ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 8, September 2018

International Conference on Emerging Trends in Science, Technology and Mathematics (ICETSTM-2018)"

22nd August 2018

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

A Comparative Analysis on Security Using Data Mining Techniques

V.Karthikeyani[1], P.Kanagavalli[2]

Department of Computer Science, NKR Government Arts and Science College for Women, Namakkal Tamilnadu, India [1]

Research Scholar,Department of Computer Science, Periyar University, Salem Tamilnadu, India[2]

ABSTRACT: Computers are vulnerable to threats and are required to be handled. Security needs to be handled at various levels as system, data, network and software. In spite of the techniques as cryptography, steganography, the security mechanisms are provided by the techniques as data mining, soft computing, biometrics, artificial intelligence and neural network.This paper discusses the threats and challenges in security. Finally, a survey on the security algorithms is made.

KEYWORDS*:* Network Security,Data Mining Techniques, Machine learning, Threats.

I. INTRODUCTION

Security is the key issue in this digital world. The computer and information are still prone to attack. The attack are of different types as denial of service (DoS), Phishing attacks, malware attacks etc., Computer security is the protection of computer systems from theft or damage to their hardware, software or electronic data, as well as from disruption or misdirection of the services they provide [1].The information in the network or in internet is required to be handled in a secured way. There were various approaches prevailing to handle the security as data mining, soft computing, artificial intelligence, neural networks etc., and this is also an ongoing research area. Techniques like cryptography, steganography are widely used in network security. While passing information over the network the cipher text are transmitted. In this approach, the text is encrypted and transmitted. At the receiving end the text is decrypted and used. In the secured sites, while establishing the secured session, the algorithms like RSA, MD5 are widely used. In steganography the text information is hidden in the images and transmitted over the internet. Authorization and authentication issues should also be considered to provide security. Even the physical features such as iris, finger, face and DNA are considered for security in biometrics.

In this paper, the section II discusses about the challenges in security and it deals with the types of security as system security, Network security, Software Security and Data Security. In section III, the approaches such as data mining, soft computing, neural networks, artificial intelligence and biometrics to handle the security are given. In section IV, the comparative study over the data mining techniques for network security are explained. The data mining approaches such as classification, clustering, machine learning, Kernal based Intrusion detection, signature-based Intrusion detection etc., are considered. In section V,the comparative study over the data mining techniques used for misuse and anomaly techniques. In section VI, the parameters considered in security such as detection rate and false prediction rate are considered. [10]Detection and hindrance of such aggressions loud intrusions typically depends on the talent potency of Intrusion Detection Arrangements (IDS).Misuse detection has also been used more generally to refer to all kinds of computer misuse. Finally, in section VII the conclusion is given.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 8, September 2018

International Conference on Emerging Trends in Science, Technology and Mathematics (ICETSTM-2018)"

22nd August 2018

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

II. CHALLENGES IN SECURITY

The security is broadly classified into four types as System Security, Network Security, Software Security, Data Security and it is shown in figure 1.

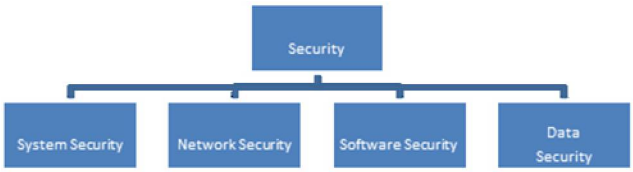

Figure 1. Types of Security

A. System Security

The system refers to the standalone computer at home or in an office which is not connect to a network. Security threat is very low in this case of computers. This may be misused by a person who may be a visitor to the place. This can be handled by setting up a password. Auto password lock is essential and enough for these types of systems. The BIOS (Basic Input-Output System) password can also be given to ensure the secured access.

