ISSN: 2277 – 9043
International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)
Volume 3, Issue 7, July 2014

Enhanced Intrusion Detection System using Hybrid Machine Learning Approach

Pavan Singhal[1], Gajendra Singh[2]
Dept of Computer Science & Engg.
Sri Satya Sai Institute of Sc. & Technology
Sehore, India

Abstract— **Modern business has run on technology and it is based on communication and consequently the gigantic speed of the today's internet or communication is the cause of the advancement in telecommunication and semiconductor technologies together. Billions of users are accessing the internet hundreds of time in a day. Due to flexibility and ease of networking services security is the chief concern. To get protected Intrusion Detection and Preventions System are the best option to assure security. In this article Anomaly based IDPS has been proposed and evaluated using hybrid machine learning approach. Machine learning sub branch of the soft computing had evolved since last decade has present more promising solution in the field of the security (host and network). Various methods of machine learning have been tested to produces better results in detection of intrusive activities. Classification (KNN) and evidence theory (DS) is types of machine learning approach and support to provide better solution in this direction. Proposed method has adopted the idea of KNN and DS Theory to fasten the detection speed, achieving better efficiency and accuracy with low false positive and negative ratio. Obtained results have achieved the accuracy about 97.47% and false ratio has minimized and limited it to 1.2 and 1.3.**

Keywords- DS, DST, IDS, IDPS, KDD, KNN, Machine Learning.

I. INTRODUCTION

According to NIST (National Institute of Standard and Technology) the security is the major concerned and its aim to assure 3 properties onto the information itself- Confidentiality, Integrity and Availability (CIA) also termed as the pillar of the security.

Firewall has deployed as first line of defense to provide security by the means of preserving privacy. Firewall has work as obstacle to the unauthorized persons (or users) to entering onto the house (network or host). But what happened when someone has just crossed the wall or enter as legitimated user (made fool to the security guard on to the gate) how to secure the assets inside. Then there is urgent necessity of the second defense of line as security perimeter.

Intrusion detection and prevention system are works as second line of defense that make secured the assets and shield against the intrusion which has intrude from firewall (or no firewall) to take the advantages of the vulnerability of the firewall ACL (Access Control List). Intrusion detection is an engineering and scientific approach to detect and prevent the unauthorized entrance (intrusion) onto the network or host itself before any harmed action to be happened.

Intrusions were first categorized by J. P. Anderson mentioned in Lunt et. al. [2]. They can be largely classified into three types: external intrusions, internal intrusions, and misfeasors. An external intrusion tries to break into a computer system without appropriate access rights. An internal intrusion originates from a valid user inside a computer system. A masquerader is an internal intruder who logs into the system by use of other users' accounts. A clandestine is also an internal intruder who deceives the system and performs illegal operations. A misfeasor usually abuses his or her authority on the use of a computer system.

Where according to Mukherjee et. al. [3], intrusion detection can be defined as detecting outside intruders "who are using a computer system without authorization" and inside intruders "who have legitimate access to the system but are abusing their privileges". Intrusion detection systems are usually built to identify these unauthorized behavior of outside or inside intruders and to enforce the security of computer systems.

IDS (Intrusion Detection System) have work onto the two detection approach – signature based and anomaly based. Signature based is fast but unable to detect unknown attack while anomaly is overcome the flaws of signature based but suffered from high false ratio.

To reduce the false ratio of anomaly detection various methods has proposed and experiments with integration various methods like data mining, rule based etc.

Machine learning sub branch of the soft computing had evolved since last decade has present more promising solution in the field of the security (host and network). Various methods of machine learning have been tested to produces better results in detection of intrusive activities. Classification (KNN) and evidence theory (DS) is types of machine learning approach and support to provide better solution in the direction.

1.

Rest of the article is organized as follow, Section II describes related works. Section III presents proposed RTT based wormhole detection mechanism for AODV protocol. Section IV discusses the obtained results and the performance analysis and finally Section V concludes the papers with the future directions of this work.

II. RELATED WORKS

In this section we review the current literature and related work in the areas of Intrusion system through examination of various online resources, journals, and texts; we have attained the necessary information for the rest of this system.

Intrusions were first categorized by J. P. Anderson mentioned in Lunt et. al. [2]. They can be largely classified into three types: external intrusions, internal intrusions, and misfeasors. An external intrusion tries to break into a computer system without appropriate access rights. An internal intrusion originates from a valid user inside a computer system. A masquerader is an internal intruder who logs into the system by use of other users' accounts. A clandestine is also an internal intruder who deceives the system and performs illegal operations. A misfeasor usually abuses his or her authority on the use of a computer system.

