ISSN: 2581-3404 (Online)
*International Journal of Innovative Research in Technology and Management, Vol-5, Issue-4, 2021.*

# A Review of Effective Intrusion Detection System using Data Mining Techniques

**Saleha Khan[1], Prof. Zuber Farooqui[2]**

**[1]M. Tech. Scholar, Department of CSE, ASCT, Bhopal, M.P. (India)**

**[2]Assistant Professor, Department of CSE, ASCT, Bhopal, M.P. (India)**

**Abstract-** *Cyber security plays an important role in maintaining the vital national infrastructure. Critical infrastructures are selected on the basis of whether or not their disabilities have a debilitating effect on security, the national economy, public health or public safety or a combination of these. Maintaining the security of industrial systems is important for any organization. As history shows us, the consequences of not doing so can be potentially damaging. Intrusion detection systems (IDS) are commonly categorized into misuse based, anomaly based and specification based IDS. Both misuse based IDS and anomaly based IDS are extensively researched in academia and industry. However, as critical infrastructures including smart grids (SG) may often face sophisticated unknown attacks in the near future, misuse based attack detection techniques will mostly miss their targets, this article will provide a survey for the different techniques for intrusion detection in smart rid and nay other industry or an organization using some techniques such as the data mining, classification techniques and, hybrid techniques.*

**Keywords:-** Intrusion detection systems, Data mining, Machine learning, Cyber security, Smart grid.

## Introduction

Detecting malicious network intrusions has been a subject of study for decades. As data scientists can appreciate, however, when the scale of a problem grows by an order of magnitude, existing approaches often are no longer effective; the problem is sufficiently different that it requires a new solution. As the volume of network traffic has grown through orders of magnitude, the field of intrusion detection has had to re-invent itself around big data techniques. An intrusion detection system (IDS) monitors either networks or other systems for malicious or anomalous behaviors. Complementing preventative technologies such as firewalls, strong authentication, and user privilege, IDSs have become an essential part of enterprise IT security management. They are typically classified as either misuse based or anomaly-based systems. Data Mining (DM) techniques are increasingly being used to identify attacks, anomalies or intrusions in a protected network environment. DM can be defined as the process of discovering interesting patterns in databases that are useful in decision making. On the other hand, machine learning is the attempt to automate the process of knowledge acquisition from examples [7].

Smart grid is a combination of electrical, computational and communication network. The incorporation of communication technology into power grid improves the quality and reliability of power system. The communication networks and smart grid devices exposes the power grid to cyber threats. Several security mechanisms like cryptographic algorithms and secure protocols are available to transmit the smart meter data in a secured way, but the attacks on smart meter and its communication network may affect the

functioning of smart grid. Hence, appropriate intrusion detection system is required to detect attacks and malicious activities in the smart grid. An intrusion detection system (IDS) is a passive monitoring system that monitors the suspicious activity of network for detecting attacks and eavesdropping of data by intruders. It acts as second layer of defense during the failure of cryptographic mechanisms [3].

An intrusion detection system (IDS) finds functionalities that abuse the security agreement of a system program or a computer network. IDS must agree to preserve security mechanisms. Firewall for IDS is used to identify threats that cause program design bugs. IDS allow forensic suspicion to acknowledge the program admin's actions against cyber threats. Intrusion detection systems detect possible intrusions in the network or system. Specifically, IDS tools aim to detect system threats or system misuse, and to alert the proper personals upon detection. An IDS inspects all incoming and outgoing network tasks and finds suspicious patterns that may indicate a network or system attack from someone trying to compromise a computer. An IDS on a system/network provides the same application as a burglar alarm system fixed in a
house. Both detects faults when an intruder is available, and both issues alert [4]. IDS is based on three distinct approaches in detecting abnormalities: signature-based, specification-based, and anomaly-based. The first approach consists in detecting patterns of malicious activities using a database of well known attack signatures. The second approach consists in identifying deviation from normal behavior profiles using logical specifications. The third approach consists in looking for deviations from normal behavior profiles using statistical measures [2]. Anomaly-based technique is more applicable and suitable approach for smart grid than signature-based and specification-based techniques.

To prevent intrusion, the smart grid confides on classical security strategy which includes firewall and password protection. Intrusion detection Mechanism (IDM) is capable to generate alarms for viable intrusions via constantly monitoring operations. Although there are several research on well-known IDS in system safety, limited effort has been made especially to the smart grid. Generally, two types of IDM system is used named as: data sourced based and detection based method [5].

