IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661, ISBN: 2278-8727 Volume 5, Issue 6 (Sep-Oct. 2012), PP 37-41
www.iosrjournals.org

Critical analysis of genetic algorithm based IDS and an approach for detecting intrusion in MANET using data mining techniques

IshaShingari
Department of Computer Science Mody Institute of Technology and ScienceLakshmangarh, India

Abstract: *Intrusion can be defined as any action which leads to the compromise of availability, confidentiality or integrity of the system. Intrusion detection system is the mechanism which detects those malicious actions and takes the appropriate security actions. Due to the varied number of network behaviors and the rapid development of attack fashions, it has become essential to use the fast machine-learning-based intrusion detection algorithms with high detection rates and low false-alarm rates. Most of the Intrusion Detection Systems (IDSs) are rule based. In this paper, we have explained genetic algorithm approach to the intrusion detection system. Similarly the data mining approach in intrusion detection is also explained. Both the approaches are analyzed with their respective merits and demerits. The principles behind intrusion detection in MANET (Mobile Adhoc Networks) have also been given. A hybrid approach has also been proposed which considers the principles of the various IDSs. An imperative study of the various intrusion detection systems has been proposed.*

Keywords*-Data mining, Genetic algorithm, Intrusion Detection, IDS, Intrusion Detection System,,MANET*

I. Introduction

Intrusion is defined by Ludin [2] as “any set of actions that attempts to compromise the integrity, confidentiality or availability of a resource”. He also notes that an intrusion is “threat of a person or proxy attempting to break into or misuse one‘s system in violation of an established policy.” A threat or intrusion can also be defined as the potential possibility of a deliberately unauthorized attempt to get hold of the information or to manipulate it. The technique of intrusion detection can be termed as the mechanism which is used for detecting the suspicious activities at host as well as at the network level. The two major intrusion detection techniques are the misuse detection and the anomaly detection [26]. In the anomaly based technique the audit data is utilized to differentiate the abnormal data from the normal data. Whereas in case of misuse detection, patterns of the well-known attacks are used to differentiate the abnormal data from the normal data. This is carried out by matching the audit data and identifying them as the intrusions.

This suggests that, the functioning of misuse detection models can be considered very much similar to that of the anti-virus applications. The misuse intrusion detection system can analyze the system or network and compare its activities against the network and system behaviors. Audit data used for testing and creating rules or defining patterns can be collected from the diverse sourceslike system logs from hosts ,network traffic data andsystem calls from various processes.

A traditional intrusion detection system has a sensor on which the intrusion detection is installed and run. The work of the network sensor is to monitor the network packets like the number of bytes transferred, duration of connection, TCP/IP headers etc. Whereas the host sensor, monitors the memory usage on host and the system logs[3]. The sensor machine generates the security events management console and it also controls the sensors. The IDS (Intrusion Detection System) engine records the logged events by the sensor into the database and generates the alerts based on the rules mentioned upon the happening of security alerts.

A traditional intrusion detection system can be explained as follows as mentioned in Figure 1[1].:

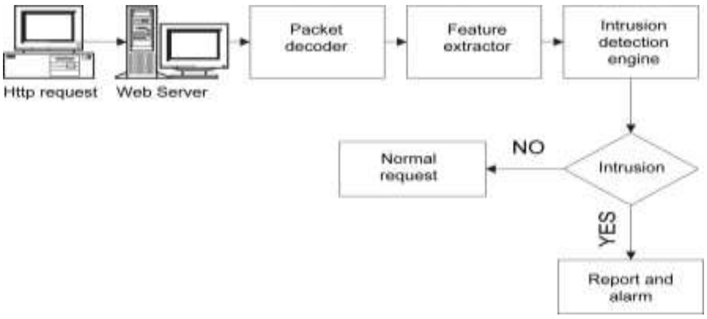

Fig 1. An example of a traditional Intrusion Detection System

II. Genetic Algorithm Approach In Ids

Genetic algorithm [4] employs the metaphor which is brought from genetics as well as biology to iteratively evolve a population of initial individuals to a population of high quality individuals, where each individual represents a solution of the problem to be solved and is composed of a fixed number of genes. The possible number of values of every gene is known as the cardinality of the gene [5]. Each individual is known to be the chromosome. The set of chromosomes thenfurther form the population.