B. Network Security

a) Local Network Security:

The network here refers to the local area network of any size but not connected internet. The security threat is medium in this case. In addition to setting up the password, the security measures should be taken. Proper access rights should be given by the administrator depending on the user level.

b) Cyber Security:

The security threat in cyber security is very high. The intruders may hack the system. Research work is going on in this field, to provide a secured environment.

C. Software Security

The application software is also prone to vulnerabilities. In [2], the comparative study of various algorithms for software vulnerabilities is discussed. The machine learning algorithms and data mining techniques are used for software vulnerability detection. Software penetration testing is done to ensure the software security.

D. Data Security

The data stored in the system should also be secured. The sensitive data can even be encrypted and stored. The unauthorized access of data should be handled.

III. APPROACHES TO HANDLE SECURITY

There are various approaches available and under research to handle security.

B. Data Mining

Data mining refers to extraction of relevant information from a large data set. Data mining is used in various applications. Security issues are also handled by applying the data mining techniques. Classification, feature selection, and clustering approaches are used to handle the security issues in the network. By applying the data mining steps the security is ensured in the network. The sequence of steps followed in the data mining approach are data collection, data cleansing, data analysis and data interpretation and it is depicted in figure 2.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 8, September 2018

International Conference on Emerging Trends in Science, Technology and Mathematics (ICETSTM-2018)"

22nd August 2018

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

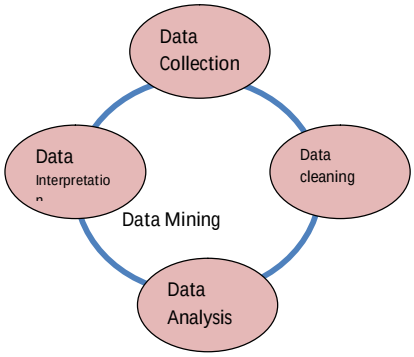

Figure 2. Data Mining Process.

C. *Soft Computing*

Soft computing approaches can be used when there is no proper sequence of steps are available to arrive a result. The soft computing algorithms such as Ant Colony Optimization(ACO), fuzzy logic, Genetic algorithms are used for intrusion detection.

D. *Artificial Intelligence and Neural Network*

AI (*artificial intelligence*) is the simulation of human *intelligence* processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions) and self-correction [8]. Various machine learning techniques and ANN (Artificial Neural Network) are used to handle security issues.

E. *BioMetrics*

Biometric security is a *security* mechanism used to authenticate and provide access to a facility or system based on the automatic and instant verification of an individual's physical characteristics [9]. Face recognition, iris scanning, finger prints, DNA are considered as input for biometric security. In which used for identifying intrusions, signature-based and [11] anomaly-based intrusions detection.

IV. COMPARATIVE STUDY OVER APPLICATION OF DATAMINING TECHNIQUES FOR SECURITY

Data mining techniques are widely used for security. The table 1. Lists the algorithm used.

Table 1. Data Mining Techniques for security

S.No	Algorithm Proposed	Author
1	Classification Tree and SVM	Ektefa et al
2	Machine Learning Algorithms	Buczak et al
3	Behavioural Pattern	Goliwale et al
4	Classification algorithms such as J48.	Aljawarneh et al
5	Clustering Algorithm	Portnoy et al
6	Kernal based Intrusion detection algorithm	Sree, Shivalingari Bhanu et al
7	Misuse detection approach	Helman,paul,Liepins et al

Ektefa et al. discusses the application of classification tree and SVM (Support Vector Machine) for intrusion detection. The detection rate and false prediction rate are considered [3].In 4, the author discusses about the application of machine learning algorithms and data mining methods for cyber intrusion detection. The cyber data sets are used for analysis. The anomaly detection approaches are proposed in [5]. The normal usage pattern is considered and whenever the malicious user enters, it is identified by the behavioural pattern. In [6], the author proposes an approach to detect the anomalies. Initially, the data is filtered using vote algorithm. Various classifiers such as J48, Meta pagging, random tree, REPTree, AdaBoostM1, DecisionStump and NaiveBayes are used to handle high false negative rate and

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 8, September 2018

International Conference on Emerging Trends in Science, Technology and Mathematics (ICETSTM-2018)"

22nd August 2018

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

low false positive rates. Portnoy et al, proposes the clustering approach for intrusion detection. The unlabeled data set is considered in this clustering [7]. [12] Schwaederle, Maria, et al. "Detection rate of actionable mutations in diverse cancers using a biopsy-free (blood) circulating tumor cell DNA assay." *Oncotarget* 7.9 (2016): 9707. Shivalingari Bhanu. [11]"Kernel Based Intrusion Detection Using Data Mining Techniques." (2018). Ref: Helman, Paul, Liepins, Gunar, and Richards, Wynette, "Foundations of Intrusion Detection,"[13]

V. MISUSE AND ANOMALY USING DATA MINING TECHINQUES

a) ***Misuse Detection***

Misuse detection is an approach to detecting attacks.In a misuse detection approach,abnormal system behaviour is defined first,and then all other behaviour is defined as normal.With misuse detection, anything not known is normal. An example of misuse detection is the use of attack signatures in an intrusion detection system. Misuse detection has also been used more generally to refer to all kinds of computer misuse

b) ***Anomaly detection***

Anomaly detection is an approach which utilizes the reverse defining normal system behaviour first and defining all other behaviour as abnormal.

VI. PARAMETERS CONSIDERED

A. ***Detection Rate***

The detection rate refers to the rate at which the malicious user or transaction identification.

B. ***False Prediction Rate***

The false prediction rate refers to the rate at which the legal user is wrongly classified to be intruder.

VII. CONCLUSION

In this paper, a comparative analysis over various security algorithms for network is considered by applying data mining techniques. The data mining techniques such as classification and clustering are considered. The parameters such as detection rate and false prediction rate are analyzed. In which for kernel based intrusion detection method. In which of the used as misuse and anomaly approaches are data mining techniques. Thus, this paper analyses the application of data mining techniques for security.

REFERENCES

[1] https://wikipedia.org/wiki/Computer_security
[2] Ghaffarian, Seyed Mohammad, and Hamid Reza Shahriari. "Software vulnerability analysis and discovery using machine-learning and data-mining techniques: a survey." ACM Computing Surveys (CSUR) 50.4 (2017): 56.
[3] Ektefa, Mohammadreza, et al. "Intrusion detection using data mining techniques." Information Retrieval & Knowledge Management,(CAMP), 2010 International Conference on. IEEE, 2010.
[4] Buczak, Anna L., and Erhan Guven. "A survey of data mining and machine learning methods for cyber security intrusion detection." IEEE Communications Surveys & Tutorials 18.2 (2016): 1153-1176.
[5] Goliwale, Paresh, et al. "Intrusion Detection System Using Data Mining." management 5.03 (2018).
[6] Aljawarneh, Shadi, Monther Aldwairi, and Muneer Bani Yassein. "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model." Journal of Computational Science 25 (2018): 152-160.
[7] Portnoy, Leonid, Eleazar Eskin, and Sal Stolfo. "Intrusion detection with unlabeled data using clustering." In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001. 2001.
[8] https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence
[9] https://www.techopedia.com/definition/6203/biometric-security.
[10] Ref: Helman, Paul, Liepins, Gunar, and Richards, Wynette, "Foundations of Intrusion Detection,"
[11] Sree, Shivalingari Bhanu. "Kernel Based Intrusion Detection Using Data Mining Techniques." (2018).
[12] Schwaederle, Maria, et al. "Detection rate of actionable mutations in diverse cancers using a biopsy-free (blood) circulating tumor cell DNA assay." Oncotarget 7.9 (2016): 9707.
[13] Ref: Helman, Paul, Liepins, Gunar, and Richards, Wynette, "Foundations of Intrusion Detection,"