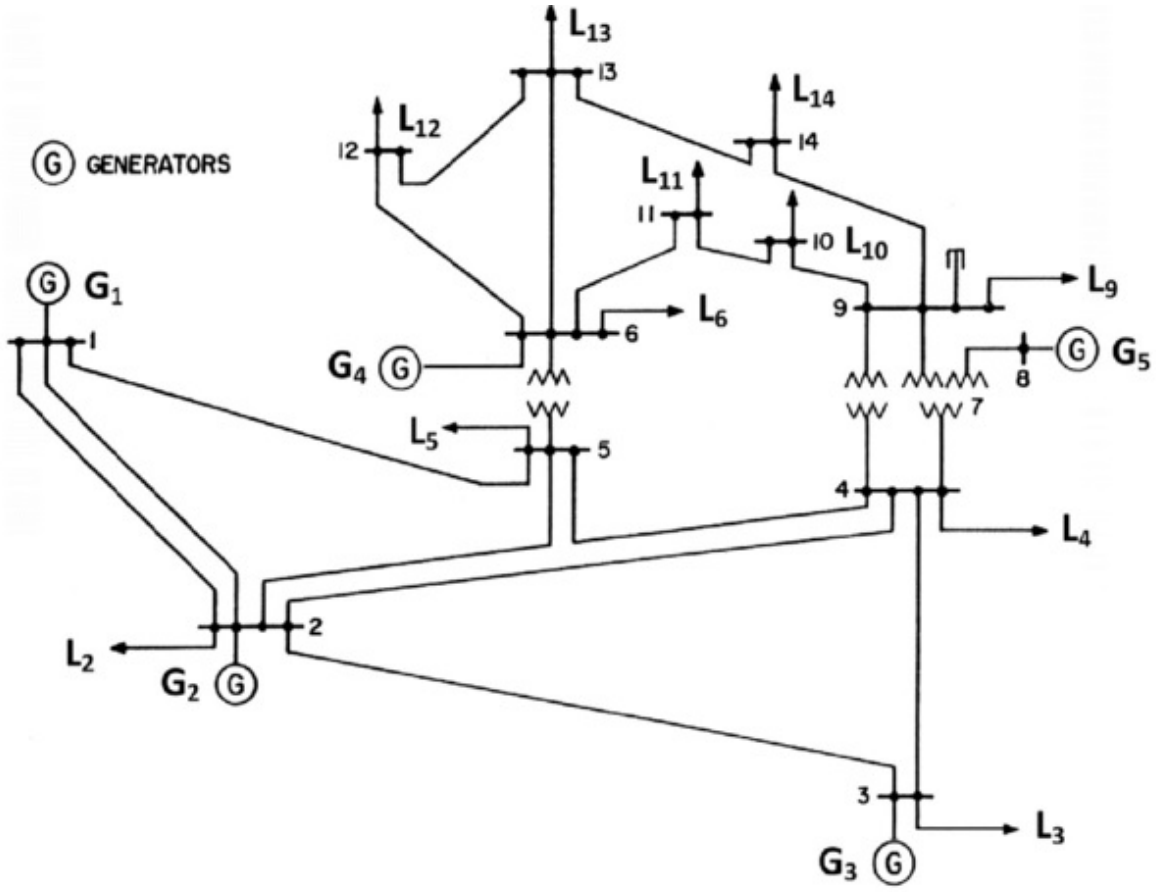

Figure 9.1: IEEE 14-bus system used for simulation

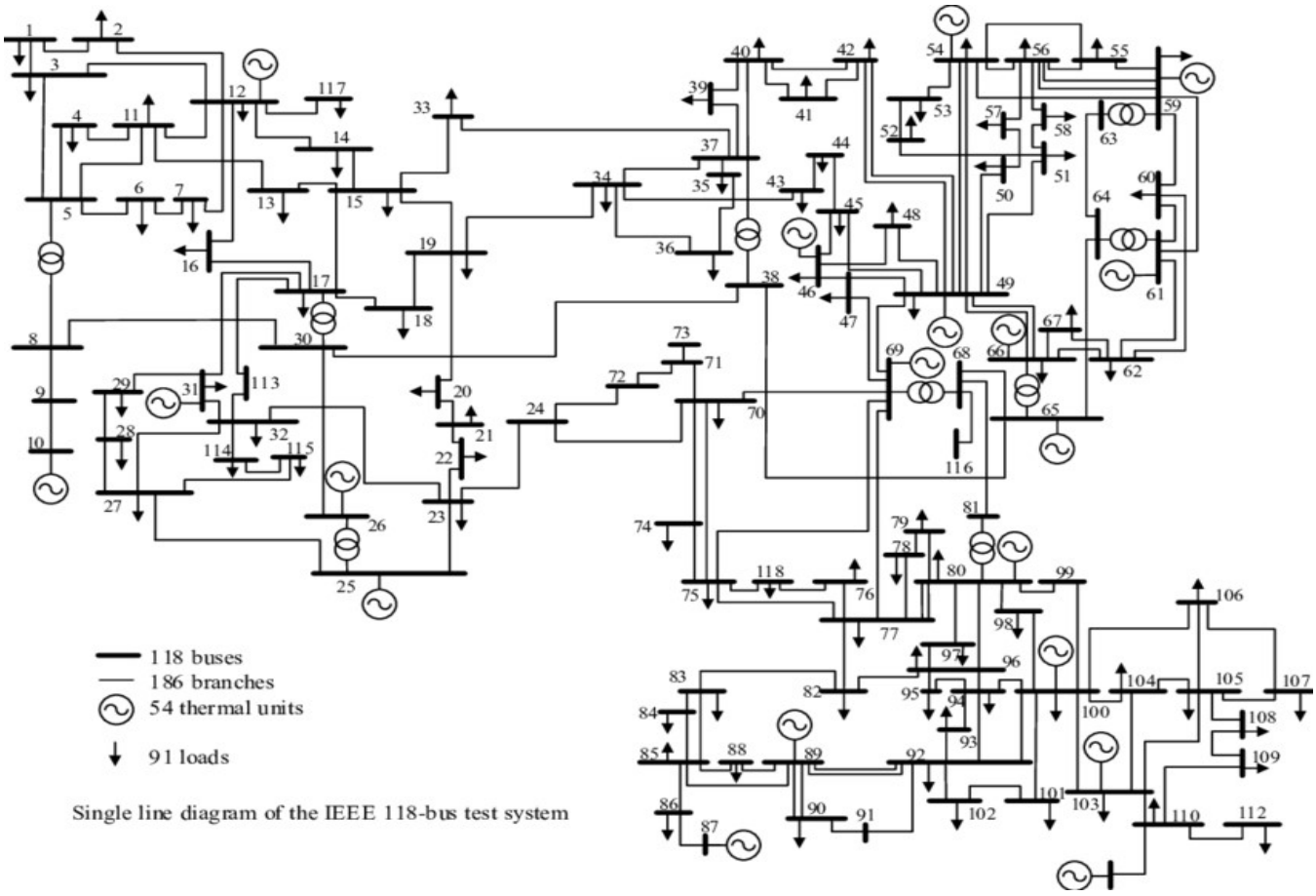

Figure 9.2: IEEE 118-bus system used for simulation