Where according to Mukherjee et. al. [3], intrusion detection can be defined as detecting outside intruders "who are using a computer system without authorization" and inside intruders "who have legitimate access to the system but are abusing their privileges". Intrusion detection systems are usually built to identify these unauthorized behavior of outside or inside intruders and to enforce the security of computer systems.

Various methods of misuse detection has been proposed, some them are hybrid approach viz. Association Rule Mining and bio inspired behavioral connection designing it includes the connected networks of cognitive etc has been given by Mansour et. al. [4]. In this paper a hybrid misuse based intrusion detection method has been proposed by the author. The main concept of this hybrid model is connectionist model – which is the combination of psychological and modern computational techniques of artificial intelligence in conjunction with neurons behavior has been included to solve and optimize the best solutions to detect the intrusive activities. Author has applying the connectionist theory due to leverage the association rule mining that's maps the associated data with frequent item and connectionist paradigm. Hence the extraction of the useful and touch of the machine learning concept comes together that more accurately classifies the intrusions. Author has gives the selected data input and applied association rule mining to fasten the classification of data using relations among them with their frequency of occurrence. According to article proposed hybrid approach given by author has promises the better classification of attacks distinctively for U2R and R2L types of intrusion. An author also promises that it performs better than neural based IDS techniques in conjunction with superior Detection Ratio (DR) along with minimum false alarm rate.

Another concept of IDS detection using fuzzy logic mentioned in R.Shanmugavadivu et.al [5]. Proposed a system in which he designed fuzzy logic-based system for identifying the intrusion activities within a network. The fuzzy logic-based system able to detect an intrusion behavior of the networks since the rule base contains a better set of rules. In this method author has applied the automated procedure to procreation of fuzzy rules that have distinct in nature by employing frequent items concepts on it. Author has examined and estimates the proposed IDS method on to the standard KDD Cup'99 intrusion dataset. In this an effective set of fuzzy rules for inference approach were identified automatically by making use of the fuzzy rule learning strategy, which is appropriate to detect intrusive activities on to network. Initially distinct rules were generated by employing data mining methods onto the single length frequent items onto both normal and intrusive (attack). After that, such rules are considered as fuzzy rules were recognized by fuzzifying process the distinct rules and inputted onto the fuzzy system for further classification testing dataset. The experimental results showed that the proposed method is effective in detecting various intrusions in computer network.

Lunt et. al. [2] has revealed the idea of IIDS describes a prototype intelligent intrusion detection system (IIDS) to demonstrate the effectiveness of data mining techniques to facilitate make use of fuzzy logic theory. Proposed method of author has is hybrid approach that integrates both intrusion detection approaches: first anomaly based intrusion detection using fuzzy data mining techniques, and second misuse detection using traditional rule-based expert system techniques. First one i.e. anomaly-based method introspect the divergence pattern from normal patterns or behavior. Whereas second one i.e. misuse detection examines heretofore identified patterns or signature (of behavior) susceptible to be an attack or intrusion. System logs and network flows mutually treated as inputs for the IDS. This system architecture support both anomaly detection and misuse detection components at both the individual workstation level and at the network level. Both fuzzy and non-fuzzy rules are supported within the system. Furthermore author has also employed genetic for better tuning the association of the membership among function to form membership functions for the fuzzy variables used by this system to and select the most effective set of features for particular types of intrusions.

Consequently Nannan Lu et. al. [6] has adopted the hybrid approach in role of a hybrid rule mining algorithm based on Fuzzy GNP and probabilistic classification for Intrusion Detection is Hybrid rule mining uses fuzzy class association rule mining algorithm to dig out relevance or interesting rules with diverse classes or groups. The classification method is based on the probability distribution of the average matching degree between data and different class rules. As a result, simulations show higher DR (Detection Rate), Accuracy and lower PFR (Positive False Rate), NFR (Negative False Rate), which means that Fuzzy based GNP class association rule

mining has better performance than the conventional class association rule mining. A hybrid misuse based ids which uses the combined structure of an association rule mining algorithm and a connectionist model , is presented .The key idea is to take advantage of different classification abilities of knowledge based and machine learning approaches for different attacks. To lower the load of the association rule mining, the inputs of rule mining algorithm are selected based on the results of a feature relevance analysis. Authors has promises based upon the obtained results mentioned in this article has improved the detection and intrusion classification especially remote to local (R2L) and user to root (U2R) attack classes. Author's proposed (hybrid method) approach out-performed well and as compared to existing neural based detection methods in context of detection rate (DR) and cost per example (CPE). False alarm rate of the proposed model is comparable with other ids, ame idea has also mentioned in Mansour et. al. [4].