The functioning of the genetic algorithms is initiated by the randomly generated populations of the individuals. After the various generations, the successive quality of the population is improved. The three basic operators of the genetic algorithm which are the selection, mutation and crossover are applied to every individual [17]. The process of crossover involves the changing of the genes between the two chromosomes. The process of mutation involves the random alteration of values of the randomly selected gene of a chromosome. The positions of the varied individuals can be represented in the form of characters, bits and numbers. When genetic algorithm is used for problem solving, three factors will have impact on the effectiveness of the algorithm, they are [4][26]:

a. The selection of fitness function
b. The representation of individuals and
c. The values of the genetic parameters

Functioning of Genetic AlgorithmGenetic algorithm is used for evolving new rules forIDS [6]. Using these rules normal network traffic or audit data is differentiated from abnormal traffic/data. Figure 2, is an example of the genetic algorithm based intrusion detection system. It denotes the detection of audit data and their corresponding counter measures [21].

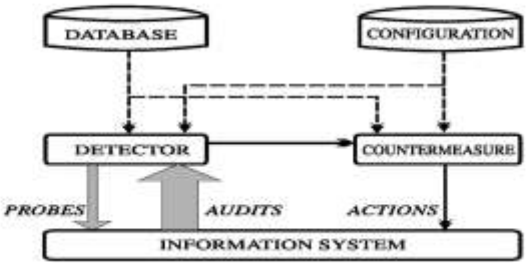

Fig 2. An example of the genetic algorithm based IDS

Most of the intrusion detection systems are rule based. The rules followed in the genetic approach for the intrusion detection system are of the if-then-else syntax [26].

If (condition) then (action)

The condition can be used to check for the port numbers in the network and access other network information; the action corresponds to the step of actions performed if the condition happens to be true. Benefits of using genetic algorithm based approach [16]:

a. *Because of the parallel nature of genetic algorithms, they can explore the various solutions in multiple directions at once.*
b. *The solutions to truly huge problems can be evaluated at once.*
c. *The genetic algorithm based systems are re-trained easily. This helps in adding up of new rules and thus evolves the IDS.[4]*

II. Data Mining In Ids

Data mining is the process to cull out the important information from the huge databases. It picks up the huge amounts of data and extracts the relevant information. It identifies the trends within the data which go beyond the simple analysis. The modern technologies of the sensors, computer networks have the made the data collection and implementation much easier. Data mining contributes in a lot of ways to offer a systematic approach to the system. Currently, many IDSs [4] are rule-based systems where the performances highly rely on the rules identified by security experts [23]. The network traffic is increasing day by day and the process involved in encoding the rules is slow and expensive. Moreover the security associated people have to introduce certain alterations or deploy new rules using a specific rule driven language [9]. In order to overcome the hurdles of rule-based intrusion detection systems, a huge number of intrusion detection systems are employing the data mining techniques [6]. The process of data mining involves the analysis of the data sets for discovering the understandable patterns or the models. Data mining can efficiently find out the patterns of the intrusions for the misuse and identify the profiles involved in the anomalies [28]. This in turn helps in building the classifiers to detect the attacks. The systems which are based on data mining are more deployable and flexible.

The security people only have to label the audit data and indicate the intrusions instead of finding out the hand coding rules for detecting the intrusions [22]. One of the major challenges faced by the Intrusion Detection Systems is feature selection. Many algorithms are very sensitive to the number of features used.

