http://www.inosr.net/inosr-applied-sciences/
Kalulu and Adabara
INOSR APPLIED SCIENCES 9(1):1-8, 2022.

International Network Organization for Scientific Research ISSN: 2705-165X

Developing of A Smart Fraud Detection System in Advanced Metering Infrastructure Using Deep Learning

Kalulu Mathias and Adabara Ibrahim

Department of Electrical, Telecommunications & Computer Engineering School of Engineering & Applied Sciences Kampala International University Uganda.
Email:kalumathias@gmail.com

ABSTRACT

Modern power systems worldwide are facing a rising appeal for the upgrade to a highly intelligent generation of electricity networks commonly known as the Smart Grid. Advanced monitoring and control systems like Supervisory Control And Data Acquisition (SCADA) and Advanced Metering Infrastructure (AMI) systems have been widely deployed and management based on them provides more flexible and achievable optimal control of power generation, transmission and distribution. However, the growing integration of power system with communication networks also brings increasing challenges to the security of the modern power grid, from both cyber and physical space. Malicious attackers can take advantage of the increased access to the monitoring and control of the system and exploit some of the inherent structural vulnerability of power grids. Motivated by these security challenges, the goal of this thesis is to facilitate the understanding of power grid outages and blackouts triggered by these attacks, to analyze the cascading process that leads to the impactful events, and to support the decision making in defence and protection for a reliable and secure Smart Grid around the corner. Simulation results from real-world power system benchmarks have been analytically discussed from both the spatial and temporal perspectives and important decision-support information have been revealed through this research.
Keywords: Smart, Fraud, Detection system and Infrastructure.

INTRODUCTION

Since its birth over a century ago, the power system in the United States has been evolved into one of the largest complex networks ever witnessed in human history [1,2,3,4]. With the increasing consumption demand, the modern electrical power grids are now growing into a mammoth system with numerous interconnected regional grids, owned and operated by power corporations at all levels and scales [5]. However, complex interests, operations and managements among different power companies often complicate cross-region transmission tasks and sometimes result in an inefficient or poorly-coordinated power delivery [6]. And the traditional power grids in modern times are facing some rising challenges. As the need and variety of consumption increases, more and more latest technologies have been incorporated into the power system, such as the distributed renewable energy generation, the electric vehicle (EV) charging system, smart meters, etc., which all contribute to the complexity of modern power delivery [7]. The ever increasing reliance on electricity and request of power quality have been constantly calling for better power delivery, more flexible pricing, faster power restoration, among others [8]. The challenges above have motivated the industry and society for a new generation of power system with more informative, intelligent and automatic operations. This new generation of power grid, commonly referred to as the Smart Grid, will modernize the traditional power grid and improve its reliability, flexibility and

efficiency. The Smart Grid is expected to have more distributed controls and consumer-based interactions, and some key applications of this upgrade include:

1. Installing advanced metering infrastructure (AMI) and other intelligent de- vices to reduce extra unnecessary demand, as well as the operation and main- tenancy cost.
2. Implementing distributed automation to enhance the reliability of power sys- term.

Utilizing automated controls for better and flexible power management The most significant feature of the Smart Grid, in contrast to the traditional power system, is the large scale of implementation of a two-way communication network connecting both the power plants, the transmission network sensors and the consumers [9]. This enhanced and interactive system will optimize the power delivery quality, efficiency and stability at a lower cost via the computer-based automation. With the incorporation of communication and computer networks in assistance to the traditional power delivery, the Smart Grid is already an emerging gigantic intelligent network system in which power flows are flexibly directed by highly automatic systems like the Supervisory Control and Data Acquisition (SCADA) system [10, 11].

However, the Smart Grid yields not only a boost of economic benefits but also a growing number of potential threats from the cyberspace [1]. While the distributed control can reduce the criticality of some control centre and thus weaken the impact of attacks targeted on them, the distributed access from the system-on- a-chip (SoC) electrical devices can possibly allow malicious attackers penetrating into the power systems with increased difficulties to detect and track them. Mean- while, with the huge volume of data flowing along the power transmission network, they are becoming more vulnerable to data or command interception and unauthorized modification, which can be utilized to either cheat the meters for an unfair price or disrupt power system operations. More seriously, knowledge of the power Grid can be learned and the information or intelligence could be used to hack into the distributed control units that may be less protected than the centralized operation and management hubs, resulting in unpredictable security risks. Hence, it becomes crucial to realize and react to the vulnerability of Smart Grid in the new forms of potential attacks. These “smart” attacks, if deliberately designed and launched successfully on some critical components, can cast a disastrous impact on the power grid transmission and significantly jeopardize the interest of both the public, the industry as well as the economics. The traditional power grids have transformed into smart grids by the rapid integration of technologies According to [2]. Smart Grids is a concept regarding digital technology application and electric power network. Smart Grid includes electric network, digital control appliance, and intelligent monitoring system. All of these, can deliver electricity from producers to consumers, control energy flow, reduce the loss of what, and make the performance of the electric network more reliable and controllable [2].