Whereas in the field of section of fuzzy class association rule mining proposed in Nannan Lu and et. al [6] presents a novel fuzzy class association rule mining method based on Genetic Network Programming (GNP) for detecting network intrusions.GNP uses directed graph structures as genes instead of strings (Genetic Algorithm) or trees (Genetic Programming), leading to create compact programs and implicitly memorizing past action sequences. By the combination of fuzzy set theory with GNP, the proposed method can deal with mixed database which contains both discrete and continuous attributes. And it can be flexibly applied to both misuse and anomaly detection in Network Intrusion Detection Problem. The author did experiments with practical data provided by KDD99Cup and DAPRA98.The experiment results show that for misuse detection, the proposed method provide high detection rate and low positive false rate and for anomaly detection the method provide high detection rate and reasonable positive false rate even without pre-experienced Knowledge.

Used of Mining algorithm classification have important role as concern of association mining as ARMAGA (Association rules mining algorithm based on a novel Genetic Algorithm), is mine the association rules from an image database in which every image is represented by the ARMAGA representation. In this author uses genetic algorithm for discovering association rule and secondly propose the algorithm compared to the Fuzzy association rules and the extended mining algorithms, and the ARMAGA algorithm avoids generating impossible candidates. The author compare results of the ARMAGA with the results of the Fuzzy association rules and the extended mining algorithms it is better than GA and ARMA through the theoretic analysis, the experimental results and also more efficient in terms of the execution time presents in Shangping Dai et. al [7].

Proposed Hybrid Machine Learning Approach for IDS has influenced from author's [1] in which author has proposed a statistical based IDPS method concentrating flow based method limited it to the peer-to-peer network. Author has used the Knn (N-nearest neighbour) machine learning method to detect and inspect intrusion on to the traffic flow or data set such as KDD'99 and DARPA.

Following problem has been identified from [1]-

1. Author's method has limited to the peer-to-peer Network - Author has not evaluated the hybrid traffic with proposed method. Because huge amount of data has came from different types of network.
2. KNN based method has already exist for packet inspection and intrusion detection – Knn based method is not sufficient to detect the modern abnormality in the traffic. It needs to be strengthening by integrating another classifier.

III. ENHANCED IDS USING HYBRID MACHINE LEARNING APPROACH

Above mentioned two problems inferred from [1] has been the main motivation for the proposed hybrid approach with following modification on it-

1. ***KNN + Dempster–Shafer theory***– As mentioned above knn alone is not sufficient to detect unknown (new) attacks of modern technique using anomaly detection approach. For strengthening the classifier of IDS Dempster–Shafer theory method has been integrated on with the KNN that's make traffic classification (anomaly checker) more accurate and efficient, resulting low false positive ratio.
2. ***Hybrid traffic data set*** – Just limiting the detection method to peer-to-peer network traffic, propose method has expanding it to whole traffic flows (every type) so that wider applicability of the propose method will be achieve and evaluated with modern attack trends that will help to give new research direction in the network security filed.

A. Proposed Algorithm:

1. Load Training data
2. Load Testing data (KDD'99)
For i= 0 to 5 //0= NORMAL, 1 = DoS …etc
{
3. Applying KNN method for all 5 attacks of KDD data
Calculate metrics and check winner
Put in the neighbour list of K
4. Apply Dempster Theory of Evidence onto the results obtained on step 3
Check evidence of attack
Calculate False positive and Negative
} End for;

IV. RESULTS AND PERFORMANCE ANALYSIS

Proposed Hybrid machine learning approach for IDS has provides better solutions. Obtained results shows that

proposed systems outperform better than existing one. Obtained results have achieved the accuracy about 97.47% and false ratio has minimized and limited it to 1.2 and 1.3.

On experimenting with different dataset, the number of normal/abnormal packets is being monitor. We have examined five different dataset in our experiment, with each having corresponding number of rejected or normal packets. In our conducted test the packets could either fall under normal packet type or in the category of attack (DOS, R2L.U2R.PROB).

We have supervised five data set with each 1000 instances of data under .the result of ratio of attacks is represented in tabular format below-

CATEGORY	DATA SET 1	DATA SET 2	DATA SET 3	DATA SET 4	DATA SET 5
Normal	650	645	652	643	647
Probs	50	52	49	53	50
DoS	150	160	148	144	148
R2R	100	90	97	109	100
R2L	50	53	54	51	55

As seen from the output of performance on data set (DS1).it can be made out that when KNN is combined with DS method, the performance gets significantly improved.

Earlier application of isolated KNN on dataset has much greater False Alarm Rate, than later by integrating both KNN and DS Methods. Also there is a considerable enhancement in the true positive and true negative detection ratio. Thus this gives the direct improvised accuracy in the result.

