International Journal of Engineering Research and Science & Technology

ISSN : 2319-5991
Vol. 3, No. 4
November 2014

www.ijerst.com

Email: editorijerst@gmail.com or editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 3, No. 4, November 2014

Research Paper

A SURVEY ON INTRUSION DETECTION IN NETWORK TRAFFIC USING DATA MINING TECHNIQUES

Shruti Yadav[1]* and Gireesh Kumar Dixit[1]

Corresponding Author:* ***Shruti Yadav* ✉ *shrut.yadav@gmail.com*

Today it is exceptionally essential to give an abnormal state security to secure profoundly touchy and private data. The Intrusion Detection System (IDS) is a vital innovation in Network Security. These days scientists have intrigued on interruption discovery framework utilizing data mining strategies as a guileful aptitude. IDS is a product or fittings gadget that arrangements with assaults by gathering data from an assortment of frameworks and system sources, then breaking down manifestations of security issues. This paper incorporates a diagram of interruption identification frameworks and acquaints the peruser with some central ideas of IDS technique. We likewise examine the essential interruption location procedures. A noteworthy test in giving a viable guard component to a system edge is being able to identify interruptions and execute countermeasures. Parts of the system border protection fit for distinguishing interruptions are alluded to as IDS. In this paper in the wake of concentrating on the exploration done in the field, we proposed a model in which first measurement decrease method is connected and then we accentuate data mining algorithms to actualize IDS by the help of one of the improved data mining algorithm.

Keywords: Data mining, KDD, Intrusion Detection System, Traffic analysis, etc.

INTRODUCTION

With the fast improvement of hardware and data engineering, machine system assumes a critical part in our everyday lives. Notwithstanding, more individuals use it as an apparatus for wrongdoing, and it has brought incredible misfortunes to numerous organizations and nations. So keeping these demonstrations is on the highest point of our plan. An interruption into a machine framework is any movement that damages framework honesty, secrecy, or information availability. So as to meet this test, Intrusion Detection System (IDS) is continuously intended to secure the accessibility, privacy and uprightness of discriminating organized data frameworks (Guanguin *et al.*, 2007; Ya *et al.,* 2007).

There are two types of intrusion detection techniques; they are misuse detection and

[1] Department of Computer Science & Engineering, MPM College, Bhopal.

anomaly detection. Misuse detection discovers attacks based on the patterns extracted from known intrusions. Anomaly detection detects attacks based on the significant deviations from the established profiles of normal activities (Zhang and Zulkermine, 2008). Intrusion detection techniques using data mining as an important application area to analyze the huge volumes of audit data and realizing performance the optimization of detection rules (Nguyen, 2008).

In the conventional IDS focused around principle identification, interruption mode is normally predefined by the security masters. The focal point of this methodology is that the standards could be planned to distinguish particular assaults in subtle element. Subsequently, it is ensured to catch known assaults and produce few false cautions. Notwithstanding, confronting with expanding and changing system information stream, it is implausible to find different interruption modes in time. The irregularity discovery is displayed because of this reason. New assaults might be recognized in time, which the framework has never seen previously as they veered off from typical conduct. On the other hand, current peculiarity recognition plans still experience the ill effects of a high rate of false cautions. In this way, right now IDS have a lot of people false alerts and repetition cautions. Therefore, understanding and taking care of IDS alerts get to be much more troublesome (Hu *et al.,* 2008).

LITERATURE SURVEY

Sumaiya Thaseen *et al.* (2013) discussed their research by aiming towards the main goal which is to evaluate eight tree based classifier algorithms to classify network events based on supervised discretization and feature selection. They summarized the experiments conducted using NSL-KDD data set and explored the models based on performance and error metrics resulting in higher accuracy and decreased resource utilization. There study shows that classification model integrated with discretization and feature selection method results in better accuracy, error rate and reduced false alarm rate. Considering the advantage and simplicity of Random Tree model, it can also be applied for dependent attributes such as NSL-KDD intrusion detection dataset.

Anazida Zainal *et al.* (2008) in paper has talked about the Efficiency is one of the real issues in interruption location. Wastefulness is frequently ascribed to high overhead and this is brought about by a few reasons. The motivation behind the paper is to address the issue of ceaseless location by presenting activity observing system. In movement checking, another distinguishment ideal model is proposed in which it minimizes unnecessary distinguishment. In this manner, the motivation behind movement observing is two-folds; to lessen measure of information to be perceived and to evade unnecessary distinguishment. For this Adaptive Neural Fuzzy Inference System and Linear Genetic Programming to structure group classifiers that demonstrates a little change utilizing the gathering methodology for Dos and R2l classes (assaults).

Zhai *et al.* (2010) in paper has examined that Id3 calculation was an excellent order of information mining. It generally chose the property with numerous qualities. The characteristic with numerous qualities wasn't the right one, it would made shortcoming caution and exclusion alert. To this blame, an enhanced choice tree calculation was proposed. The choice tree was made after the information gathered arranged

effectively. With the assistance of utilizing decision tree calculation, it demonstrates the most extreme assaults furthermore builds the caution level after changed the choice tree.

Jorge Blasco *et al.* (2010) in paper has mulled over that one of the focal zones in system interruption location is the means by which to construct viable frameworks that can recognize typical from nosy movement. To evade the visually impaired utilization of GP, it gives the pursuit by method for a wellness capacity focused around late advances on IDS assessment. For the test work utilization of a well-known dataset (i.e., KDD- 99) that has turned into a standard to analyze research in spite of the fact that its downsides. Comes about obviously demonstrate that a savvy utilization of GP gives better correctness furthermore analyze the Hit rate and False Rate to discover the quantity of assaults.

Ahmed Youssef *et al.* (2011) in paper has considered that Intrusion recognition has turned into a basic part of system organization because of the incomprehensible number of assaults per severingly undermine our machines. Conventional interruption discovery frameworks are restricted and don't give a complete answer for the issue. Then again, by and large, they neglect to recognize noxious practices (false negative) or They fire alerts when nothing wrong in the system (false positive). For this mix of Data Mining Techniques and Network conduct investigation were connected and conquer the confinements of customary Intrusion Detection System.

Mohd. Junedul Haque *et al.* (2012) in paper has said that the Intrusion Detection framework is a dynamic furthermore driving secure engineering to bargain the privacy, trustworthiness, accessibility, or to side step the security systems of a system. The fundamental piece of Intrusion Detection Systems (Idss) is to produce tremendous volumes of alerts. The fascinating alerts are constantly blended with undesirable, non-intriguing and copy cautions. For this Data mining calculation, K means bunching, Distributed IDS are connected to enhance the recognition rate and abatement the false alert rate.

Joshi *et al.* (2013) in paper has given that the enormous development in data engineering, system security is one of the testing issue thus as Intrusion Detection framework (IDS). The customary IDS are not able to oversee different recently emerging assaults. To defeat this sort of issue Data Mining systems, Feature Selection, Multiboosting were connected. With information mining, it is not difficult to distinguish legitimate, valuable and justifiable example in extensive volume of information. Gimmicks are chosen utilizing twofold classifiers for more exactness in each one sort of assault. Multiboosting is utilized to diminish both the change and predisposition. Subsequently the proficiency and correctness of Intrusion Detection framework are expanded and security of system so is likewise upgraded.

INTRUSION DETECTION SYSTEM

IDS are one of several most effective, expanding systems from the safety measures room, seeing that safety measures with system, method can be a massive matter in addition to for this function invasion detection notion enter lifetime. The industry is currently loaded largely simply by rule-based IDS options looking on sensing currently

acknowledged attacks simply by examining targeted visitors flow in addition to in search of acknowledged signatures. This specific truth demands these kinds of IDS of being below frequent building, updating in addition to changing invasion signatures in addition to needing to pay a significant fiscal sum for assisting. On the other hand, you are able to utilize anomaly based IDS options sensing not just acknowledged attacks, but unknown attacks in addition to educating system technical engineers with regards to feasible system troubles or maybe aiding these phones troubleshoot them. There is no apparent response which often the solution is more preferable when they have their advantages and disadvantages, but there's an opportunity to put the particular rule -based IDS options available as if these were anomaly based. This specific report details feasible ways of studying system targeted visitors dependant on widely recognized altered kdd99 information fixed or maybe NSL-KDD information fixed in addition to using some detailed fact, any safety measures a flag is harmonized while using the freshly forthcoming in addition to deviations is regarded as seeing that anomalous behavior in addition to these kinds of technique anomalous embarrassing activity is found.

Intrusion Detection Provides

1. Keeping track of along with considering both users along with technique.
2. Things to do a couple of inspecting technique adjustments along with vulnerabilities.
3. Accessing technique along with data file honesty.
5. Power to acknowledge patterns normal connected with episodes.
6. Analysis connected to the unnatural action structure.
7. Pursuing user plan infringement.

Misuse Detection

Misuse detection is also known as signature-based or knowledge-based systems. They follow the same principle as most anti-virus software and rely on the knowledge accumulated about previous attacks and vulnerabilities to detect intrusion attempts. Misuse detection systems compare current activities of the host or the network monitored with "signatures" of known attacks. If the current activities match any of the known signatures, an alarm is triggered.

Advantages and Limitations

Low Rate of False Alarms: The main advantage of misuse detection systems is their ability to detect known attacks and the relatively low false alarm rate when rules are correctly defined. It is important to note that, as said above, the signatures which are used in rules must be as specific as possible to prevent false alarms.

Only Known Attacks Detection: The foremost drawback of misuse detection systems is their complete inability in detecting unknown attacks.

Anomaly Detection

Anomaly detection systems are also known as behavior-based systems. They rely on the fact that intrusions can be detected by observing deviations from the expected behavior of the system monitored. These "normal" behaviors can either correspond to some observations made in the past or to some forecasts made by various techniques.

Everything that does not correspond to this "normal" pattern will be flagged as anomalous.

Therefore, the core process of anomaly detection is not to learn what is anomalous, but to learn what is normal or expected. Learning the normal behaviors and detecting deviations.