Objective

The general objective is to develop a smart fraud detection system in Advanced metering infrastructure.

METHODOLOGY

Power Grid Modelling

This chapter describes the model of power grid and cascading failure adopted in this study, which provides the simulation model for the purpose of cascading failure analysis.

Topological Model

Most of the topological information on the structure of a power grid can be obtained in the form of geospatial dataset. It has been standardized thanks to the effort of power and energy industry, geographic information system (GIS) industry, as well as government supported academic institutes like the Power Systems Engineering Research Center (PSERC) and the National Institute of Standards and Technology (NIST). In contrast to the wide area dynamic operational states and parameters stored and protected in the control centers, these topological data are more accessible from commercial companies or academic institutes. They provide the time-invariant information that does not subject to the transient state of power systems, which makes it a useful source of knowledge in the vulnerability assessment of power grids. In order to represent the power grid as a topological network, there are a few assumptions to be specified: First, as part of a transmission network, a substation in our power grid cascading model is referred to as a node, regardless of its type as a generator, a load or simply a pass-through transmission substation; also, a transmission line which connects one substation at each end will be regarded as a branch in the network. Hence the power grid is regarded as a bidirectional unweighted graph, a simplification that helps to reduce the computational cost significantly marks the end of a cascading failure. The redistribution can cause overloading on some of surviving neighbours of victim v in the grid and can lead to subsequent cascading failures. So, considering a non-recoverable scenario, when a node is overloaded to a certain degree, it will be regarded as fatally overloaded and cut off from the network with all the branches that directly linked to it. The threshold of overloading ratio, above which a node is considered failed, is then referred to as the system tolerance, and currently we assign a universal system tolerance, denoted as T, for all the nodes in the network. The failure propagation will continue as long as new fatally overloaded nodes emerge in the grid, leading to a cascading failure across the network. If the initial victims are well-selected, the malicious attacker will be able to create a remarkable blackout in its scale or speed of propagation in the power system. Note that there is no ground-truth for the practical value of T, and so in the simulation different values of T will be tested for this important factor, which will also affect the cascading process significantly. Finally, when a number of nodes are failed, we use the concept of "round" to help describing the progress of failure cascading. The very first set of failed nodes consists of the victims in the initial attack. Then the nodes knocked down by the cascading failure of initial victims will be regarded as the victims of second round, so on and so forth. In this way, failed nodes at different rounds of a cascading process form a tree-like structure where the "child" nodes are the direct victims of their parent node's failure, and the root nodes are the initial multi-victims set of attacks. In this structure a node may have more than one parent if it is affected by multiple nodes failure at the same time, which represents the overlapping of multi-victim cascading, As the failure propagates through the surviving networked nodes, the load of the remaining system will be shed due to the impaired transmission capacity. At a certain moment, the load will be likely to drop to an extent low enough that the load carried by each branch will all stay below the fatal overloading threshold and thus the cascading process will come to a final stabilized state. The simulator will then stop the simulation and return all the data recorded during the cascading process for further analysis. One more fact to notice is that the load and status

of each node is only updated once at each round, which means the nodes failed in the same round will not have instant effect on others. Instead, the failure of all nodes of last round will simultaneously affect the remaining active nodes in next round. In this sense, our cascading model better matches a simultaneous failure process.

RESULTS

Spatial Analysis for Cascading Failure in Power Grids

In most researches, the information on the power systems are usually represented with the digitalized measurement, for instance generator voltage, load, phase angle, active and reactive power, among others. They are recorded as numerical values and tables, which is convenient for quantitative studies. Although many of them are accompanied by some visualization tools, it is noted that most of them did not take it as a formal form of security analysis. In many cases, the power transmission network is only represented as an abstract grid which only emphasizes the information on the substation type (generation, transmission and distribution) and the existing transmission lines connecting them. Their exact locations and lengths are usually not considered since acquiring and maintaining such information with their detailed status are either difficult (with remote locations) or expensive (to install meters and keep track of the huge amount of data generated). However, with the development of geographic information system (GIS), more and more up-to-date power grid data become available with improved accuracy. As a result, a growing number of GIS based analyses that have become popular among the study of smart grid security.

