An Innovative Amalgamation Classification Algorithm for Intrusion Detection System: A Review

Akansha Malviya[1], Dr. Amit Shrivastava[2]

[1]M Tech Scholar, [2] HOD

Computer Science Department, Sagar Institute of Research and Techology Ayodhya Bypass Road Bhopal, India(462001)

Abstract* - As due to swift development of internet, handling and ensuring network traffic security is becoming a key concern of computer network system. Various types of attack came in notices that are not easy to scan. Intrusion is also among them. Various intrusion detection systems are proposed for providing security to the data and information. Data mining is used to extract the useful information from large databases. Data mining techniques can be used to scrutinize and analyze large amount of intrusion over network & classify network data as per the strategy to find out normal and affected data. Various data mining techniques such as classification and clustering are used to make Intrusion detection system. An efficient algorithm in intrusion detection system must have high detection rate, low false alarm rate and high accuracy. This proposed paper give a review on IDS and diverse Data mining techniques used on IDS for the Instant detection of pattern in the network like malevolent , suspicious or normal activities that will be helpful to expand secure information system.

***Keywords* - Intrusion detection system; accuracy; false alarm rate; data mining, database.**

I. INTRODUCTION

Data mining deals with theories, methodologies, and in computer systems for knowledge extraction or mining from large amounts of data. It is used to extract the information from a large number data such as incomplete, noisy and random. In today's world classification is an important technique in data mining. Classification Methods are considered, it focuses on a study on different classification techniques that are most commonly used in data-mining. Algorithms are divided into three forms K-NN classifier, Bayesian network and Decision tree which show the accuracy in performance efficiency and time complexity. Data mining can be divided into two categories: descriptive and predictive Classification techniques.

Intrusion detection systems (IDSs) play a vital role in network security. Only conventional approaches were used for network such as encryption, firewalls, virtual private network etc but are not enough to secure network completely. It is difficult to depend completely on static defense techniques. This increases the need for dynamic technique, which can be monitors system and identify illegal activities. Thus to enhance the network security dynamic approach is introduced and known as Intrusion Detection System. Intrusion Detection system collects online information from the network after that monitors and analyzes this information and partitions it into normal & malicious activities provide the result to system administrator. IDS is the area, where Data mining is used extensively, this is due to limited scalability, adaptability and validity. In IDS data is collected from various sources like network log data, host data etc. Since the network traffic is large, the analysis of data is too hard. This give rise to the need of using IDS along with different Data mining techniques for intrusion detection The success of an IDS can be characterized in both detection rates (DR) and false positives (FP) for different types of intrusions. Intrusion is a type of attack that attempts to evade the security method of a computer system. It is the process of monitoring and analyzing the measures occurring in a system to identify the signs of security problems. There are two main types of IDS: misuse detection and anomaly detection. Misuse detection attempts to equivalent the patterns and signatures which is already known by the invader in the network. This technique is based on the recognition of Traffic anomalies.

A Bayesian network is used to model a province containing uncertainty. It is also said to be directed acyclic graph (DAG) where each node represents a detached random variable of interest. Each node contains the random variable that represents and a conditional probability table (CPT). The CPT of a node contains probabilities of the node being in a specific state given the states of its parents.

II. LITERATURE SURVEY

Tsern-Huei Lee, Nai-Lun Huang [1] “A Pattern-Matching Scheme With High Throughput Performance and Low Memory Requirement” author present here a pattern matching scheme for provide a hogh throughput and less memory spaces, the details are a pattern-matching architecture consisting of a stateful pre-filter and an AC-based verification engine. The stateful pre-filter is optimal in the sense that it is equivalent to utilizing all previous query results. In addition, the filter can be easily realized with bitmaps and simple bitwise-AND and shift operations. T Well- known pattern-matching algorithms include Knuth–Morris–Pratt (KMP), Boyer–Moore (BM), Wu Manber (WM), and Aho Corasick (AC). The KMP and BM algorithms are efficient for single-pattern matching, but are not suitable for matching multiple patterns.

In Cheng-Hung Lin, and Shih-Chieh Chang [2] "**Efficient Pattern Matching Algorithm for Memory Architecture**" author present here a pattern matching algorithm. Network intrusion detection system is used to inspect packet contents against thousands of predefined malicious or suspicious patterns. Because traditional software alone pattern matching approaches can no longer meet the high throughput of today's networking, many hardware approaches are proposed to accelerate pattern matching. Among hardware approaches, memory-based architecture has attracted a lot of attention because of its easy reconfigurability and scalability. In order to accommodate the increasing number of attack patterns and meet the throughput requirement of networks, a successful network intrusion detection system must have a memory-efficient pattern- matching algorithm and hardware design. In this paper, they propose a memory-efficient pattern-matching algorithm which can significantly reduce the memory requirement.

Cheng-Hung Lin, Yu-Tang Tai, [3]Shih-Chieh Chang in "Optimization of Pattern Matching Algorithm for Memory Based Architecture" described here a due to the advantages of easy re-configurability and scalability, the memory-based string matching architecture is widely adopted by network intrusion detection systems (NIDS). In order to accommodate the increasing number of attack patterns and meet the throughput requirement of networks, a successful NIDS system must have a memory-efficient pattern-matching algorithm and hardware design. In this paper, they propose a memory-efficient pattern-matching algorithm which can significantly reduce the memory requirement.

Majid Nezakatolhoseini and Mohammad Amin Taherkhani "a framework for performance evaluation of asips in network-based ids[4]" describe nowadays efficient usage of high-tech security tools and appliances is considered as an important criterion for security improvement of computer networks. Based on this assumption, Intrusion Detection and Prevention Systems (IDPS) have key role for applying the defense in depth strategy. In this situation, by increasing network bandwidth in addition to increasing number of threats, Network-based IDPSes have been faced with performance

challenge for processing of huge traffic in the networks. A general solution for this bottleneck is exploitation of efficient hardware architectures for performance improvement of IDPS. In this paper a framework for analysis and performance evaluation of application specific instruction set processors is presented for usage in application of attack detection in Network based Intrusion Detection Systems (NIDS). By running this framework as a security application on V850, OR1K, MIPS32, ARM7TDMI and PowerPC32 microprocessors, their performance has been evaluated and analyzed.

Benjamin C. Brodie, Ron K. Cytron, and David E. Taylor [5] in "A Scalable Architecture for High-Throughput Regular-Expression Pattern Matching" present and evaluate an architecture for high throughput pattern matching of regular expressions. Our approach matches multiple patterns concurrently, responds rapidly to changes in the pattern set, and is well suited for synthesis in an ASIC or FPGA. Our approach is based on a new and easily pipelined state machine representation that uses encoding and compression techniques to improve density. We have written a compiler that translates a set of regular expressions and optimizes their deployment in the structures used by our architecture. We analyze our approach in terms of its throughput, density, and efficiency. We present experimental results from an implementation in a commodity FPGA, showing better throughput and density than the best known approaches. They presented an architecture that solves the more general problem of regular-expression pattern matching with throughput and density rivaling the best known solutions to the simpler problem of string matching. Our experiments, conducted under modest technology assumptions, show that we can sustain a throughput of 16 Gbps at a density that supports nearly 1,000 regular-expression engines on a die.

Zachary K. Baker and Viktor K. Prasanna [6] "High throughput Linked Pattern Matching for Intrusion Detection Systems" Author presents a hardware architecture for highly efficient intrusion detection systems. In addition, a software tool for automatically generating the hardware is presented. Intrusion detection for network security is a compute-intensive application demanding high system performance. By moving both the string matching and the linking of multi-part rules to hardware, our architecture leaves the host system free for higher-level analysis. The tool automates the creation of efficient Field Programmable Gate Array architectures (FPGA).

Young H. Cho and William H. Mangione Smith "A Pattern Matching Coprocessor for Network Security" [7] Author estimated that computer network worms and virus caused the loss of over $55B in 2003. Network security system use techniques such as deep packet inspection to detect the harmful packets. While software intrusion detection system running on general purpose processors can be up dated in response to new attacks. They lack the processing power to monitor gigabit networks. We present a high performance pattern matching co-processor architecture that can be used to monitor and identify a large number of intrusion signatures. The design consists of a bank of pattern matchers that are used to implement a highly concurrent filter. The pattern matchers can be programmed to match multiple patterns of various lengths, and are able to leverage the existing databases of threat signatures. We have been able to program the filters to match all the payload patterns defined in the widely used Snort network intrusion detection system at a rate above 7 Gbps, with memory space left to accommodate threat signatures that become available in the future.

[8] Tran Ngoc Thinh, Surin Kittitornkun "Massively Parallel Cuckoo Pattern Matching Applied For NIDS/NIPS" author described a Cuckoo-based Pattern Matching (CPM) engine based on a recently developed hashing algorithm called Cuckoo Hashing. We implement the improved parallel Cuckoo Hashing suitable for hardware-based multi pattern matching with arbitrary length. CPM can rapidly update the static pattern set without reconfiguration while consuming the lowest amount of hardware. With the power of massively parallel processing, the speedup of CPM is up to 128X as compared with serial Cuckoo implementation. Compared to other hardware systems, CPM is far better in performance and saves 30% of the area.

[9] Zhongqiang Chen, Yuan Zhang, Zhongrong Chen and Alex Delis "A Digest and Pattern Matching Based Intrusion Detection Engine" author described here about a pattern matching based intrusion detection and the details are an Intrusion detection/prevention systems (IDSs/IPSs) heavily rely on signature databases and pattern matching (PM) techniques to identify network attacks. The engines of such systems often employ traditional PM algorithms to search for telltale patterns in network flows. The observations that real world network traffic is largely legitimate and that telltales manifested by exploits rarely appear in network streams lead us to the proposal of Finger printer. This framework integrates fingerprinting and PM methods to rapidly distinguish well-behaved from malicious traffic. Finger printer produces concise digests or fingerprints for attack signatures during its programming phase. In its querying phase, the framework quickly identifies attack-free connections by transforming input traffic into its fingerprint space and matching its digest against those of attack signatures.

To protect intranets and computer systems from being compromised, IDSs/IPSs employ PM techniques to identify intrusions often with the help of an attack signature database. By matching the incoming streams against each signature with exact PM algorithms such as Boyer Moore and Aho Corasick-an IDS/IPS generates a positive verdict if a match occurs and a negative verdict otherwise. Clearly, a positive verdict can be delivered by scanning on average half of the signatures while a negative one necessitates the involvement of the entire signature database. The latter is obviously more computationally intensive and consequently legitimate traffic gets heavily penalized. In this paper, we propose the Finger printer whose aim is to accelerate the attack identification process of IDSs/IPSs based on the observations that the vast majority of the Internet traffic is legitimate and telltale patterns in signatures are often only unique to attacks. The Finger printer integrates fingerprinting and PM techniques to generate negative verdicts very quickly for attack-free streams. At first, the framework develops a concise and compact fingerprint for each attack signature. Then, it transforms the incoming traffic into the fingerprint space and matches its digest against those derived from the signatures. Traffic is exploit-free if no FM exists. Otherwise, Finger printer resorts to the Boyer–Moore method to ascertain that the input indeed satisfies conditions specified in the signatures with matching fingerprints. We combine multiple fingerprinting approaches such as Bloom–Filter and Rabin–Fingerprint in order to reduce false matches that occur when the input shares the same fingerprints with signatures but fails to match the exact patterns specified by the signatures. We have implemented the Finger printer as a PME in the open-source IDS/IPS Snort and experimentally evaluated with a number of traces.

[10] Giorgos Vasiliadis, Michalis Polychronakis, Sotiris Ioannidis "MIDeA: A Multi Parallel Intrusion Detection Architecture" author present here a Network intrusion detection systems are faced with the challenge of identifying diverse attacks, in extremely high speed networks. For this reason, they must operate at multi-Gigabit speeds, while performing highly complex per packet and per-flow data processing. In this paper, we present a multi-parallel intrusion detection architecture tailored for high speed networks. To cope with the increased processing throughput requirements, our system parallelizes network traffic processing and analysis at three levels, using multi-queue NICs, multiple CPUs, and multiple GPUs. The proposed design avoids locking, optimizes data transfers between the different processing units, and speeds up data processing by mapping different

operations to the processing units where they are best suited. Our experimental evaluation shows that our prototype implementation based on commodity off-the-shelf equipment can reach processing speeds of up to 5.2 Gbit/s with zero packet loss when analyzing traffic in a real network, whereas the pattern matching engine alone reaches speeds of up to 70 Gbit/s, which is an almost four times improvement over prior solutions that use specialized hardware.

Abhaya1, Kaushal Kumar2 in "Data Mining Techniques for Intrusion Detection: A Review [11]", they explain most publicized attack on network traffic is considered as Intrusion. Intrusion detection system has been used for ascertaining intrusion and to preserve the security goals of information from attacks. Data mining techniques are used to monitor and analyze large amount of network data & classify these network data into anomalous and normal data. Since data comes from various sources network traffic is large. Data mining techniques such as classification and clustering are applied to build Intrusion detection system. An effective Intrusion detection system requires high detection rate, low false alarm rate as well as high accuracy. This paper presents the review on IDS and different Data mining techniques applied on IDS for the effective detection of pattern for both malicious and normal activities in network, which helps to develop secure information system.

