Intelligent Systems and Computer Technology
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC200127

Cyber-Attacks Detection Using Novel IDDMS Framework

Gadekar Ganesh Bhivsen [a], Udayabhanu N P G[b], Dange Bapusaheb Jalindar[c], Vengatesan K[d] and Abhishek Kumar[e]
[a,c,d] *Dept of CSE, Sanjivani College of Engineering, Kopargaon, India*
[b] *Dept of CSE, Vignan Institute of Tech. and Science, Hyderabad, India*
[e] *Dept of Computer Science, Banaras Hindu University, Varanasi, India*

Abstract. Security of a data system is a significant property, particularly today when PCs are interconnected by means of the internet. Since no system can be totally secure, the opportune and precise detection of intrusions is essential. Cyber security is the region that manages shielding from cyber terrorism. Cyber-attacks incorporate access control infringement, unapproved intrusions, and disavowal of service just as insider risk. For this reason, IDS were planned. The IDS in the mix with DM can give security to the next level. DM is the way toward presenting inquiries and separating designs, frequently already ambiguous from huge amounts of data utilizing design coordinating or other thinking techniques. This Paper gives the IDDMS (Intrusion Detection with Data Mining system) Framework which is a mix of data mining techniques with the Intrusion detection system, this can be utilized in Cyber-security for accomplishing the next level of service.

Keywords. Intrusion detection system, Data mining techniques, cyber-attacks

1. Introduction

Identifying cyber-attacks without a doubt has become a major data issue. The space of cyber-security is innately a progressively evolving one. More up to date attacks, for example, multi-organize endeavours and zero-day attacks, can be altogether more differing than old attacks as far as specialized execution just as the hidden strategies themselves in the progressing astuteness race among attackers and protectors. As cyberattacks have developed and developed in complexity, cyber-attack detection techniques have likewise gotten significantly more modern, by checking a regularly expanding measure of various heterogeneous security occasion sources. [1] Presents a diagram upon different techniques under various detection models, specifically abuse detection and peculiarity detection, separately, and furthermore presents the level of attack coerce as a method for describing intrusion detection exercises. So to prevent attacks, consciousness of an attack is fundamental to having the option to respond and safeguard against attackers. Cyber resistances can be additionally improved by using security investigation to search for shrouded attack examples and patterns. This is obviously the inspiration for applying DM for cyber security. Data mining is the way toward extricating valuable and already unnoticed models or examples from huge data stores [2]. The objective of data mining process is to find designs that are tucked away among the gigantic arrangements of data. In this manner, data mining is likewise taken as information revelation. [3] Presents an outline on various data mining and machine

[1]Gadekar Ganesh Bhivsen, Dept of CSE, Sanjivani College of Engineering, Kopargaon, India
Email: bapudange@gmail.com

learning calculations being applied to abuse and anomaly detection. [4]Presents a dispersed design for data mining based ID that adventures the benefits of both abuse and anomaly detection methodologies.

2. Novel IDDMS Framework

Data mining based cyber-attack detection includes five general stages, as showed in Figure 1, that is, system checking and data catching by means of different sensors, organize/system/process logging and sniffing daemons/specialists, and security gadgets, data pre-preparing (e.g., normalisation, cleansing, filtering, etc.) at local data stores, event correlation and feature extraction (e.g., via Map Reduce, big data processing, and HDFS), data mining (clustering, dimensionality reduction, classification) to detect misuse or anomaly, interpretation and visualisations of mining results. Data mining based detection, when appropriately designed, has the ability to become focal sensory system of system. Data mining based detection can give some valuable determined capacities, for instance, continuous checking and episode management for security related occasions which are gathered from arrange, security gadgets, system, applications. It gives a work process which tracks and heighten the occurrence.

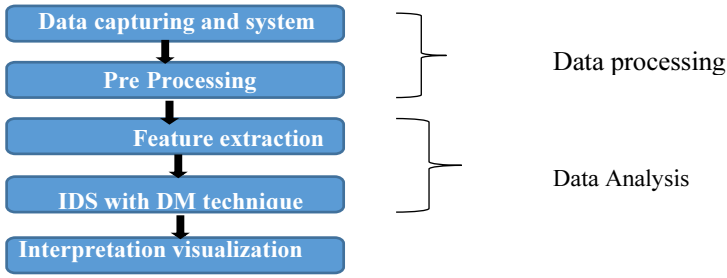

Figure 1., Proposed IDDMS

Data Analysis**:** This is main thing of data mining based cyber-attack detection. The gigantic data gathered by procurement layer is for incorporated stockpiling and examination in investigation level, in order to remove the significant data concerned. The gathered data is broke down in this progression to decide if the data is bizarre or not. Here, it might include highlight choice and danger connection. Highlight choice produces from huge dataset include vectors. Danger connection utilizes the computerized reasoning to figure out various logs and log passages to recognize attackers.