The GIS information of power grids can also be a potential intelligence re- source for the attackers. These data are usually more accessible than the internal data guarded within the power control centers; yet they provide some informative details like the power grid topology and connectivity, the voltage categories, the power plants and their ownership, the length of transmission lines that could be used to approximate the admittance and resistance to solve power flows and simulate effectiveness of potential attacks. Therefore, we are interested in visualizing the power grids in the space and assessing their structural vulnerability through the spatial patterns. Specifically, in this chapter the geospatial location and connectivity of power grid substations will be visualized for cascading analysis.

VISUALIZE THE CASCADING ATTACK

The Cascading Simulator

MATLAB is chosen as the simulation platform and the simulator will record both the stabilized grid after cascading as well as every intermediate process during the failure propagation. The failure sequence of nodes (in forms of substation IDs) will be forwarded to a MATLAB interface, which searches the victim nodes and branches in GIS database and updates the database with a new attribute as an identifier, form of layers, which could be overlapped to provide customized visualization for analysis. In our platform, we deploy 2 layers to represent the transmission network of a Smart Grid: substations and transmission lines. Note that the GIS databases are stored in the “shape file” format in our experiments. In the cascading model, as described in this paper, we assume that when a node (substation) is down, its load is redistributed to its neighbours. Also, we assume that there is only one transmission line (branch) between two nodes, and multiple branches between two nodes will be treated as one. Under these assumptions, we start the attack by picking a single victim node in the power grid. Once a node is knocked down, all the branches connected to it will also be considered as failed at the same time. Using a constant tolerance for the grid, the algorithm recalculates the load distribution iteratively and a cascading procedure will be generated throughout the grid until the grid reaches a final stabilized status. A cascading failure is

assessed by the fraction of failed nodes, denoted by P of at each round. In other words, in every new round a set of new victim nodes emerges. Victims in each round could be updated to the visualization interface, which is called an online model; or their information could be stored in a data sequence until the cascading process finalized, and then the failures of all rounds could be exported as a whole. In general, the process of this attack simulation and visualization could be described as below:

1. Build up the topology for a power grid;
2. Calculate the metrics value of each node to choose the most vulnerable node to be the first victim node;
3. Recalculate the distribution of load and find out new victims at each round, store the victim substation IDs;
4. Communicate the results from MATLAB to GIS database, and update the visual information in the ArcGIS;
5. Repeat the process until no more failure nodes can be identified, i.e., the end of the cascading failure process.

Interface between MATLAB and ArcGIS

In ArcGIS, our GIS database is stored in a format called "shape files", defined by Environmental Systems Research Institute (ESRI) as a popular geospatial data format for GIS systems. A dataset usually consists of three shape files for each Single layer: a .ship file containing primary geographic reference data, a .dbf file storing all the attribute values and a .six file saving the shape index table. It may also contain a projection file (.pry) or a spatial index file (.son), which is not used in our current experiments. The database format is ideal for visualization as it can be imported into two sets of fields in MATLAB, one fixed set containing the geometry information such as the coordinates and types of object (e.g., point, link or polygon), and another set of attributes includes all other specific information of the object. For entries like power plants and substations, the attribute set stores information including information of location, load, owner, voltage cat- gory, generator/fuel type, etc. For branches (transmission lines) the attribute set could provide voltage category, length in mile or kilometres, IDs and names of nodes connected, among others. These set attributes could be easily customized and updated, which is ideal for real-time monitoring and management through a visualization platform. The shape file database can be imported to MATLAB as structure arrays by using MATLAB mapping toolbox. The mapping toolbox is a set of tools and utilities to process geographic data analysis and map displaying. It was first introduced in MATLAB 6.0 and is updated with every following release of MATLAB. An interface in MATLAB is also developed to update the cascading information into shape files, which can be described as four steps, the data pre-loading, the initialization, the search and update of victims, and the export.

Pre-load Data

At the first step, MATLAB loads a sequence of a cascading failure generated by the attack algorithm. The failure sequence is an M-by-N matrix of which all element values are initialized to zero, where M is the number of rounds of cascading and N is the number of final victim nodes. For each row in the matrix, the row in failure sequence contains the IDs of all victim nodes that have already been knocked down at the jth round. In addition, MATLAB also loads the shape files from both databases of subsections and transmission lines, matching up the victim nodes information with the failure sequence. The import is done by a function called shape read from MAT-LAB's mapping toolbox, which reads geo-information stored in shape file format and arrange them into an array of geo-structures. The compact geo-structure array is MATLAB-friendly and easy to customize, providing fast calculation as well as good compatibility. After this step, we store three arrays in MATLAB: an M-by-N failure sequence called F ail sew,

a K-by-1 structure array of nodes called NODE and an L-by-1 structure array of branches called LINK, where K and L are the numbers of substations and transmission lines in our database, respectively. In our case, K =553 and L =726.