[12] Christopher Kruegel by "Bayesian Event Classification for Intrusion Detection" author explain that Intrusion detection systems (IDSs) attempt to categorize attacks by comparing collected data to predefined signatures known to be malicious (misuse-based IDSs) or to a model of legal behavior (anomaly-based IDSs). Anomaly-based approaches have the improvement of being able to detect previously unknown attacks, but they suffer from the complexity of building robust models of acceptable behavior which may result in a large number of false alarms. Almost all current anomaly-based intrusion detection systems classify an input event as normal or anomalous by analyzing its features, utilizing a number of different models.

[13] Shyara Taruna R et al, in "Enhanced Naïve Bayes Algorithm for Intrusion Detection in Data Mining" explain that Classification is a classic data mining technique based on machine learning. Classification is used to classify each item in a set of data into one of predefined set of classes. Naïve Bayes is a commonly used classification supervised learning method to predict class probability with belonging data set. This paper proposes a new method of Naïve Bayes Algorithm in which we tried to find effective detection rate and false positive rate of given data. They tested the performance of the proposed algorithm by employing KDD99 benchmark network intrusion detection dataset.

[14] Jaina Patel1,in "Effective Intrusion Detection System using Data Mining Technique", explain that Network Security has become the important foundation with the tremendous increase in usage of network-based services and information sharing on networks. Intrusion poses a serious risk to the network security and compromises integrity, confidentiality & availability of the computer and network resources. Intrusion Detection System (IDS) is one of the looms to detect attacks and anomalies in the network. In this paper a hybrid model is proposed that integrates Anomaly based Intrusion detection technique with Signature based Intrusion detection technique is divided into two stages. In first stage, the signature based IDS SNORT is used to create alerts for anomaly data. In second stage, data mining techniques "k-means + CART" is used to cascade k-means clustering and CART (Classification and Regression Trees) for classifying normal and abnormal activities.

[15] Mrutyunjaya Panda1, in "NETWORK INTRUSION DETECTION USING NAÏVE BAYES" explains that with the tremendous growth of network-based services and sensitive information on networks, network security is more important. Intrusion poses a serious security risk in a network environment. The ever growing new intrusion types poses a sober problem for their detection. In this paper, they apply one of the efficient data mining algorithms called naïve bayes for anomaly based network intrusion detection. Experimental results on the KDD cup'99 data set show the novelty of our approach in detecting network intrusion.

[16]By Hesham Altwaijry, "Bayesian based intrusion detection system", in this paper an intrusion detection system is developed using Bayesian probability. The system developed is a naive Bayesian classifier that is used to identify possible intrusions. The system is trained on a priori subset of the KDD dataset. The trained classifier is then tested using a larger subset of KDD dataset. The Bayesian classifier was able to detect intrusion with a superior detection rate.

[17] Dr. Saurabh Mukherjeea, in "Intrusion Detection using Naive Bayes Classifier with Feature Reduction" in this paper the Intrusion detection is the process of monitoring and analyzing the measures occurring in a computer system in order to detect symbols of security problems. Today most of the intrusion detection approaches focused on the issues of feature selection or reduction. The purpose of this study is to identify important features in building IDS that is computationally efficient and effective. For this they study the performance of three standard feature selection methods using

Correlation-based Feature Selection, Information Gain and Gain Ratio.

[18] Ahmed Youssef, in "NETWORK INTRUSION DETECTION USING DATA MINING AND NETWORK BEHAVIOUR ANALYSIS" explain the Intrusion detection has become a critical component of the administrator due to the vast number of attacks patiently threaten our computers. A traditional intrusion detection system is limited and does not provide a complete solution for the problem. They search for potential malicious activities on network traffics; they sometimes succeed to find true security attacks and anomalies.

III. ISSUES

In network security various techniques have proposed to detect unauthorized use, misuse and abuse of computer systems by both system insiders and external intruders. Due to the rapidly increasing unauthorized activities, Intrusion Detection System (IDS) as a component of defense-in- depth is very necessary because traditional techniques cannot provide complete protection against intrusion for example:

- The correlation of alarm is not précised.
- The detection and prediction of false positive and false negative rate is high.
- The measurement of abnormal behaviour using the pattern structure has limited scope.
- The failure rate of this type of intrusion detection systems is very high.

The nature of intruder file nature is adaptive, so detection and classification is very difficult. Now modern technology of data mining and biological inspires function effect the classification and detection rate of intrusion. In the field of intrusion detection are open research are for algorithm approach and feature optimization technique. In this dissertation we perform following contribution of work in the field of intrusion detection system based on network.