Hence, it is very important to select the features when it comes to improve the detection rates [21]. The date available in raw format is not exactly suitable for intrusion detection. So, the Intrusion detection systems must detect the features from the raw network traffic date which involves a lot of efforts in computation [20]. Feature selection can thereby reduce the computation cost for the feature construction by the reduction of the features. However, in the case of most of the data mining based intrusion detection systems, the feature selection is based on the domain knowledge. The feature selection algorithm can give the estimates of the important data features in the classification [19].

The other challenge which comes across while working for intrusion detection system is the imbalanced intrusion. One of the intrusions is Denial OfService (DoS) [27]. Most of the data mining algorithms are there to minimize the overall error rate of the system. But this in turn tends to increase the error rate of the minority intrusions [18]. However, it is also a fact that the minority intrusions are more dangerous than the majority attacks. One of the problems of supervised anomaly intrusion detection approaches [26] is the high dependency on training data for normal activities.

As the training data only consists of the historical activities, the profile regarding the normal activities can only be inclusive of the historical patterns of the normal behavior. Therefore the new activities which are involved due to the change in the network environment or the services are considered to be the deviations from the previously built profiles and are known as the attacks [12].

Whereas on the other hand, the training data which is attack free is very difficult to obtain as there is no guarantee to protect all the networks present in the real-world [11]. The intrusion detection systems are trained by the data with the hidden intrusions which usually tend to lose the ability to detect the varied intrusions [10]. To overcome these limitations of the supervised anomaly based systems and the anomaly based systems, number of intrusion detection systems go for the unsupervised learning approach.

The unsupervised approach for anomaly detection does not need the training data which is attack free. It detects the attacks by determining the unusual activities under the two assumptions: the majority of activities are normal andthe unusual activities are outliers that are inconsistent with the remainder of data set as per BernetLewis[18].

Thus, the outlier detection techniques are applied in the unsupervised anomaly detection approach. Actually the outlier detection has been utilized in a varied number of practical applications such as voting irregularity analysis, severe weather prediction, fraud detection etc.[3].

III. Intrusion Detection In Manet

A mobile ad-hoc network (MANET) is a self-configuring infrastructure less network of mobile devices connected by wireless. *ad hoc* is Latin and means "for this purpose”. Each device in a MANET is free to move independently in any direction, and will therefore change its links to other devices frequently [30]. Each must forward traffic unrelated to its own use, and therefore be a router. The primary challenge in building a MANET is equipping each device to continuously maintain the information required to properly route traffic. Such networks may operate by themselves or may be connected to the larger Internet.

Due to the restrictions of most MANET routing protocols, nodes in MANETs assume that other nodes always cooperate with each other to relay data [24]. This assumption leaves the attackers with the opportunities to achieve significant impact on the network with just one or two compromised nodes. In order to solve this problem IDS should be added to enhance the security level of MANETs. If MANET can detect the attackers as soon as they enter the network, we will be able to completely eliminate the potential damages caused by compromised nodes at first time. But the truth is that the IDSs usually act as the second layer in MANETs, and it is a great complement to existing proactive approaches [27]. Jie et al. [4] presented a very thorough survey on contemporary IDSs in MANETs.

The difference between wired infrastructure networks andmobile ad hoc networks raises the need for new IDS models that can handle new security challenges [21]. Due to the emerging security needs in MANET, cooperative intrusion detectionModel has been proposed in [27], where every node participates in running its IDS in order to collect and identify possible intrusions. If an anomaly is being detected with weak evidence, then a global detection process is initialized for the further investigation about the intrusions throughout the secure channel. An enhanced version of this process was proposed in

[8] where a set of intrusions can be identified with their corresponding sources. Moreover, the authors have also addressed the problem occurred due to runtime resource constraints through the modeling of the repeatable and ransom election framework. An elected leader is a concept in which a node is responsible for detecting the intrusions for the predefined phase of time. A random forest generated algorithm is employed for selecting the elected leader nodes randomly.

