**RESEARCH ARTICLE**

# A Study on Data Mining Techniques for Cyber Security

V.JeyaKumar* and S.Deepankumar

Assistant Professor, Department of Computer Science, KG College of Arts and Science, Coimbatore, Tamil Nadu, India.

Received: 23 May 2021 Revised: 30 May 2021 Accepted: 11 Jun 2021

***Address for Correspondence**
**V.JeyaKumar**
Assistant Professor,
Department of Computer Science,
KG College of Arts and Science,
Coimbatore, Tamil Nadu, India.
E.Mail: jeyakumar28@gmail.com

## ABSTRACT

Cyber security is that the world that deals with protecting from cyber terrorism. Cyber-attacks include access control violations unauthorized intrusions and denial of service also as insider threat. Security of a data or information system is its vital property especially today when computers are interconnected via internet. Because no system is often absolutely secure the timely and accurate detection of intrusions is important. For this purpose, Intrusion Detection Systems were designed. The IDS together with data processing can provide the safety with next level data processing is that the process of posing queries and extracting patterns, often previously unknown from large quantities of data using pattern matching or other reasoning techniques. This Paper gives the over view of the various data processing techniques which may be utilized in Cyber security for intrusion detection.

**Keywords:** Cyber security, Intrusion Detection System, Data Mining.

## INTRODUCTION

Cyber security cares with protecting computer and network systems from corruption owing to malicious software including Trojan horses and viruses. Data processing for cyber security applications for instance, anomaly detection techniques might be want to detect unusual patterns and behaviours. Data processing or mining is that the process of identifying patterns in large datasets. Data processing techniques are heavily utilized in research project also as in business, mostly to collect statistics and valuable information to improve customer relations and marketing strategies. In this paper, we specialise in data processing application for cyber security. To grasp the mechanism to be adopted so as to safeguard the computers and network, it's imperative to know the kinds of threats that endanger the cyber network.

## Cyber Security

Cyber security is about of rules and technologies which are mean to guard our systems, network, and data from unauthorized access, attacks, and unwanted interrupts. they're aim to take care of the confidentiality, integrity, and availability of data and knowledge management systems through various cyber defence systems. To safe the cyber infrastructure against possibly malicious threats, a rising collaborative effort between cyber security professionals and researchers from institutions, private industries, and various agencies has engaged in abusing and designing a spread of cyber defence systems. Cyber security systems are composed of network security systems and host security systems. Each of those has, firewall, antivirus software, and an intrusion detection system (IDS). IDS discover, determine, and identify unauthorized use, duplication, alteration, and destruction of data systems [1].The second line of cyber defence consists of reactive security solutions, like intrusion detection systems (IDSs). IDSs detect intrusions supported the knowledge from log files and network flow, in order that the extent of injury is often determined, hackers are often tracked down, and similar attacks are often prevented within the future. data processing or knowledge discovery (KDD) may be a method want to analyse data from a target source and compose that feedback into useful information. In cyber security data processing techniques are getting used to spot unsure conditions.

## Cyber Terrorism, Threats and Malicious Software

Now a days internet has allowed for a huge exchange of data. Thus, has created a cyber space during which terrorists can implement attack. Cyber-terrorism, consistent with the O' Leary (2010) is committed through the utilization of cyberspace or computer resources. This use of cyber space leads to there not being simply a physical threat of terrorism. Janczewski, & Colarik (2008) defines cyber terrorism as: "Cyber terrorism means pre-mediated, politically motivated attacks by sub national groups or clandestine agents or individuals against information and computer systems, computer programs, and data that leads to violence against non-combatant targets." Cyber Terrorism is one among the main threat to world now. Over recent decades, it's become apparent that our society is becoming increasingly information technology dependant.

For instance, of banking system. If surprise attack such a system and deplete accounts of funds, then the bank could lose millions or billions of dollars. Crippling the pc system many hours of productivity might be lost, which is ultimately like money loss. Even an easy power outage at work could cause several hours of productivity loss which ends in loss. Therefore, it's imperative that our data system might be secured. Threats can occur from outside or inside a corporation. Malicious software are the codes or procedures or programs which are mean to damage the systems, networks, clients and servers, databases. The most common sorts of this are virus, warms, trojan horses. Intruders try to tap into network and get vital information. It is often a person's or malicious software set by humans.