Initialization

At the second step, our interface customizes the database by introducing a new field FAIL to each element of both structure arrays NODE and LINK, which would be exported as one attribute for each database. In our current development, this attribute will simply record the round number when the current object is taken down by the failure cascading. A zero value of FAIL means the current object (node or link) has not been affected by the cascading, while a positive value k indicates that the current object is already taken down in the round. In this way, our VBA script could effectively read the FAIL attribute and visualize its value in ArcGIS. Grid in the Bay Area, which possibly means that they could be part of another power grid not fully covered in this regional snapshot.

Visualization in ArcGIS

The objective of the Visualization is to acquire the GIS database containing data for substation and transmission line failure and then display the results in ArcGIS in an animated way, so that the cascading failure effects are effectively displayed and the critical nodes can be quickly identified. In addition, the interface should allow user interaction, meaning that users can manually select attack victim on the platform and see the cascading effect due to the failure of such nodes. Therefore, users can compare and contrast the vulnerability of nodes visually and identify the nodes that are most critical. The objective is achieved by Visual Basic for Applications (VBA) program- Ming in ArcGIS Desktop, which is one of the application components of ArcGIS. It consists of Arc Objects, a set of platform-independent software components written in C++, which provides services to support GIS applications on the desktop in the form of thick and thin clients and on the server. The results of the visualization program are displayed in Arc Map, the main component of ArcGIS Desktop, together with geographical base map. The flowchart for visualization is shown in Fig.1.The VBA script has 4 major steps:

1. Retrieve selected attack victim node from user and pass the node information to MATLAB;
2. Read shape files to obtain the node and link information after Matlab finishes calculation and returns the result;
3. Draw all categories for both substation and transmission line data according to the value in the FAIL attribute;
4. Apply colour for better visualization.