In the modular IDS system based on the mobile agents is proposed and the authors point out the impact of limited computational and battery power on the network monitoring tasks. In this approach, the solution does not pay much attention towards the remaining nodes and the selfishness node issue in the MANET. In order to motivate the selfish nodes in the routing, CONFIDANT [6] proposes a system in which every node keeps track of the misbehaving or the selfish nodes. This reputation system is being built up upon the negative evaluations instead of the positive impression. Wherever there is specified threshold exceeded, then an appropriate action is performed against the node. Therefore the nodes are being encouraged in order to participate in punishing the misbehaved or selfishly behaving nodes through the negative reputation. As this consequence, the malicious node can then broadcast the negative impression about the misbehaving node to be punished.

CORE is a proposed mechanism for the cooperative enforcement which is based on the monitoring and reputation systems [25]. The major goal of this system is to find out the selfish behaving nodes and then enforcing them to cooperate. Every node keeps a track of the other nodes which uses the reputation as the metric. CORE gives the assurance that the misbehaving nodes are punished gradually by keeping them away from the communication services. In this model, the reputation is been calculated on the basis of the monitored data by the local nodes and the information being provided by the neighboring nodes in the network.

IV. Hybrid Approach In Intrusion Detection In Manet

The systems which are based upon the anomaly detection are used for detecting the unknown attacks. These systems are designed for the analysis which is done offline [25]. The analysis is often done offline due to their memory overheads and expensive processing [13]. The systems which are signature based leverage the characterized attack signatures for detecting the known attacks in the real time traffic scenario [29]. The hybrid system approach or the Hybrid Intrusion Detection System (HIDS) integrates the flexibility of the ADS and the accuracy of the signature based SNORT intrusion detection system. The custom designed ADS isconnected in cascade with the SNORT [7]. Then these two subsystems combine together to merge all the traffic events being initiated by the both malicious as well as the legitimate users [15].

V. Conclusion

In this paper we have done a critical analysis of the genetic based approaches in the intrusion detection systems. We have given a brief introduction of IDS and then we have discussed the various security measures. The merits of genetic algorithm based approach have also been mentioned. Next the data mining approach is explained. The data mining strategies have been explained which are being used in the intrusion detection system. A brief introduction about MANET and then intrusion detection system in MANETs has been highlighted. The various techniques have been analyzed in this paper regarding the intrusion detection systems. The hybrid approach is also mentioned in which two IDSs are combined to counter the misuse detection and anomaly detection based attacks.

References

Journal Papers:

[1] T. Anantvalee and J. Wu."A Survey on Intrusion Detection in Mobile Ad hoc Networks".In Wireless/Mobile Security, Springer, 2008.

[2] Vivek K. Kshirsagar, Sonali M. Tidke& Swati Vishnu" Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview"International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012

[3] vakani and R.S. Rajesh, "Genetic Algorithm for Framing Rules for Intrusion Detection" IJCSNS International Journal of Computer Science and Network Security, Vol. 7 No. 11, November 2007.

[4] K. Tan, K. Killourhy, and R. Maxion, "Undermining an anomaly based intrusion detection system using common exploits," in Proc. Recent Adv.Intrusion Detect. (RAID), Zurich, Switzerland, Oct. 2002, pp. 54–73.

[5] K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," in Proc. 28th Australasian CS Conf., Newcastle, Australia, Jan. 2005, vol. 38, pp. 333–342.

[6] Q. Zhou, L. Gu, C.Wang, J.Wang, and S. Chen, "Using an improved C4.5 for imbalanced dataset of intrusion," in Proc. 4th Annu. Privacy Secure.Trust Conf., Markham, Canada, Oct. 2006, pp. 481–484.

[7] W. Lee and S. Stolfo, "A framework for constructing features and models for intrusion detection systems," ACM Trans. Inf. Syst. Secur., vol. 3, no. 4, pp. 227–261, Nov. 2000.

[8] Kai Hwang, Min Cai, Member,Ying Chen" Hybrid Intrusion Detection with WeightedSignature Generation over Anomalous Internet Episodes"IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 4, NO. 1, JANUARY-MARCH

[9] ZoranaBankovic, Jose M. Moya, Alvaro Araujo, Slobodan Bojanic and Octavio Nieto-Taladriz, "A Genetic Algorithmbased Solution for Intrusion Detection", Journal of Information Assurance and Security 4 (2009) 192-199.