Response: Clearly all things considered, moves should be made because of the recognized attacks. Reaction can be set to be performed consequently. If there should arise an occurrence of manual investigation circumstances it very well may be done physically, which implies the last ascertainment by human managers for assurance about cyberattacks and relief and choice. Detection system cautions the system head that an attack has happened utilizing a few strategies like email, alert symbols and

perception techniques. Detection system can likewise stop or control attack by shutting system ports or executing forms.

Reduction of Dimensionality**:** Dimensionality decrease is the way toward diminishing the quantity of irregular factors viable. Two techniques for dimensionality decrease are highlight extraction and highlight determination. [7] Stated that a near numeric investigation on dimensionality decrease calculations related to detection techniques under abuse and anomaly detection models. Highlight extraction alludes to the mapping of the first high-dimensional data onto a lower dimensional space. Through element extraction, all the first highlights are consolidated into another decreased arrangement of highlights. Instances of highlight extraction calculations are ICA and PCA.

Table1: Training Model dataset

Kind of Attack	Name of Attack	No. of Instances
Probe	Portsweep	813
	SATAN	467
	IPSWEEP	372
U2r	Buffer overflow	11
	Rootkit	4
	Perl	2
	Load module	1
Dos	Teardrop	31
	POD	9
	Back	68
	Neptune	36728
	SMURF	85893
Normal		32570

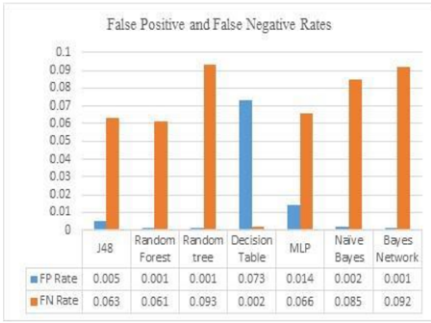

Figure 2. False Negative Rate and False Positive Rate.

3. Implementation

As to the KDD dataset there are 12 kind of attacks ordered into three gatherings (PROBE, DOS, and U2R) with various number of cases and event in data-set. After the KDD dataset imported to SQL server 2012. 107223 examples of records have been extricated and exhibited in Table 1 as preparing data. In light of a profound examination of KDD dataset the conveyance event of various sorts of attacks was spared. As such 69% of separated dataset present DOS attacks and 29% for typical traffic while 2% for different sorts of intrusions (PROBE and R2U).

WEKA (Waikato Environment for Knowledge Analysis) is a machine learning tool is written in JAVA. It is an open-source instrument and accessible for nothing. The numerical arrangement models showing up in this paper are given by the WEKA tool compartment. The Most well-known machine learning classifiers are utilized in this analysis (Bayes Network, Random woods, J48, Random Tree, Multilayer Perceptron (MLP), Decision Table, and Naive Bayes). In view of 107223 occurrences of records, it was effective to make the preparation models for all the chose machine learning classifiers. All the contemplated models are arranged and thought about for a thorough investigation of machine learning classifiers' proficiency.

As indicated by Figure.2, which delineates the FP and FN execution parameters, it tends to be reasoned that the arbitrary tree classifier accomplished the most elevated FN rate of 0.093. Consequently, there is an enormous number of attacks named ordinary packet. Despite what might be expected, the choice table classifier is accomplished at the least FN rate of 0.002. In a similar time, the choice table classifier came to the most elevated FP rate of 0.073 and that implies there is huge amount of ordinary packet named attack packets.

4. Conclusion

Data mining has extraordinary potential as a malware detection tool. It enables you to break down colossal arrangements of data and concentrate new information from it. While determining the viability of the strategies, there isn't just a single foundation yet a few that should be considered. Contingent upon a specific IDS some may be a higher priority than others. Another essential perspective Data mining for cyber intrusion detection is the significance of the data sets for preparing and testing the systems. The fundamental advantage of utilizing data mining techniques for distinguishing malignant programming is the capacity to recognize both known and unknown attacks. In any case, since a formerly obscure however real movement may likewise be set apart as conceivably deceitful, there's the likelihood for a high rate of false positives.

References

[1] S. Noel, D. Wijesekera and C. Youman (2002). Modern intrusion detection, data mining,and degrees of attack guilt. In: D. Barbará and S. Jajodia (eds.), Applications of Data Mining in Computer Security, series Advances in Information Security, Volume 6, Springer Science+Business Media New York, 2002, pp. 1-31.

[2] W. Lee and S. J. Stolfo (2000). A framework for constructing features and models for intrusion detection systems. Information and System Security, Vol. 3, No. 4, pp. 227-261.

[3] V.D.Ambeth Kumar, V.D.Ashok Kumar, Dr.S.Malathi and P.Jagaeedesh, (2014) "Intruder Identification using Footprint Recognition with PCA and SVM Classifiers" for the International Journal of Advanced Materials Research Vols.1345, PP 984-985 (2014) pp 1345-1349. [DOI:10.4028/www.scientific.net/AMR.984-985.1345]

[4] I. Brahmi, et al. (2012). Towards a multiagent-based distributed intrusion detection system using data mining approaches. In: L. Cao, et al. (eds), Agents and Data Mining Interaction (ADMI'2011), Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Vol 7103, pp. 173-194.

[5] G. P. Tadda and J. S. Salerno (2010). Overview of cyber situation awareness. In: S. Jajodiaet al, (eds), Cyber Situational Awareness-Issues and Research, vol. 46, Springer, 2010, pp. 15-35.

[6] H. Tianfield (2016). Cyber security situational awareness. In: Proceedings of 2016 IEEE International Conference on Smart Data (SmartData'2016), Chengdu, China, 15-18 December 2016, pp. 782-787

[7] Ambeth Kumar, V. D., Malathi, S., Venkatesan, R., Ramalakshmi, K., Vengatesan, K., Ding, W., & Kumar, A. (2019). Exploration of an innovative geometric parameter based on performance enhancement for foot print recognition. Journal of Intelligent & Fuzzy Systems, 1–16. https://doi.org/10.3233/jifs-190982

[8] Singh, V. K., Singhal, A., Rai, K. N., Kumar, A., & Dwivedi, A. N. D. (2019). Randomized key-based gmo-bcs image encryption for securing medical image. International Journal of Recent Technology and Engineering. https://doi.org/10.35940/ijrte.C4453.098319

[9] Rai, A. K., Agarwal, S., & Kumar, A. (2018). A novel approach for agile software development methodology selection using fuzzy inference system. Proceedings of the International Conference on Smart Systems and Inventive Technology, ICSSIT 2018. https://doi.org/10.1109/ICSSIT.2018.8748767

[10] Kesavan, S., Kumar, E. S., Kumar, A., & Vengatesan, K. (2019). An investigation on adaptive HTTP media streaming Quality-of-Experience (QoE) and agility using cloud media services. International Journal of Computers and Applications. https://doi.org/10.1080/1206212X.2019.1575034

[11] Rai, A. K., Agarwal, S., Khaliq, M., & Kumar, A. (2019). Quantitative analysis of development environment risk for agile software through machine learning. International Journal of Recent Technology and Engineering. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-85067942900&partnerID=MN8TOARS

[12] Kumar, A., Vengatesan, K., Rajesh, R., Parthibhan, M., & Singhal, A. (2018). Review of gene subset selection using modified k-nearest neighbor clustering algorithm. Proceedings of the International Conference on Smart Systems and Inventive Technology, ICSSIT 2018. https://doi.org/10.1109/ICSSIT.2018.8748667

[13] Vengatesan, K., Kumar, A., Naik, R., & Verma, D. K. (2019). Anomaly based novel intrusion detection system for network traffic reduction. Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018. https://doi.org/10.1109/I-SMAC.2018.8653735

[14] Daniel S. Berman, Anna L. Buczak and Jeffrey S. Chavis and Cherita L. Corbett (2019). A Survey of Deep Learning Methods for Cyber Security, journal of MDPI, vol 10, p. 110-122.

[15] Dhanabal. E, Harish G, YogaDinesh S, V. D. Ambethkumar, M. Rajendiran .Cryptographic Enhanced Shared Data On Conserve Encipher Text Update In Cloud Computing .Advances in Natural and Applied Sciences. 11(6); p:145-151, 2017.