A good number of soft computing based methods and solutions have been adapted in IDS in cyber security research works to improve detection accuracy and efficiency. Seeking to imitate human intelligence, to improve learning and decision-making processes and thus to solve real-world problems, soft computing as a general term describes a set of optimization techniques including fuzzy logic, artificial neural networks, probabilistic reasoning, association rule mining, genetic algorithms, particle swarm intelligence, ant colony optimization etc. Often their applications in IDS are inspired by the successful use cases in other scientific fields like medicine, bioinformatics, economics, computer networks etc. What makes them attractive to be applied in intrusion detection is that soft computing techniques are capable of handling uncertain and partially true data, which are oftentimes seen in cyber security research field. Inexpensive solutions are achieved with an acceptable trade-off between robustness and forbearance for inaccuracy and partial truth. When used in intrusion detection, soft computing techniques can be complemented by rule based expert knowledge described as a set of IF-THEN rules.

## II. Data Mining

Data mining is the extraction of interesting patterns or knowledge from huge amount of data. It can be known by different names like knowledge discovery (mining) in Databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence and others. The term "data mining" [2] is nothing but analysis of data in a database using tools which

look for trends or anomalies without the knowledge of meaning of the data and is primarily used by statisticians, database researchers and business communities. A data mining software does not just change the presentation, but discovers previously unknown relationships among the data. The information on which the data mining process operates is contained in a historical database of previous interactions.

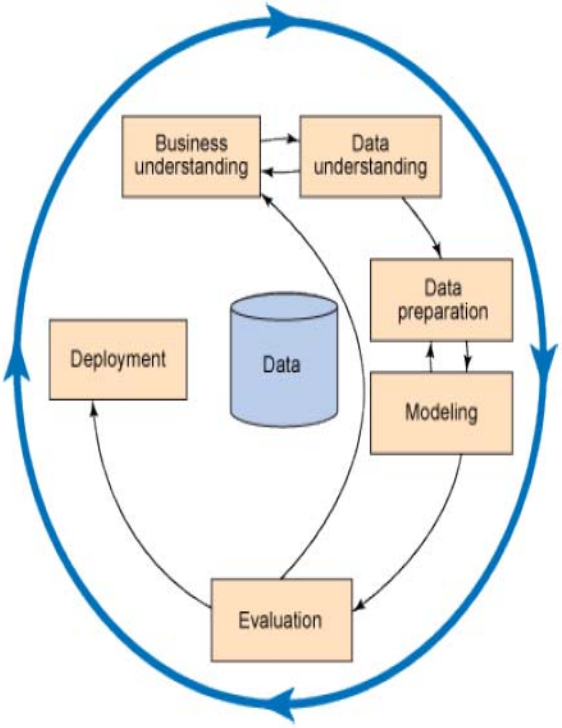


**Fig 1:** The data mining process.

### III. Related Work

With plenty of studies performed to review the current state of intrusion detection systems, the foci vary from one to another in analyzing, comparing and summarizing the investigated intrusion detection techniques and identifying research gaps and future research directions. For instance, In Smart environments there are uses of security devices and smart appliances, sensors and energy meters. New requirements in security and privacy are driven by the massive growth of devices numbers that are connected to IoT which increases concerns in security and privacy. The most ubiquitous threats to the security of the smart grids (SG) ascended from infrastructural physical damages, destroying data, malwares, DoS, and intrusions. Intrusion detection comprehends illegitimate access to information and attacks which creates physical disruption in the availability of servers. This work [1] proposes an intrusion detection system using data mining techniques for intrusion detection in smart grid environment. The results showed that the proposed random forest method with a total classification accuracy of 98.94 %, F-measure of 0.989, area under the ROC curve (AUC) of 0.999, and kappa value of 0.9865 outperforms over other classification methods. In addition, the feasibility of our method has been successfully demonstrated by comparing other classification techniques such as ANN, k-NN, SVM and Rotation Forest.

Smart grid is an emerging and promising technology. It uses the power of information technologies to deliver intelligently the electrical power to customers, and it allows the integration of the green technology to meet the environmental requirements. Unfortunately, information technologies have its inherent vulnerabilities and weaknesses that expose the smart grid to a wide variety of security risks. The Intrusion detection system (IDS) plays an important role in securing smart grid networks and detecting malicious activity, yet it suffers from several limitations. Many research papers have been published to address these issues using several algorithms and techniques. Therefore, a detailed comparison between these algorithms is needed. This paper [2] presents an overview of four data mining algorithms used by IDS in Smart Grid. An evaluation of performance of these algorithms is conducted based on several metrics including the probability of detection, probability of false alarm, probability of miss detection, efficiency, and processing time. Results show that Random Forest outperforms the other three algorithms in detecting attacks with higher probability of detection, lower probability of false alarm, lower probability of miss detection, and higher accuracy.