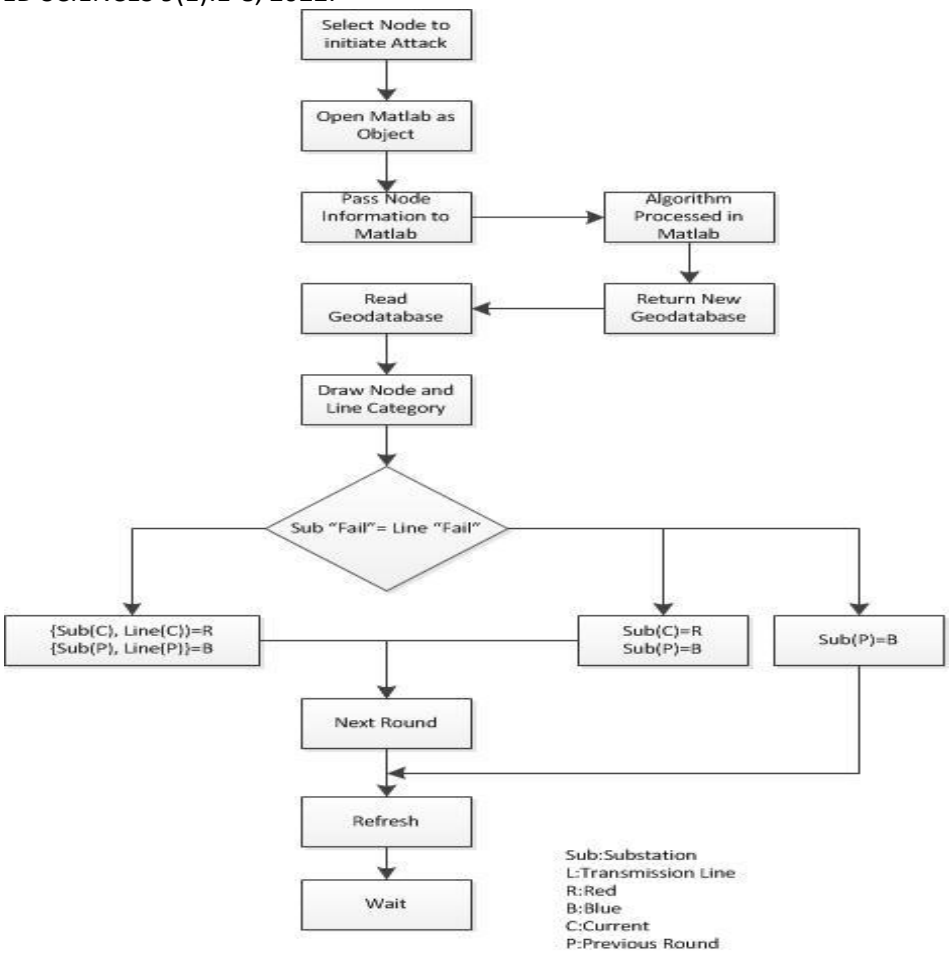

Figure 1: Visualization platform

Simulation Result

The visualization platform has been developed with MATLAB R2010b and ArcGIS 10.0. The data are a snapshot extracted from the Powernap GIS dataset provided by PLATTS, including 726 transmission lines and 553 substations to represent the major power grid of upper Bay area in San Francisco, California, as shown in Fig. Each link and node contains attributes that could be imported into MATLAB in forms of structure arrays.Are not around the edge of most recent cascading failure areas. We refer to the critical moment as the round when new victim nodes become disconnected in a remote region in comparison to the existing victim nodes, a phenomenon that poses more challenges to the protection of power grids and requires more inter-regional cooperation among different power companies and operators. Our experiments show the existence of such critical moments of failure migration, which are also observed and described in [2] in a power grid structured in a theoretical rather than practical way, whose observation was that "the consecutive cascading failure can occur at an arbitrary long distance".

DISCUSSION AND CONCLUSION

AMI in the grids is very important part, which is always a target of cyber criminals for various purposes including energy theft. The cyber defence for AMI to detect and prevent these vulnerabilities is always a challenging task. In this research, we presented a novel technique utilizing unsupervised and supervised machine learning Techniques to detect suspicious data. We also devised an algorithmic approach to find out suspicious nodes in the AMI network that

may be responsible for the data manipulation while forwarding energy consumption data from smart meters.

REFERENCES

1. Javier Hernando and Climent Nadeu, Speech Recognition In Noisy Car Enviornment Based On OsaLPC Representation And Robust Similarity Measuring Techniques, Signal Theory and comm.. Dept., Spain, 2004
2. Zhu, S., Lee, J. S., Guo, F., Shin, J., Perez-Atayde, A. R., Kutok, J. L., Rodig, S. J., Neuberg, D. S., Helman, D., Feng, H., Stewart, R. A., Wang, W., George, R. E., Kanki, J. P., and Look, A. T. (2012) Activated ALK Collaborates with MYCN in Neuroblastoma Pathogenesis. Cancer Cell. 21(3):362-373.
3. Ghori, K. M., Abbasi, R. A., Awais, M., Imran, M., Ullah, A. and Szathmary, L. (2019). Performance Analysis of Different Types of Machine Learning Classifiers for Non-Technical Loss Detection. IEEE Access, 8, pp.16033-160484.
4. Mohammed, H., Tonyali, S., Rabieh, K., Mahmoud, M. and Akkaya, K. (2016). Efficient privacy-preserving data collection scheme for smart grid AMI networks, in GLOBECOM. IEEE, pp. 1–6.5
5. Jokar, P., Arianpoo, N. and Leung, V. C. (2015). Electricity theft detection in AMI using customers' consumption patterns. IEEE Transactions on Smart Grid, 7(1), pp.216-226.
6. Kadurek, P., Blom, J., Cobben, J. and Kling, W. L. (2010). Theft detection and smart metering practices and expectations in the Netherlands. IEEE, pp. 1–6.
7. Masisani William Mufana and Adabara Ibrahim (2022). Monitoring with Communication Technologies of the Smart Grid. *IDOSR Journal of Applied Sciences* 7(1) 102-112.
8. Nabiryo Patience and Itodo Anthony Ejeh (2022). Design and Implementation of Base Station Temperature Monitoring System Using Raspberry Pi. *IDOSR Journal of Science and Technology* 7(1):53-66.
9. Masisani William Mufana and Adabara Ibrahim (2022). Implementation of Smart Grid Decision Support Systems. *IDOSR Journal of Scientific Research* 7(1) 50-57, 2022.
10. Natumanya Akimu (2022). Design and Construction of an Automatic Load Monitoring System on a Transformer in Power Distribution Networks. *IDOSR Journal of Scientific Research* 7(1) 58-76, 2022.
11. Kisakye Rebecca (2022).Simulation and Analysis of Dipole Transmitter Antenna (KIU Laboratory) *IDOSR Journal Of Computer And Applied Sciences* 7(1):119-135.