1. Feature reduction of intruder attribute
2. Improve the rate of classification of classifier using optimization technique
3. Reduce the false detection of classifier.

IV. CONCLUSION

A review of diverse classification and clustering data mining method for intrusion detection is analyzed in this paper. After calculating detection rate, accuracy, execution time and false alarm rate, we conclude that execution time with Support vector machine is reduced and also show high accuracy with small dataset. As Naive Bayes is a classifier that is easy to implement, but decision tree show high detection rate for large dataset. In clustering techniques, less execution time of KMeans with numerous data point K-Medoids is having better result. If decision tree work with genetic algorithm then it performs better and it can be a best approach. Efficiency of k-nn can be improved if number of data sets increases. We conclude from the review that rapid algorithm for classifiers are: Naive Bayes algorithm, Decision tree and lastly k-nn algorithm.

REFERENCES

[1] Tsern-Huei Lee, Nai-Lun Huang "A Pattern-Matching Scheme With High Throughput Performance and Low Memory Requirement" IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 21, 2013. Pp 1104-1116.

[2] Cheng-Hung Lin, and Shih-Chieh Chang "Efficient Pattern Matching Algorithm for Memory Architecture" IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL. 19, 2011. Pp 33-41.

[3] Cheng-Hung Lin, Yu-Tang Tai, Shih-Chieh Chang"Optimization of Pattern Matching Algorithm for Memory Based Architecture" ACM 2007. Pp 11-17.

[4] Majid Nezakatolhoseini and Mohammad Amin Taherkhani "a framework for performance evaluation of asips in network-based ids" International Journal of Network Security & Its Applications (IJNSA), Vol.4, 2012. Pp 43-55.

[5] Benjamin C. Brodie, Ron K. Cytron, and David E. Taylor "A Scalable Architecture For High-Throughput Regular-Expression Pattern Matching" Proceedings of the 33rd International Symposium on Computer Architecture IEEE 2006. Pp 1-12.

[6] Zachary K. Baker and Viktor K. Prasanna "High throughput Linked Pattern Matching for Intrusion Detection Systems" ACM October 2005. Pp 101-112.

[7] Young H. Cho and William H. Mangione Smith "A Pattern Matching Coprocessor for Network Security" ACM June 2005. Pp 234-239.

[8] Tran Ngoc Thinh, Surin Kittitornkun "Massively Parallel Cuckoo Pattern Matching Applied For NIDS/NIPS" Fifth International Symposium on Electronic Design, Test & Applications, IEEE 2010. Pp 217-221.

[9] Zhongqiang Chen, Yuan Zhang, Zhongrong Chen and Alex Delis "A Digest and Pattern Matching Based Intrusion Detection Engine" Computer Journal Advance Access 2009. Pp 1-25.

[10] Giorgos Vasiliadis, Michalis Polychronakis, Sotiris Ioannidis "MIDeA: A Multi Parallel Intrusion Detection Architecture" ACM October 2011. Pp 1-12.

[11] Abhaya1, Kaushal Kumar2 "Data Mining Techniques for Intrusion Detection: A Review", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 6, June 2014.

[12] Christopher Kruegel "Bayesian Event Classification for Intrusion Detection"

[13] Shyara Taruna R et al, " Enhanced Naïve Bayes Algorithm for Intrusion Detection in Data Mining" (JCSIT)

International Journal of Computer Science and Information Technologies, Vol. 4 (6) , 2013, 960-962.

[14] Jaina Patel1, “Effective Intrusion Detection System using Data Mining Technique” June 2015, Volume 2, Issue 6 JETIR (ISSN-2349-5162) JETIR1506034 Journal of Emerging Technologies and Innovative Research (JETIR).

[15] Mrutyunjaya Panda1, “NETWORK INTRUSION DETECTION USING NAÏVE BAYES” IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007, 258.

[16] Hesham Altwaijry, “Bayesian based intrusion detection system” Journal of King Saud University – Computer and Information Sciences (2012) 24, 1–6

[17] Dr. Saurabh Mukherjeea, “ Intrusion Detection using Naive Bayes Classifier with Feature Reduction” Procedia Technology 4 (2012) 119 – 128 2212-0173 © 2012 Published by Elsevier Ltd.doi: 10.1016/j.protcy.2012.05.017 C3IT-2012.

[18] Ahmed Youssef, “NETWORK INTRUSION DETECTION USING DATA MINING AND NETWORK BEHAVIOUR ANALYSIS”International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011DOI : 10.5121/ijcsit.2011.3607 87.