The process of learning the normal behavior of a system or a network and detecting deviations from these behaviors is an active area of research ever since the idea was first raised by Denning. Most of the methods currently investigated fall in any of the following five categories:

Statistics-based detection, Payload-based detection, Protocol-based detection, Graph-based detection and Machine-learning based detection.

Advantages and Limitations

Unknown Attacks Detection: The main advantage of anomaly detection systems is that, contrary to misuse detection systems, they can detect unknown or novel attacks. They do not rely on any a priori knowledge concerning the intrusions. It is also important to note that anomaly detection systems have not for the main purpose to replace misuse detection systems. The very good efficiency of misuse systems in detecting known attacks makes them a perfect complement to anomaly detection systems.

High Rate of False Alarms: Two factors may lead to a very high rate of false alarms or to a very poor accuracy of anomaly detection systems.

DATA MINING

Establishments of Data Mining and Knowledge Discovery" contains the most recent results and new bearings in Data mining exploration. Data mining, which coordinates different innovations, including computational insights, database and learning administration, machine adapting, delicate registering, and detail, is one of the quickest developing fields in software engineering. Albeit numerous data mining methods have been created, further advancement of the field obliges a nearby examination of its establishments.

Data mining determines its name from the likenesses between hunting down important data in an expansive database - for instance, discovering co-connection among the information present there. Given databases of sufficient size and quality, Data mining engineering can create new open doors by giving these capacities:

- Automated forecast of patterns and practices. Data mining robotizes the procedure of discovering prescient data in expansive dataset. Addresses that generally obliged broad active examination can now be addressed specifically from the information — rapidly.
- Automated disclosure of formerly obscure examples. Data mining apparatuses clear through dataset and distinguish formerly shrouded examples in one stage.

CONCLUSION

Internet and local networks have become everywhere. So organizations are increasingly employing various systems that monitor IT security breaches because intrusion events are growing day by day. This paper describes the different types of intrusion detection system and highlights techniques of intrusion detection. It is demonstrated in the paper that there will be a few interruption discoveries devices with contending gimmicks which are produce for location of assaults like known assaults and obscure assaults furthermore directed and Un-regulated methodologies are utilized to discover the assaults. Unsupervised learning systems can locate the interruptions that have not been adapted

by directed methodologies. Combining more than one data mining algorithms may be used to eliminate disadvantages of one another.

REFERENCES

1. Ahmed Youssef and Ahmed Emam (2011), "Network Intrusion Detection using Data Mining and Network Behavior Analysis", *International Journal of Computer Science & Information Technology (IJCSIT),* Vol 3, No. 6.
2. Anazida Zainal, Mohd Aizaini Maarof and Siti Mariyam Shamsuddin (2008), "Data Reduction and Ensemble Classifiers in Intrusion Detection" in 2008 IEEE.
3. Guangjun Song, Zhenlong Sun, Xiaoye Li (2007), "The Research Of Association Rules Mining And Application In Intrusion Alerts Analysis", *Second International Conference On Innovative Computing, Information and Control* (ICICIC 2007), pp. 567.
4. Guangqun Zhai and Chunyan Liu (2010), "Research and Improvement on ID3 Algorithm in Intrusion Detection System" in 2010 IEEE
5. Hu Zhengbing, Li Zhitang and Wu Junqi (2008), "A Novel Network Intrusion Detection (NIDS) Based On Signatures Search Of Data Mining", 2008 Workshop On Knowledge Discovery And Data Mining, pp. 10-16.
6. Joshi S A and Varsha S Pimprale (2013), "Network Intrusion Detection System (NIDS) based on Data Mining", *International Journal of Engineering Science and Innovative Technology* (IJESIT), Vol. 2, Issue 1.
7. Jorge Blasco, Agustin Orfila and Arturo Ribagorda (2010), "Improving Network Intrusion Detection by Means of Domain-Aware Genetic Programming", DOI 10.1109/ ARES.2010.53 in IEEE 2010.
8. Mohd. Junedul Haque, Khalid.W. Magld and Nisar Hundewale (2012), "An Intelligent Approach for Intrusion Detection Based on Data Mining Techniques" in 2012 IEEE.
9. Nguyen H A and Choi D (2008), "Application Of Data Mining To Network Intrusion Detection: Classifier Selection Model", In *Challenges For Next Generation Network Operations And Service Management,* Springer Berlin, Heidelberg.
10. Sumaiya Thaseen and Ch. Aswani Kumar (2013), "An Analysis of Supervised Tree Based Classifiers for Intrusion Detection System", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME) February 21-22.
11. Ya-Li Ding, Lei Li and Hong-Qi Luo (2009), "A Novel Signature Searching For Intrusion Detection System Using Data Mining", Machine Learning and Cybernetics", 2009 International Conference On Vol. 1, pp.122-126.
12. Zhang J and Zulkernine M (2008), "Random-Forest-Based Network Intrusion Detection Systems", I*EEE Transactions On System Man And Cybernetics,* Vol. 38, pp. 649-659.

International Journal of Engineering Research and Science & Technology
Hyderabad, INDIA. Ph: +91-09441351700, 09059645577
E-mail: editorijlerst@gmail.com or editor@ijerst.com
Website: www.ijerst.com