Security of communication network is essential for the smooth functioning of smart grid. In this paper [3], an intrusion detection system is proposed for early detection of threats in advanced metering infrastructure of smart grid. The proposed intrusion detection system has a multi-support vector machine classifier with mutual information based feature selection technique to detect attacks in Neighborhood Area Network (NAN) of smart grid. Mutual information technique selects the input features of classifier by analyzing the relation between different features with attacks. The developed classifier is the integration of multiple support vector machine classifiers in which each classifier detect specific attack only. The performance of developed intrusion detection system is analyzed by training and testing the classifier with ADFA-LD dataset.

The power industry is not an exception in the technological advancement which makes everything smarter. Smart grid is the advanced version of the traditional grid, which makes the system more efficient and self-healing. Synchrophasor is a device used in smart grids to measure the values of electric waves, voltages and current. The phasor measurement unit produces immense volume of current and voltage data that is used to monitor and control the performance of the grid. These data are huge in size and vulnerable to attacks. Intrusion Detection is a common technique for finding the intrusions in the system. In this paper [4], a big data framework is designed using various machine learning techniques, and intrusions are detected based on the classifications applied on the synchrophasor dataset. In this approach various machine learning techniques like deep neural networks, support vector machines, random forest, decision trees and naive bayes classifications are done for the synchrophasor dataset and the results are compared using metrics of accuracy, recall, false rate, specificity, and prediction time. Feature selection and dimensionality reduction algorithms are used to reduce the prediction time taken by the proposed approach. The proposed classifier outperforms the other machine learning approaches like artificial neural network in the detection of attacks.

The smart grid is an intelligent and complex system designed to work more efficiently, reliable, and economical with the help of computational technologies, advanced communication infrastructure, and state-of-the-art monitoring stations. This paper [5] proposes an Intelligent Loop Based Artificial Neural Network (ILANN) based detection technique for the detection of cyber intrusion in a smart grid against False Data Injection Attack (FDIA). This method compares the deviation of a system with the equipment load profile present on the system node(s) and any deviation from predefined values generates an alarm. Every 2 milliseconds (ms) the data obtained by the measurement is passed through the attack detection system, in case if the deviation is continuously for 5 measurement cycles i.e. for 10 ms and it does not match with the load combination the operator will get the first alert alarm. In case the deviation is not fixed after 8 measurement cycles then the system alerts the control centre.

Intrusion Detection System (IDS) is one of the significant ways to provide secure and reliable services in a smart grid environment. In this paper, [6] they propose intrusion detection framework for the smart grid. They consider the three-layer architecture of smart grid system. The proposed framework has an IDS in each HAN and NAN and many IDS sensors in WAN. Any malicious activity will be sent to the central management unit; the central management unit correlates and investigates alerts produced by various distributed sensors using anomaly based detection methodology. IDS management system will collect and preprocess the alerts of all sensors and correlate these alerts to distinguish symptoms of attack and contravention of security policy.

The continued ability to detect malicious network intrusions has become an exercise in scalability, in which data mining techniques are playing an increasingly important role. In this article [7], they

survey and categorize the fields of data mining and intrusion detection systems, providing a systematic treatment of methodologies and techniques. They apply a criterion-based approach to select 95 relevant articles from 2007 to 2017.We identified 19 separate data mining techniques used for intrusion detection, and our analysis encompasses rich information for future research based on the strengths and weaknesses of these techniques.

The smart grid is a revolutionary, intelligent, next-generation power system. Due to its cyber infrastructure nature, it must be able to accurately and detect potential cyber-attacks and take appropriate actions in a timely manner. This paper [9] creates a new intrusion detection model, which is able to classify the binary-class, triple-class, and multi-class cyber-attacks and power-system incidents. The intrusion detection model is based on a whale optimization algorithm (WOA)-trained artificial neural network (ANN). The WOA is applied to initialize and adjust the weight vector of the ANN to achieve the minimum mean square error. The proposed WOA-ANN model can address the challenges of attacks, failure prediction, and failure detection in a power system.

Based on the rapid evolution of the cyber-physical systems (CPS), both academia and industry have developed appropriate measures for enhancing the security surface of the SG paradigm using, for example, integrating efficient, lightweight encryption and authorization mechanisms. Nevertheless, these mechanisms may not prevent various security threats, such as denial of service (DoS) attacks that target on the availability of the underlying systems. An efficient countermeasure against several cyber attacks is the intrusion detection and prevention system (IDPS). In this paper [10], they examine the contribution of the IDPSs in the SG paradigm, providing an analysis of 37 cases. More detailed, these systems can be considered as a secondary defense mechanism, which enhances the cryptographic processes, by timely detecting or/and preventing potential security violations. For instance, if a cyber attack bypasses the essential encryption and authorization mechanisms, then the IDPS systems can act as a secondary protection service, informing the system operator for the presence of the specific attack or enabling appropriate preventive countermeasures.

## IV. Problem Identification

The smart grid improves the efficiency, reliability and economics of current energy systems through the integrated use of both information technology (IT) systems used for data-centric computing and operational technology (OT) systems used to monitor events, processes and devices. Using the two-way flow of electricity and information, the smart grid builds an automated highly distributed energy delivery network, which supports real-time data exchange with the aim to balance supply and demand. Intrusion detection is a critical security service to protect smart grid systems, alerting the system operator for the presence of ongoing attacks. Hence, there has been lots of research conducted on intrusion detection, especially anomaly-based intrusion detection, to address security concerns in smart grids. However, some of that prior research is based on imbalanced data, which have much more data instances belonging to normal behaviors than to attack ones, and problems emerge when common approaches of pattern recognition are used for those imbalanced data; these approaches cause low detection rates for minority classes.

## V. Conclusion

Cyber-attacks might appear as natural events. Therefore, discriminating between malicious and non-malicious data in the communication system is difficult and challenging. An Intrusion Detection System (IDS) and even its evolution, the Intrusion Prevention System (IPS), can operate as a second line of defense in a communication network, by enhancing the operation of the encryption and authorization mechanisms. For instance, if a cyber attack bypasses the encryption and authorization mechanisms, the IDS or IPS can timely inform the security administrator or perform appropriate preventive countermeasures.

**References**

[1] Abdulhamit Subasi, Khloud Al-Marwani, Reem Alghamdi, Aisha Kwairanga, Saeed M. Qaisar, Malak Al-Nory, Khulood A. Rambo, "Intrusion Detection in Smart Grid Using Data Mining Techniques", IEEE 2018, pp. 1-6.

[2] Zakaria El Mrabet, Hassan El Ghazi, Naima Kaabouch, "A Performance Comparison of Data Mining
Algorithms Based Intrusion Detection System for Smart Grid", 2018, pp 1-6.

[3] R. Vijayanand, D. Devaraj, B. Kannapiran, "Support Vector Machine Based Intrusion Detection System with Reduced Input Features for Advanced Metering Infrastructure of Smart Grid", International Conference on Advanced Computing and Communication Systems, 2017, pp. 2-7.

[4] Vimalkumar K, Radhika N, "A Big Data Framework for Intrusion Detection in Smart Grids Using Apache Spark", IEEE, 2017, pp. 198-205.

[5] P. K. Gupta, N. K. Singh, V. Mahajan, "Intrusion Detection in Cyber-physical Layer of Smart Grid using Intelligent Loop Based Artificial Neural Network Technique", International Journal of Engineering,
2021, pp. 1250-1256.

[6] Imtiaz Ullah, Qusay H. Mahmoud, "An Intrusion Detection Framework for the Smart Grid", 30th Canadian Conference on Electrical and Computer, IEEE 2017, pp. 1-6.

[7] Fadi Salo, Mohammadnoor Injadat, Ali Bou Nassif, Abdallah Shami, Aleksander Essex, "Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review", IEEE 2018, pp. 56046-56058.

[8] Anzar Iqbal, Mohammad ummer chopan, Pooja, "Intrusion Detection in Smart Grid", International Journal of Innovative Science and Research Technology, 2019, pp. 54-57.

[9] Lida Haghnegahdar, Yong Wang, "A whale optimization algorithm-trained artificial neural network
for smart grid cyber intrusion detection", Neural Computing and Applications, Springer 2020, pp. 1-16.

[10] Panagiotis I. Radoglou-Grammatikis, Panagiotis G. Sarigiannidis, "Securing the Smart Grid: A Comprehensive Compilation of Intrusion Detection and Prevention Systems", IEEE Access, 2019, pp. 46595-46620.