A Survey on Intrusion Detection Technique over the Web Data

Bhagwat P. Dwivedi, Shiv Kumar, Babita Pathik

Abstract: ***The intrusion detection systems (IDSs) generate large number of alarms most of which are false positives. Fortunately, there are reasons for triggering alarms where most of these reasons are not attacks. In this work, a new data mining technique has been developed to group alarms and to produce clusters. we have monitored a paper IDS over web mining – up approach which is efficient and determined to visualized the intrusion data and optimize according to the user requirement and monitored the data efficiently, here we would like to further enhance research work on analyzing and using the entropy data as input and to use them in technique to visualize and to optimize according to the user requirement in the web entropy visualization.***

Keywords: Network intrusion, web mining scenario, web intrusion data, Data Mining Algorithms.

I. INTRODUCTION

The purpose of Web mining is to develop methods and systems for discovering models of objects and processes on the World Wide Web and for web-based systems that show adaptive performance. Web Mining integrates three parent areas: Data Mining (we use this term here also for the closely related areas of Machine Learning and Knowledge Discovery), Internet technology and World Wide Web, and for the more recent Semantic Web. The World Wide Web has made an enormous amount of information electronically accessible. The use of email, news and mark-up languages like HTML allows users to publish and read documents at a world-wide scale and to communicate via chat connections, including information in the form of images and voice records. The HTTP protocol that enables access to documents over the network via Web browsers created an immense improvement in communication and access to information. For some years these possibilities were used mostly in the scientific world but recent years have seen an immense growth in popularity, supported by the wide availability of computers and broadband communication.[1]
Over the last decade, our society has become technology dependent. People rely on computer networks to receive news, stock prices, email and online shopping. The integrity and availability of all these systems need to be defended against a number of threats. Amateur hackers, rival corporations, terrorists and even foreign governments have the motive and capability to carry out sophisticated attacks against computer systems.

Revised Version Manuscript Received on December 16, 2016

Bhagwat P. Dwivedi, M.Tech. Scholar, Department of Computer Science and Engineering, Lakshmi Narain College of Technology Excellence, R.G.P.V. University, Bhopal (M.P)-462021, India.

Dr. Shiv Kumar, Professor & Head, Department of Computer Science and Engineering, Lakshmi Narain College of Technology Excellence, R.G.P.V. University, Bhopal (M.P)-462021, India.

Babita Pathik, Assistant Professor, Department of Computer Science and Engineering, Lakshmi Narain College of Technology Excellence, R.G.P.V. University, Bhopal (M.P)-462021, India.

Therefore, the field of information security has become vitally important to the safety and economic well being of society as a whole. [3]

In this struggle to secure our stored data and the systems, IDS can prove to be an invaluable tool, where its goal is to perform early detection of malicious activity and possibly prevent more serious damage to the protected systems. By using IDS, one can potentially identify an attack and notify appropriate personnel immediately or prevent it from succeeding, so that the threat can be contained. IDS can also be a very useful tool for recording forensic evidence that may be used in legal proceedings if the perpetrator of a criminal breach is prosecuted. Nowadays, there exists an extensive growth in using Internet in social networking (e.g., instant messaging, video conferences, etc.), healthcare, e-commerce, bank transactions, and many other services.
A complete process for the IDS using data mining is demonstrated in figure 1.

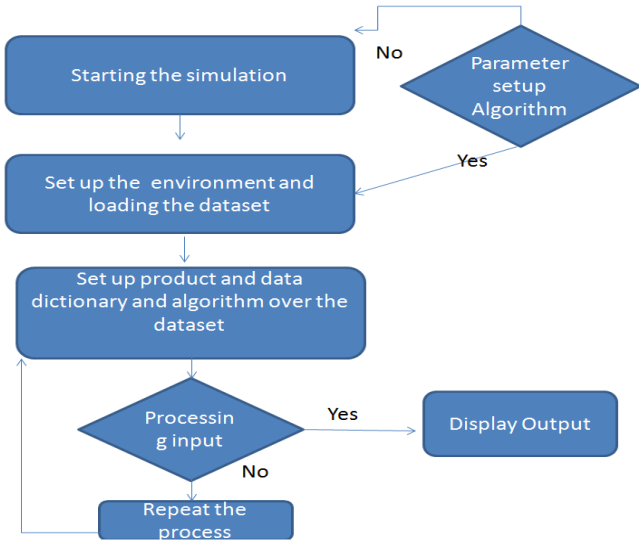

Figure 1: IDS Process Block Diagram.

The above diagram 1 shows the process of complete scenario.

II. LITERATURE SURVEY

A. Zhan Jiuhua[1]

Analyzed recent IDS models, the development of IDS (Intrusion Detection System), and the current and gives a brief introduction to DM (Data Mining) technology. Presented a framework of IDS based on data mining for resolving the current problems IDS is facing. The system that performs anomaly detection can detect intrusions known and unknown, reduce omissions and misstatements, improve accuracy and speed of intrusion detection

and has good adaptive capacity and scalability.

Problem:
Existing current paper algorithm outperforms the omission of available data and finding wrong accuracy over the simulation.
Solution:
A unique and high accuracy algorithm can be presented to propose the new model.

B. Changxin Song, Ke Ma [3]

Internet technology has developed rapidly and both software system and hardware equipment have improved greatly in recent years. However, Internet brings people not only convenience but also great potential threats. Facts show that potential safety hazards exist from the emergence of internet. As a kind of effective information security safeguard measure, intrusion detection makes up for the defects of traditional security protection techniques. As a kind of effective data analysis method, data mining is introduced into intrusion detection systems.
Problem:
Existing work on previous technique doesn't come up with effective pre-processing scheme thus a high noise and low accuracy is presented.
Solution:
This paper puts forward the idea of applying data mining technology to intrusion detection systems and then designs data preprocessing module, association analysis module and cluster module respectively.

C. Wang Pu, Wang Jun-qing [4]

Intrusion Detection Systems (IDSs) have become an efficient defense tool against network attacks since they allow network administrator to detect policy violations.
Problem:
Traditional IDs are vulnerable to original and novel malicious attacks. Also, it is very inefficient to analyze from a large amount volume data such as possibility logs. In addition, there are high false positives and false negatives for the common IDSs.
Solution:
Data mining has been popularly recognized as an important way to mine useful information from large volumes of data which is noisy, fuzzy, and random.

D. Singh et. al in [11]

Proposed an ontology agent based focused crawler (O-ABFC) which improves existing agent based focused crawlers by using ontology and contextual information in crawling. Use of ontology is emerging as a promising tool that eliminates simple keyword based crawling method as it introduces semantics or contexts in which a keyword is searched. Author proposed an intelligent & adaptive ontology mapping mechanism for providing an interface that facilitates agent interaction in homogenous as well as heterogeneous ontology's. Their work automates the ontology-mapping task using multi-agent system that not only overcomes the curse of already existing mapping mechanisms but also is time efficient. Content mining agent (CMA) periodically visits indexes maintained at different servers and provides the newly added documents to DMA and SMA, which update their table by recording appropriate features. CMA performs mining on these tables by using text-mining tools to get knowledge about the recorded documents. Let us consider, if an author name appears with 50 different files written in 6different languages, then it can derive pattern of author name, context, geographical area of publishing (if any) and language in which it written. If context of those file span across semantic web, agent technology fuzzy logic and neural networks then CMA can draw the conclusion that author's field of work is artificial intelligence and thus whenever there is some query for artificial intelligence papers, the work of this author may also be listed as part of output in user default language. Ontology database will help in this kind of context generalization.

III. EXISTING TECHNIQUE

Agent based technique works with the existing agent and dependency on the existing crawler.

A. Solution provided:

An AI based technique provide CMA based approach over the existing agent technique.

B. Yan jun Zhao, Ming jun Wei, Jing Wang[7]

On the basis of further analyzing the operational mechanism of the existing intrusion detection system model, in allusion to the existing.
Problem:
Powerless, high false negative rate, low detection efficiency and the lack of the rule base automatic extension mechanism to unknown aggressive behavior for existing detection mechanisms, Combining the relevant knowledge of data mining technology, then to design one improved network intrusion detection system model based on data mining, combined misuse detection and anomaly detection.
Solution:
In the model, we select the K-means algorithm in clustering analysis and the Apriori algorithm in association rule mining and improve it.

C. Abdullah, I. Abd-algafar, G.I. Salama, A. Abd-alhafez [8]

An Intrusion Detection System (IDS) is a software application or device that monitors the system or activities of network for policy violations or malicious activities and generates reports to the management system. A number of systems may try to prevent an intrusion attempt but this is neither required nor expected of a monitoring system. The main focus of Intrusion detection and prevention systems (IDPS) is to identify the possible incidents, logging information about them and in report attempts. In addition, organizations use IDPS for other purposes, like identifying problems with security policies, deterring individuals and documenting existing threats from infringing security policies. IDPS have become an essential addition to the security infrastructure of nearly every organization. Various methods can be used to detect intrusions but each one is specific to a specific method. The main goal of an intrusion detection system is to detect the attacks efficiently. Furthermore, it

is equally important to detect attacks at a beginning stage in order to reduce their impacts. This research work proposed a new approach called outlier detection where, the anomaly dataset is measured by the Neighborhood Outlier Factor (NOF). Here, trained model consists of big datasets with distributed storage environment for improving the performance of Intrusion Detection system. The experimental results proved that the proposed approach identifies the anomalies very effectively than any other approaches.

Problem:

Existing system doesn't provide the logging information and monitoring feature.

Solution:

The provided solution with the IDPS, provides a logging and monitoring system for the various attacks.

IV. PROBLEM IDENTIFICATION

Previous algorithm for the Intrusion detection data mining technique got various issue while working in network data environment:

1. The existing technique does not compute high accuracy and detection rate over the dataset.
2. The existing technique work over the data which is limited parameter in nature.
3. The existing techniques compute less precision while a high precision is required irrespective of number of dataset.

V. COMPARATIVE ANALYSIS

The existing algorithm for the Intrusion detection system and the classification mechanism being discussed having the following limitation on which we are working:

1. The mentioned previous techniques having the poor accuracy and detection rate for the outlier detection over the standard KDD dataset.

2. The existing technique having the high computation time.

3. In order to process heavy a relative result is not obtained using the existing technique.

Table 1: Analysis Table Previous Techniques.

Technique	Problem Formulation	Solutions
IDPS, NOF technique.	Existing system doesn't provide the logging information and monitoring feature	provides a logging and monitoring system for the various attacks
CMA based approach	High false detection rate and other parameters.	K-means algorithm in clustering analysis and the Apriori algorithm in association rule mining and improve it.
ontology agent based focused crawler (O-ABFC)	Agent and dependency on the existing crawler.	AI based CMA provided.
Traditional IDS technique	Traditional IDs are vulnerable to original and novel malicious attacks.	Data mining based technique for detection is performed.

The table 1, above shows the comparison analysis over multiple available techniques.

VI. CONCLUSION

Web mining and Intrusion over the web platform plays an important role, while in today's scenario we often more deal with the Internet and different web platform. After describing the details of the experiments of multiple algorithm which deals with the web mining and IDS algorithm over the platform. The scenario approached with the network intrusion over the web data and details over the network have been observed. The amalgamation of web mining techniques with agent technology will lead to improved performance, reduced network traffic, and better results.

REFERENCES

1. Zhan Jiuhua Intrusion Detection System Based on Data Mining Knowledge Discovery and Data Mining, 2008. WKDD 2008.
2. Bane Raman Raghunath Network Intrusion Detection System (NIDS)Emerging Trends in Engineering and Technology, 2008. ICETET '08.
3. Changxin Song Design of Intrusion Detection System Based on Data Mining Algorithm 2009 International Conference on Signal Processing Systems.
4. Wang Pu Intrusion detection system with the data mining technologies Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference.
5. Gaikwad, D.P. Sonali Jagtap, Kunal Thakare, Vaishali Budhawant Anomaly Based Intrusion Detection System Using Artificial Neural Network and fuzzy clustering International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, 1 (9.) (2012 November).
6. Goyal, C. Kumar GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System, Electrical Engineering and Computer Science, North West University Technical Report (2008).
7. Gu, P. Porras, V. Yegneswaran, M. Fong, W. Lee BotHunter: detecting malware infection through IDS-driven ialog correlation Proc. of 16th USENIX Security Symp. (SS'07) (2007 Aug), pp. 12:1–12:16.
8. G. Gu, J. Zhang, W. Lee BotSniffer: detecting botnet command and control channels in network traffic Proc. of 15th Ann. Network and Distributed Sytem Security Symp. (NDSS'08) (2008 Feb).
9. V. Jaiganesh, P. Sumathi, S. Mangayarkarasi An Analysis of Intrusion Detection System using back propagation neural network IEEE Computer Society Publication (2013).
10. Buccafurri, G. Lax, D. Rosaci and D. Ursino, _Dealing with Semantic Heterogeneity for Improving Web Usage'. Data Knowledge Eng. Vol. 58, Issue 3, pp. 436–465,2006.
11. Singh A., Juneja D. and Sharma A.K., _Design of Ontology-Driven Agent based Focused Crawlers'. In proceedings of 3rd International Conference on Intelligent Systems & Networks (IISN-2009),Organized by Institute of Science and Technology, Klawad, 14 -16 Feb 2009, pp. 178-Available online in ECONOMICS OF NETWORKS ABSTRACTS, Volume 2, No. 8: Jan 25, 2010.