## Data Mining

In general, it's a process that involves analysing information, predicting future trends, and making proactive, knowledge-based decisions supported large datasets. It is a process that involves scanning the knowledge, predicting future trends, and making the knowledge-based decisions supported large datasets. Data mining consistent with Silltow (2012) automates the detection of relevant patterns during a database, using defined approaches and algorithms to seem into current and historical data which will then be analysed to predict future trends.

While the term data processing is typically treated as a synonym for Knowledge Discovery in Databases (KDD), it's actually only one of the steps during this process. The main goal of KDD is to get useful and sometimes previously unknown information from large sets of knowledge. Due to the supply of huge amounts of knowledge in cyber infrastructure and increasing number of cyber criminals attempting to realize unauthorized access to the information, there's need of capabilities to address the challenges of cyber security.

Data mining tools predict future trends and behaviours by reading through database for hidden patterns, learning these behaviours is vital, as they will identify and describe structural patterns which helps to get the knowledge on the idea of that data, and helps the organization to answer the questions that were too time-consuming perversely. Data mining application for cyber security is that the use of knowledge mining techniques to detect cyber threats. Data mining with the mixture of machine learning is being applied to problems areas like intrusion detection and auditing in cyber security and which is extremely effective technique. In recent years, many IT industry giants like Comodo, Symantec, and Microsoft have started using data processing techniques for malware detection.

### Data Mining Methods for Cyber Security

This segment labels the different Data Mining methods for cyber security.

### Association Rule

The association rule mining discovers the relationship among variables in database. Consider an example IF (P AND Q) THEN S. This rule implies that IF P and Q are present, then it is also presence of S. Association rules have metrics that tell how often a given relationship occurs within the data. Association Rule Mining was introduced by Agrawal et al. [2] as how to get interesting co-occurrences in supermarket data. It finds frequent sets of things (i.e., combinations of things that are purchased together in a minimum of N transactions within the database), and from the frequent items sets such as {A, B}, generates association rules of the form: $A \rightarrow B$ and/or $A \rightarrow B$.

### Clustering

Clustering is employed to assign the similar data object in groups called clusters in order that the objects in one cluster are more almost like one another than objects in other clusters. In simple word this process is employed to spot data items that have similar characteristics. Clustering [4] may be a set of techniques for locating patterns in high-dimensional unlabelled data. The main advantage of clustering for intrusion detection is that it can learn from audit data without requiring the supervisor to supply explicit descriptions of varied attack classes.

### The decision tree technique

The decision tree may be a tree like structure having leaves which represent the classification and branches which represent the conjunction of features that cause those classifications. Decision tree depends on if–then rules, but it doesn't require parameters and metrics. This simple and interpretable structure allows decision trees to unravel multi-type attribute problems. Decision trees also can manage missing values or noise data. However, they cannot guarantee the optimal accuracy that other machine-learning methods can. [5] The advantages of decision trees are simple implementation.

### The neural network

Neural Networks are inspired by the brain and composed of organized artificial neurons capable of certain calculations on their inputs [6]. The input data to the first layer activate the neurons of the network whose output is the input to the second layer of neurons in the network. Neural networks re long training times and are therefore more suitable for applications where this is often feasible.

IDS use two kind of Neural Networks, they are

- Multilayered Feedforward NN
- Kohonen’s Self-Organizing NN

These techniques are used to model complex relationships between inputs and outputs and to discover new patterns. The combination of Self organizing map and back propagation neural network supply a very efficient mean for detection of new intrusions.

### Data Mining in Malware Detection

Data mining is one of the wide method used today for detecting malware. While building a security app, software developers uses the data mining methods to improve the speed and quality of malware detection.

There are three strategies used to detect malwares:

- Anomaly Detection
- Misuse Detection
- Hybrid Detection

Anomaly Detection is the identification of infrequent actions or observations which raise uncertainties by differing significantly from the majority of the data. It contains demonstrating the normal behaviour of a system or network in order to identify deviations from normal usage patterns. Anomaly based detection can also detection the previous unknown attacks and use for defining the signature for misuse detectors. The main problem with anomaly detection is that any deviation from the normal, even if it is a legitimate behaviour, will be reported as an anomaly, thus producing a high rate of false positives.

Misuse Detection is also referred as a signature-based detection, identifies only known attacks based on examples of their signatures. It states to detection of attacks by observing for specific patterns, such as byte sequences in network traffic, or identified malicious instruction sequences used by malware.

Hybrid approach is a combination of anomaly and misuse detection techniques in order to rise the number of detected intrusions while reducing the number of false positives. It does not build any models, but in its place uses information from both harmful and clean programs to create a classifier – a set of rules generated by the data mining algorithm.

## CONCLUSION

In this paper we have studied the different data mining techniques for cyber security. It is a young interdisciplinary, drawing from areas such as database systems, data warehousing, statistics, machine learning, data visualization, information retrieval, and high-performance computing. Data mining have great potential in place of malware detection tool which allows you to analyse enormous sets of information and extract new knowledge from it. When determining the efficiency of the methods, there is not only one principle but several that need to be taken into account. Depending on a particular Intrusion Detection System some might be more important than other [9]. Another vital aspect data mining for cyber intrusion detection is the importance of the data sets for training and testing the systems. The main advantage of using data mining techniques for detection malicious software is the ability to identify the both known and zero day attacks.

## REFERENCES

1. A. Mukkamala, A. Sung, and A. Abraham, “Cyber security challenges: Designing efficient intrusion detection systems and antivirus tools,” in Enhancing Computer Security with Smart Technology, V. R. Vemuri,Ed. New York, NY, USA: Auerbach, 2005, pp. 125–163.

2. R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in Proc. Int. Conf. Manage. Data Assoc. Comput. Mach. (ACM), 1993, pp. 207–216.
3. H. Brahmi, B. Imen, and B. Sadok, "OMC-IDS: At the cross-roads of OLAP mining and intrusion detection," in Advances in KnowledgeDiscovery and Data Mining. New York, NY, USA: Springer, 2012, pp. 13–24.
4. K.Jain and R. C. Dubes, Algorithms for Clustering Data. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.
5. Sumeet Dua and Xian Du "Data Mining and Machine Learning in Cyber security"
6. K. Hornik,M. Stinchcombe, and H.White, "Multilayer feedforward networks are universal approximators," Neural Netw., vol. 2, pp. 359–366, 1989.
7. Bolton, R. and D. Hand, Statistical fraud detection: A review. Statistical Science 17 (3),pp. 235-255, 2002.
8. https://www.apriorit.com/dev-blog/527-data-mining-cyber-security

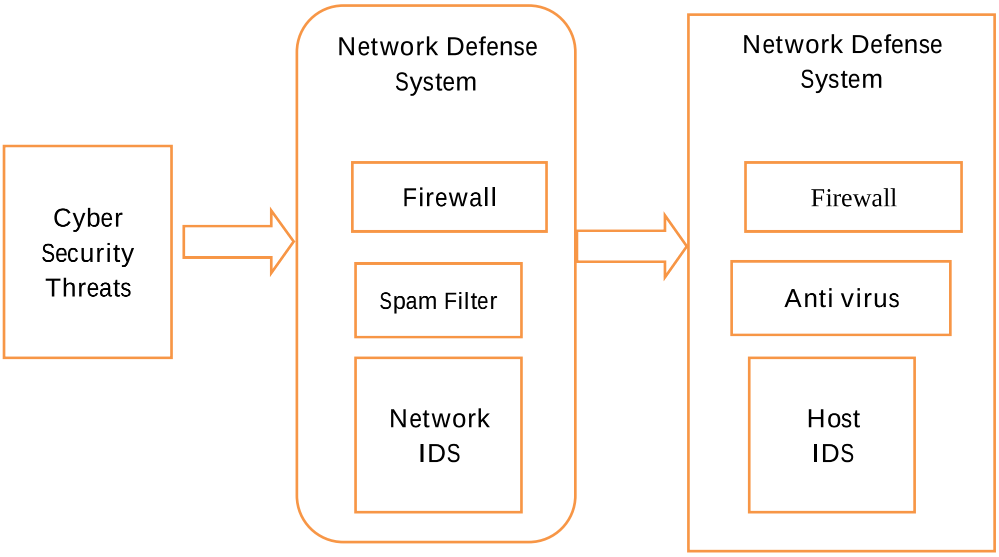

**Fig. 1. Conventional cyber security**