[10]]Tamas Abraham, "IDDM: Intrusion Detection using Data Mining Techniques", Information Technology Division,

[11] Electronics and Surveillance Research Laboratory, DSTOGD- 0286.

[12] Jiong Zhang, Mohammad Zulkernine, and Anwar Haque" Random-Forests-Based Network Intrusion Detection Systems" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 38, NO. 5, SEPTEMBER 2008

Books:

[13] Introduction Of Soft Computing Systems For Software Security Management.

[14] Snort Network Intrusion Detection System. (2006). [Online]. Availablehttp://www.snort.org

[15] Lecture Notes in Computer Science, vol. 2820, pp. 220–237.

[16] V. Barnett and T. Lewis, Outliers in Statistical Data. NewYork: Wiley,1994.

[17] B. Sun. Intrusion Detection in Mobile Ad hoc Networks. Doctoral Dissertation.Texas A&M University, 2004.

Website:

[18] Wikipedia.edu/MANET

Proceedings Papers:

[19] Lundin, E. and E. Jonsson. "Privacy versus Intrusion Detection "Analysis."Proceedings of the Second International Workshop **on** Recent Advances in Intrusion

[20] Shi-Jinn Horng·PingzhiFan,Yao-Ping Chou, Yen-Cheng Chang·Yi Pan" A feasible intrusion detector for recognizing IIS attacks based on neural networks " IN, 1999.

[21] Wei Li, "Using Genetic Algorithm for Network Intrusion Detection", Department of Computer Science and Engineering, Mississippi, State University, Mississippi State, Ms 39762.

[22] D. Hand, H. Mannila, and P. Smyth, "Principles of Data Mining."Cambridge, MA: MIT Press, Aug. 2001.

[23] M. Mahoney and P. Chan, "An analysis of the 1999 DARPA" LincolnLaboratory evaluation data for network anomaly detection," in Proc.Recent Adv. Intrusion Detect. (RAID), Pittsburgh, PA, Sep. 2003

[24] S. Bridges and R. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection," in Proc. Nat. Inf. Syst. Secur. Conf. (NISSC),Baltimore, MD, Oct. 2000, pp. 13–31.

[25] Q. Tran, H. Duan, and X. Li, "One-class support vector machine foranomaly network traffic detection," presented at the 2nd Netw. Res.Workshop 18th APAN, Cairns, Australia, Jul. 2004.

[26] D. Barbara, J. Couto, S. Jajodia, L. Popyack, and N. Wu, "ADAM: Detectingintrusions by data mining," in Proc. 2nd Annu. IEEE WorkshopInf. Assur.Secur., New York, Jun. 2001, pp. 11–16.

[27] R. Smith, A. Bivens, M. Embrechts, C. Palagiri, and B. Szymanski, "Clusteringapproaches for anomaly based intrusion detection," presented at the1st Annu. Walter Lincoln Hawkins Graduate Res. Conf., New York, Oct.2002.

[28] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, "A geometricframework for unsupervised anomaly detection: Detecting intrusions inunlabeled data," in Applications of Data Mining in Computer Security.Norwell, MA: Kluwer, 2002.

[29] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, "A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data," in Applications of Data Mining in Computer Security.Norwell, MA: Kluwer, 2002.

[30] D. Barbara, J. Couto, S. Jajodia, L. Popyack, and N. Wu, "ADAM: Detectingintrusions by data mining," in Proc. 2nd Annu. IEEE WorkshopInf. Assur.Secur., New York, Jun. 2001, pp. 11–16.

[31] C. Lu, D. Chen, and Y. Kou, "Algorithms for spatial outlier detection,"in Proc. 3rd IEEE Int. Conf. Data Mining, Melbourne, FL, Nov. 2003,pp. 597–600