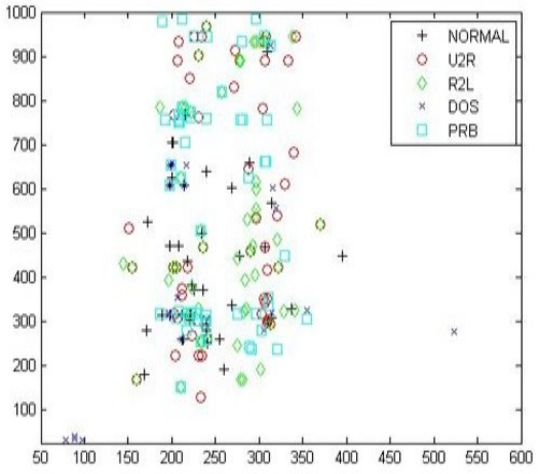
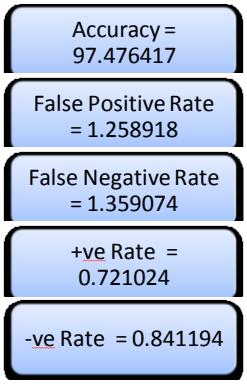


Fig. 1. Attack Detected by Proposed KNN-DS based Method

Performance of proposed KNN-DS-



V. CONCLUSION

This article has presented a enhanced hybrid machine learning approach to defend against intrusion into the network traffic. A dempester evidence theory method is discussed in chapter 4 has been integrated with proposed mechanism to detect intrusion. Which solve the problem that traditional technique of intrusion detection, these techniques are not finding a new pattern of intrusion or intrusive activities. And experiments prove that the method has the property of high classification accuracy. Proposed methods main object is to apply a manner for intrusion detection using KNN classification and Dempester theory of evidence. Through these manners we gathered a new discovered pattern of intrusion and classify Category of pattern and apply event evidence logic with the help of DS- Theory. Finned pattern of intrusion compare with the existing pattern if intrusion and generate a new schema of pattern and update a list of pattern of intrusion detection and improved the true rate of intrusion detection. we have also perform some experimental task with KDD99Cup databases from MIT Lincoln Laboratory show that the proposed method provides competitively high detection rates compared with other machine-learning techniques and crisp data mining

- Intrusion detection is a very challenging area of research in a current scenario. Now every day find a new pattern of intrusion and detection of this pattern are very challenging job. Our object is we apply a manner for intrusion detection using KNN classification and Dempester theory of evidence.
- Through this manner we gathered a new discovered pattern of intrusion and classify Category of pattern and apply event evidence logic with the help of DS- Theory.
- Fined pattern of intrusion compare with the existing pattern if intrusion and generate a new schema of pattern and update a list of

pattern of intrusion detection and improved the true rate of intrusion detection.

- Proposed Method also perform some experimental task with KDD99Cup databases from MIT Lincoln Laboratory show that the proposed method provides competitively high detection rates compared with other machine-learning techniques and crisp data mining.
- Obtained results shows that proposed hybrid machine learning approach has outperform better than existing method based on KNN concept.
- Performance of proposed mechanism has achieved the accuracy about 97.47% and false ratio has minimized and limited it to 1.2 and 1.3.

This section discusses a few areas where the current work can be taken further. In Botnet Detection and DDoS (Distributed Denial of Service) Attack Detection using proposed hybrid approach.

REFERENCES

[1] José Camacho, Pablo Padilla, Pedro García-Teodoro and Jesús Díaz-Verdejo "A generalizable dynamic flow pairing method for traffic classification", Elsevier science direct, Computer Networks 57 (2013) 2718–2732, 2013.

[2] Lunt, T. 1993. Detecting intruders in computer systems. In Proceedings of 1993 conference on auditing and computer technology. (Downloaded from http://www2.csl.sri.com/nides/index5.html on 3 February 1999.)

[3] Mukherjee, B., L. Heberlein, and K. Levitt. 1994. Network intrusion detection. IEEE Network, May/June, 26-41.

[4] Mansour Sheikhan and Zahra Jadidi, " Misuse Detection Using Hybrid of Association Rule Mining and Connectionist Modeling", World Applied Sciences Journal 7 (Special Issue of Computer & IT): 31-37, 2009.

[5] R. Shanmugavadivu Dr. N. Nagarajan "NETWORK INTRUSION DETECTION SYSTEM USING FUZZY LOGIC", Indian Journal of Computer Science and Engineering (IJCSE).

[6] Nannan Lu; Mabu, S.; Wenjing Li; Hirasawa, K.; Grad. Sch. of Inf., Waseda Univ., Fukuoka, Japan "Hybrid rule mining based on fuzzy GNP and probabilistic classification for intrusion detection", SICE Annual Conference 2010.

[7] Shangping Dai; Li Gao; Qiang Zhu; Changwu Zhu; Hua Zhong Normal Univ., Wuhan, "A Novel Genetic Algorithm Based on Image Databases for Mining Association Rules", Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS.