Indian Journal of Science and Technology Vol. 4 No. 1 (Jan 2011) ISSN: 0974- 6846

A computational intelligence for evaluation of intrusion detection system

J. Visumathi[1] and K. L. Shunmuganathan[2]
[1]*Dept. of CSE, Jeppiaar Engineering College, Chennai, India*
[2]*Dept. of CSE, RMK Engineering College, Chennai, India*
jsvisu@gmail.com

Abstract

Intrusion detection system work at many levels in the network fabric and are taking the concept of security to a whole new sphere by incorporating intelligence as a tool to protect networks against un-authorized intrusions and newer forms of attack. Intrusion detection system is one of the widely used tools for defense in computer networks. In literature, plenty of research is published on Intrusion detection systems. In this paper we present a survey of intrusion detection systems. We survey the existing types, techniques and approaches of intrusion detection systems in the literature. We propose a new architecture for intrusion detection system and outline the present research challenges and issues in intrusion detection system using SVM classifiers. Finally we carry out our experiments based on our proposed methodology using DARPA (Defense advanced research projects agency) intrusion detection data set which is used for IDS evaluation.

Keywords: IDS, data mining, network, DARPA data set, SVM.

Introduction

An intrusion detection system (IDS) is a device or software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a management station. The purpose of IDS is to detect and prevent electronic threat to computer systems. The extensive use of the computers and availability of the Internet increase the impact of problem in size. In today's world everyone is connected over networks and many services are provided over the internet. This global reach increases the risk of intrusion threats from unknown sources. According to the computer emergency response team (CERT) 32,956 vulnerabilities were reported from many sources throughout 1995 until the first quarter of 2007 (Sujatha *et al.,* 2008). Intruder can use these vulnerabilities to launch an attack against computer network or servers. Two things are certain– intrusion detection is still a long way from being mature, and intrusion prevention technology is in its infancy.

Reasons for using intrusion detection system (IDS) are: to protect network from attack and abuse, to detect the violations in security and attacks on network, to document the existing threat to an organization and to get detail information about intrusions that occurred.

Basic approaches for intrusion detection system

Approaches for Intrusion detection systems can be broadly classified as: Signature based, Classification based and Anomaly based.

Signature based (misuse detection) approach

Most of the commercial IDSs are "misuse detection systems" which are designed to detect only known attacks. This approach uses a database of known attack signatures which is developed by experts and intrusion analyst. The traffic over the network or sequence of processes within the computer is compared to the entries in this database. If there is a match with database entries, the IDS system generates an alert message. Even though such a system does not generate false positives alerts, these systems cannot identify new and novel attacks (Hu Zhengbing *et al*., 2008; Ding *et al.,* 2009). There are two advantages of misuse detection approach: It is very effective for detecting the attacks without generating an overwhelming number of false alarms and it can quickly and reliably diagnose the use of a specific attack tool. On the other hand, the disadvantages of misuse detection approach are: It can only detect those attacks that have been described in the database and the database must be constantly updated with signatures of new attacks.

Classification-based intrusion detection approach

This approach uses normal and abnormal data sets of user behaviour, and uses data mining techniques to train the IDS system. This creates more accurate classification models for IDS as compared to signature-based approaches and thus they are more powerful in detecting known attacks and their variants.

Disadvantage of classification-based intrusion detection approach: it is still not capable of detecting unknown attacks.

Anomaly intrusion detection approach

The basic assumption of anomaly detection approach is that attacks are different from normal activity and thus they can be detected by IDS systems that identify these differences. Thus this approach begins with definition of desired form or behaviour of the system and then distinguishes between that desired behaviour and undesired or anomalous behaviour. The main problem is, defining the boundary between acceptable and anomalous behaviour. So, the anomaly detector approach must be able to distinguish between the anomaly and normal.

There are 2 types of anomaly detectors: 1.Static *anomaly*

Indian Journal of Science and Technology Vol. 4 No. 1 (Jan 2011) ISSN: 0974- 6846

detectors: It is based on the assumptions that there is a portion of the system being monitored that should remain constant and 2. *Dynamic anomaly detectors:* To characterize normal and acceptable behaviour a base profile is created by a dynamic anomaly intrusion system. Building the sufficiently accurate base profile is the main difficulty with the dynamic anomaly detection system.

The advantage of anomaly intrusion detection approach is: It is possible to detect unknown attacks. The disadvantage of anomaly intrusion detection approach is: Produces a large number of false alarms due to the unpredictable behaviours of users and networks. Therefore, large and accurate training data set is the major requirement of anomaly detection approaches to define the normal behaviour patterns.

Types of intrusion detection system

Network-Based IDS: Network-based IDS (Hu Zhengbing *et al.,* 2008; Su *et al.,* 2008) monitors network traffic using techniques like packet sniffing to collect network traffic data and tries to detect malicious activity such as denial of service attacks; port scans or even attempts to crack into computers.

Host-Based IDS: Host-based IDS (Sujatha *et al.,* 2008) monitors and analyzes system calls, application logs, file-system modifications and other host activities to identify the intrusion such as unauthorized remote login attempt, attempt to access unprivileged data. It normally works with Network-based IDS.

Protocol-Based IDS: Typically protocol-based IDS (Sangeetha *et al.,* 2008) are installed on a web server, and they are used for monitoring and analysis of the protocol in use of the computing system. If there is a deviation from intended behaviour of protocol then it can be detected as intrusion.

Graph-Based IDS: Graph-based IDS (Hassanzadeh & Sadeghian, 2008) concerned with detecting intrusions that involve connections between many hosts or nodes. A graph consists of nodes representing the domains and edges representing the network traffic between them.

Techniques for intrusion detection system

Neural networks (NNs): (Mussao de Lima *et al.,* 2008) can be trained to recognize arbitrary patterns in input data and associate such patterns with an outcome, which can be a binary indication of whether an intrusion has occurred. Such models are only as accurate as the data used to train them.

State transition tables: (Agrawal & Srikant, 1994; Lee *et al.,* 2008) describe a sequence of actions an intruder does in the form of a state transition diagram. When the behaviour of the system matches those states, an intrusion is detected.

Hidden Markov models (HMMs): (Lee *et al.,* 2008) are a stochastic version of the state transition techniques discussed above, where the states and transition probabilities are modeled as a Markov process with unknown parameters. A learning phase estimates these unknown parameters from the input data.

Artificial immune systems: (Dal *et al.,* 2008) are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving. The innate system of the human immune system can be compared with the misuse detection of the IDS; both uses pattern recognition respectively on memory cells and signatures database to detect intrusions. The adaptive system can be compared with the anomaly detection where both can detect yet unseen attacks and where their sensors have to go through a training phase.

Genetic algorithms (GAs): (Owais *et al.,* 2008) Genetic algorithms mimic the natural reproduction system in nature where only the fittest individuals in a generation will be reproduced in subsequent generations, after undergoing recombination and random change.

Decision tree: (Leet *et al.,* 2008) is a model of decisions and also can be used to show possible consequences for particular occurrences where there are conditional probabilities for each occurrence. Those occurrences of attacks form a tree-based structure that contains root node and a number of leaf nodes. Decision tree generally performs very efficiently even if dealing with a large amount of data.

Bayesian network: (Huijuan *et al.,* 2008) Bayesian Network is a graphical representation of the joint probability distribution function over a set of variables. The network structure can be represented in Bayesian Network as a Directed Acyclic Graph where each node represents a random variable and each edge between nodes shows the relation between nodes (i.e. relationship between variables). Individual invents which occurs during attack are represented as nodes in the graph and relationship between those events are represented as edges of the graph and this graph is then used to detect the intrusion.

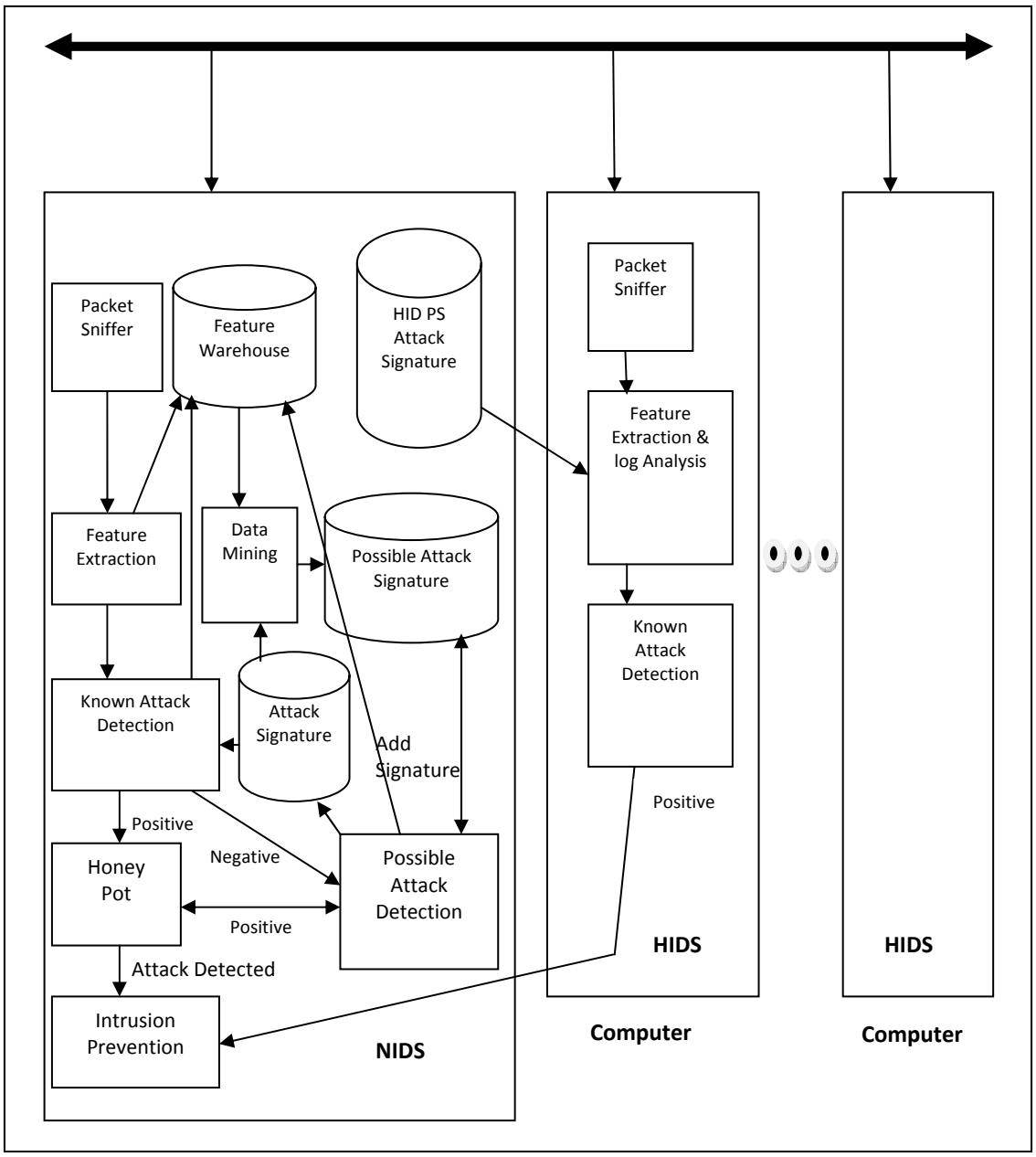
Fuzzy logic: (Su *et al.,* 2008) is a set of concepts and approaches designed to handle vagueness and imprecision. A set of rules can be created to describe a relationship between the input variables and the output variables, which may indicate whether an intrusion has occurred

Honeypot: (Khosravifar & Bentahar, 2008) is an unreal network system designed to trap crackers and intruders. The honeypot is used as bait in the form of a vulnerable system to trap hackers and keep them away from accessing the critical information in the main system. In this technique alarming adversaries, initially detected by the IDS, will be rerouted to a honeypot network for a more close investigation. If as a result of this investigation, it is found that the alarm decision made by the IDS of the agent is wrong, the connection will be guided to the original destination in order to continue the previous interaction. This action is hidden to the user. Such a scheme significantly decreases the alarm rate and provides a higher performance of IDS.

Data mining: (Hu Zhengbing *et al.,* 2008; Su *et al.*, 2008)

Indian Journal of Science and Technology Vol. 4 No. 1 (Jan 2011) ISSN: 0974- 6846

Fig. 1. System diagram.



is an analytic process designed to explore data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data.

Proposed architecture

Each type of IDS offers fundamentally different information-gathering, logging, detection and prevention capabilities. Each technology type offers benefits over the others such as detecting some events that the others cannot and detecting some events with significantly greater accuracy than the earlier technologies. In many environments, a robust IDS solution cannot be achieved without using multiple types of IDS technologies. For most environments a combination of network-based and host-based IDS technologies is needed for an effective IDS solution. Thus in our architecture we combined host-based and network-based IDS. Network-based IDS is used to detect Dos, DDoS and Probing attacks and Host based IDS are used to detect R2L and U2R attacks.

Using IDS based on data mining ((Hu Zhengbing *et al.,* 2008; Su *et al.,* 2008) is an effective method. IDS based on date mining have a behavioural model through widely checking data. So it can accurately capture the actual invasion and normal behaviour. This automated technique no longer needs manual analysis and manually coding the invasion mode and no longer needs to choose statistical methods by experience when build the normal behaviour using model. The major advantage of the data mining technique is that, it can be applied to multiple data stream. Many researchers have used fuzzy association rules effectively to design their NIDSs. Incremental fuzzy-rule mining can be very useful to meet the real-time requirements of IDS because it can produce the new rules set while detection process is going on (Su *et al.,* 2008).

1. Data warehouse is the most suitable data store for storing the data records gathered online from network. This will increase the speed of incremental fuzzy-rule mining algorithm and is the most suitable data store to analyze multiple data streams.
2. Using the honeypot technique, the system is able to

Indian Journal of Science and Technology Vol. 4 No. 1 (Jan 2011) ISSN: 0974- 6846

avoid many wrong decisions made by IDS. This will reduce the false alarm rate of the attack detection (Khosravifar & Bentahar (2008). Fig. 1 shows the block schematic of the proposed network intrusion detection system.

Feature data warehouse: It is used to store packet information extracted by feature extractor which is used to detect Intrusion.

Known attack signature database: It is used to store known attack signatures.

Possible attack signature database: It is used to store possible attack signatures which are predicted by using Known attack.

Data mining: (Agarwal *et al.,* 1993; Agarwal & Sant, 1994; Manila *et al.,* 1994) It uses attack signature database and feature data warehouse along with Apriori algorithm to predict possible attack signatures using existing attack signatures.

HIDPS attack signature database: Attack signatures for host based IDS are centrally stored at machine running NIDS.

Packet sniffer: It uses raw socket programming to fetch packets from network.

Feature extractor: It extracts information present within the packet such as, source IP address, destination IP address, values of flags present in packet header, etc... .

Known attack detector: Known attack detector module is used to detect network connections that correspond to attacks for which signatures are available.

Possible attack detector: It uses possible attack signature database to detect whether traffic matches with possible attack signature generated by data mining unit. If there is a match it forward that connection to honeypot to detect whether there is an intrusion or not.

Honeypot: It is used to detect whether the connection is trying to do intrusion in the network or not.

Algorithms

Possible_attack_signature_algorithm:

Input: Attack signature database (ASDb)

Output: Possible attack signature database (PASDb)

Steps:

1. (Agarwal *et al.,* 1993; Agarwal & Sant, 1994; Manila *et al.,* 1994) Apply Apriori algorithm on feature data warehouse to generate patterns set
2. For each Pattern in patterns set
   a. For each Signature in Known Attack Signature set
      i. Calculate Similarity between pattern and signature
      ii. If (Similarity > 0.9)
1. Add pattern to possible attack signature
3. Stop

Known_attack_detection_algorithm:

Input: Network traffic feature, attack signature database

Output: Traffic classification (Norma/Attack)

Steps:

1. For each signature in known signature set
   a. If(Traffic feature matches with signature)
      i. Forward corresponding connection to intrusion prevention module
      ii. Mark corresponding entry in feature data warehouse for attack
   b. Else
      i. Forward network traffic feature to possible attack signature detector

Possible_attack_detection_algorithm:

Input: Network traffic feature, possible attack signature database

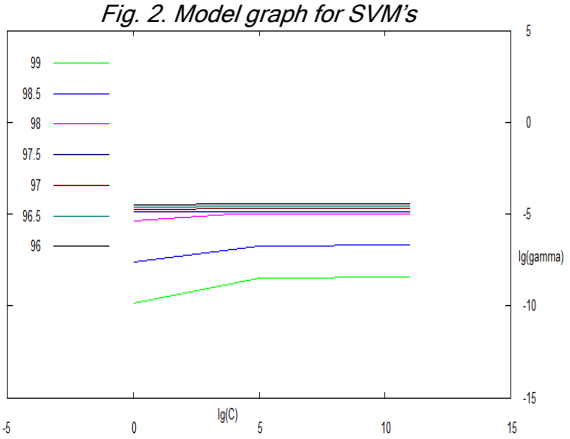
Output: Traffic classification (Norma/Attack)

Steps:

1. For each signature in possible signature set
   a. If(Traffic feature matches with signature)
      i. Forward corresponding connection to honeypot module to detect intrusion
2. If (Result from honeypot is positive)
   a. Remove corresponding signature entry from possible attack signature database
   b. Add removed signature to known attack signature database

   Else
   c. Remove corresponding signature entry from possible attack signature database
3. Mark corresponding network traffic feature entry in feature data warehouse for attack.

Fig. 2. Model graph for SVM's



SVM classifiers:

Support vector machines (SVM) are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. SVM delivers a unique solution, since the optimality problem is convex. This is an advantage compared to neural networks, which have multiple solutions associated with local minima and for this reason may not be robust over different samples. SVM are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. Since SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a

Indian Journal of Science and Technology Vol. 4 No. 1 (Jan 2011) ISSN: 0974- 6846

model that predicts whether a new example falls into one category or the other (Fig.2).

. Experiments on real time data set

Experiments were carried out based on our proposed methodology using a well know DARPA Dataset (Mukkamala *et al.,* 2004).

A sub set of the DARPA intrusion detection data set is used for off-line analysis. In the DARPA intrusion detection evaluation program, an environment was set up to acquire raw TCP/IP dump data for a network by simulating a typical US air force LAN. The LAN was operated like a real environment, but being blasted with multiple attacks. For each TCP/IP connection, 41 various quantitative and qualitative features were extracted (Huijuan *et al.,* 2008). The 41 features extracted fall into 3 categorties, "intrinsic" features that describe about the individual TCP/IP connections; can be obtained form network audit trails, "content-based" features that describe about payload of the network packet; can be obtained from the data portion of the network packet, "traffic-based" features, that are computed using a specific window (connection time or no of connections). As DOS and probe attacks involve several connections in a short time frame, whereas R2U and U2Su attacks are embedded in the data portions of the connection and often involve just a single connection; "traffic-based" features play an important role in deciding whether a particular network activity is engaged in probing or not. Attack types fall into four main categories: 1. Probing: surveillance and other probing, 2. DOS: denial of service, 3. U2Su: unauthorized access to local super user (root) privilege and 4. R2L: unauthorized access from a remote machine.

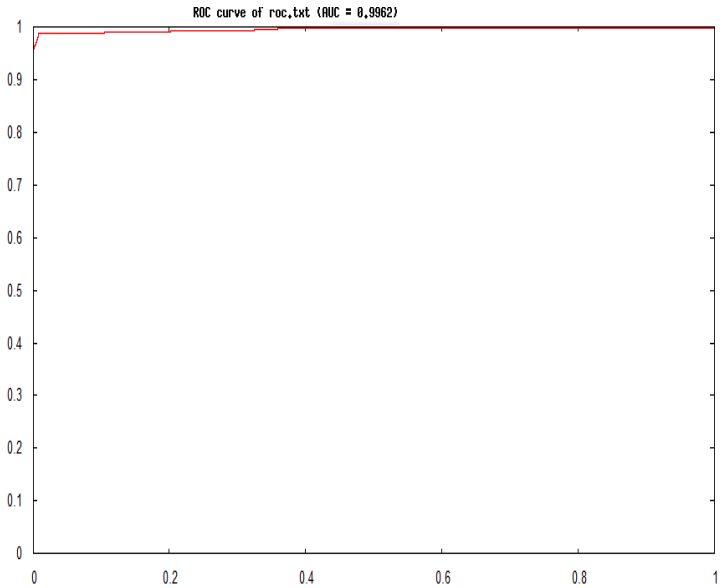
Table 1. Accuracy obtained using various machine learning methods.

Machine learning method	Accuracy
Naïve bayesian	74
Bayesian network	99
Logistic regression	99.87
SVM	97.25
RBF network	97.98
Multilayer perceptron	99.85
SMO	99.55
KNN	99.92
Random forest	99.97

The dataset consists of 11982 data points of which 7204 are attacks and 4778 are normal connections. There are 41 features for each data point. Result obtained using different machine learning techniques is given in Table 1. We used various machine learning techniques for our analysis. We show the performance of our method using receiver operating characteristic (ROC) curves. ROC curves are generated for SVMs by considering the rate at which true positives accumulate versus the rate at which false positives accumulate with each one corresponding, to the vertical axis and the horizontal axis in Fig. 3. The point (0, 1) is the perfect classifier, since it classifies all positive and negative cases correctly. Thus an ideal system will initiate by identifying all the positive examples and so the curve will rise to (0, 1) immediately, having a zero rate of false positives, and then continue along to (1, 1).

Fig. 3 gives the ROC curve obtained using current

Fig.3. Receiver operational characteristic curve using SVM's.



modified feature set. Detection rates and false alarms are evaluated for DARPA dataset described in section 6 and the obtained results are used to form the ROC curves. In each of these ROC plots, the x-axis is the false alarm rate, calculated as the percentage of normal considered as attacks; the y-axis is the classification rate, calculated as the percentage of attacks. A data point in the upper left corner corresponds to optimal high performance, i.e., high classification rate with low false alarm rate. We can see from ROC that Area under Curve (AUC) for modified feature set is higher than for previous features.

Challenges and issues

With best of our knowledge many researchers have proposed new architecture for intrusion detection system but did not comment on how their architecture will accept in real time environment. Further many of them did not marked that how much load their architecture will create on executing platform. (Future scope of our paper will include that part).

Conclusion and future scope

This paper reviews and tried to summarize different types, methods and approaches for intrusion detection system and also provides a strong platform to detect anomalies. Further this paper has proposed a new architecture for intrusion detection system which generates and test new signatures for intrusion detection without the interference of third party. Experimental results are carried out by DARPA data set. The proposed model is in its initial stage where an initial algorithm is

Indian Journal of Science and Technology Vol. 4 No. 1 (Jan 2011) ISSN: 0974- 6846

proposed. The future step for this proposal is under development where the real time analysis is going on.

References

1. Agrawal R and Srikant R (1994) Fast algorithms for mining association rules. Proc. of the 20th VLDB conf., Santiago, Chile. pp.487-499.
2. Amin Hassanzadeh and Babak Sadeghian (2008) Intrusion detection with data correlation relation graph. IEEE, The Third Intl. Conf. on Availability, Reliability and Security. pp.982-989.
3. Bane Raman Raghunath and Shivsharan Nitin Mahadeo (2008) Network intrusion detection system. IEEE, First Intl Conf. on Emerging Trends in Engg. & Technol. pp:1272-1277.
4. Creation and Deployment of Data Mining-Based Intrusion Detection Systemsin Oracle Database 10g. http://www.oracle.com/technology/products/bi/odm/pdf/odm_based_intrusion_detection_paper_1205. pdf
5. Divyata Dal, Siby Abraham, Ajith Abraham, Sugata Sanyal and Mukund Sanglikar (2008) Evolution induced secondary immunity: An artificial immune system based intrusion detection system. IEEE, 7th Computer Information Systems & Industrial Management Applications.pp:65-70.
6. Do-hyeon Lee, Doo-young Kim and Jae-il Jung (2008) Multi-Stage intrusion detection system using hidden Markov model algorithm. IEEE, Intl. Conf. on Information Sci. & Security. pp:72-77.
7. Heikki Manila, Hannu Toivonen and A. Inkeri Verkamo (1994) Efficient algorithms for discovering association rules. In: Knowledge Discovery in Databases (KDD'94). Fayyad UM & Uthurusamy R (Eds.), AAAI Press. p:81-192.
8. Joong-Hee Leet, Jong-Hyouk Leet, Seon-Gyoung Sohn, Jong-Ho Ryu, and Tai-Myoung Chungt (2008) Effective value of decision tree with KDD 99 intrusion detection datasets for intrusion detection system. IEEE, ISBN: 978-89-5519-136-3.
9. Juan Wang, Qiren Yang and Dasen Ren (2009) An intrusion detection algorithm based on decision tree technology. IEEE Asia-Pacific Conf. on Information Processing. ISBN: 978-0-7695-3699-6. pp:333-335.
10. Khosravifar B and Bentahar J 2008) An experience improving intrusion detection systems false alarm ratio by using honeypot. IEEE, 22nd Intl. Conf. on Advanced Information Networking and Applications.pp: 997-1004.
11. Kola Sujatha P, Kannan A, Ragunath S, Sindhu Bargavi K and Githanjali S (2008) A behaviour based approach to host-level intrusion detection using self-organizing maps. IEEE, First Intl. Conf. on Emerging Trends in Engg. & Technol. pp:1267-1271.
12. Lgor Vinicius Mussoi de Lima, Joelson Alencar Degaspari and Jo˜ao Bosco Mangueira Sobral (2008) Intrusion detection through artificial neural networks. IEEE, ISBN: 978-1-4244-2066-7.pp:867-870.
13. Lu Huijuan, Chen Jianguo and d Wei Wei (2008) Two stratum Bayesian network based anomaly detection model for intrusion detection system. IEEE, Intl. Symp. on Electronic Commerce & Security.pp:482-487.
14. Marimuthu and A. Shanmugan (2008) Intelligent progression for anomoly intrusion detection. IEEE, ISBN: 978-1-4244-2106-0. pp:261-265.
15. Mukkamala S, Sung AH and Abraham A (2004) Computational intelligent techniques for detecting denial of service attacks. Proc. of Innovations in Applied Artificial Intelligence, 17th Intl. Conf. on Industrial & Engg. Appl. of Artificial Intelligence & Expert Systems (IEA/AIE), Lecture Notes in Computer Science 3029 Springer, ISBN 3-540-22007-0, pp: 633-642.
16. Owais S, Snasel V, Kromer P and Abraham A (2008) Survey: Using genetic algorithm approach in intrusion detection systems techniques. *CISIM 2008, IEEE.* pp:300-307.
17. Rakesh Agrawal, Arun Swami and Tomasz Imielinski (1993) Mining association rules between sets of items in large databases. Proc. of the 1993 ACM SIGMOD Conf. Washington DC, USA, May 1993. pp:1-10.
18. Robert, Richardson (2007) Computer crime and security survey. http://i.cmpnet.com/v2.gocsi.com/pdf/ CSISurvey2007.pdf.
19. Sangeetha S, Vaidehi V, Srinivasan N, Rajkumar KV, Pradeep S, Ragavan N, Sri Sai Lokesh C, Subadeepak I and Prashanth V (2008) Implementation of application layer intrusion detection system using protocol analysis. IEEE-Intl Conf. on Signal processing, Commun. & Networking .pp:279-284.
20. Su MY, Chang KC, Wei HF and Lin CY (2008) A real-time network intrusion detection system based on incremental mining approach. IEEE.pp: 76- 81.
21. Ya-Li Ding, Lei Li and Hong-Qi Luo (2009) A novel signature searching for intrusion detection system using data mining. IEEE 8th Intl. Conf. on Machine Learning & Cybernetics. ISBN: 978-1-4244-3703-0. pp:122-126.
22. Youssif Al-Nashif, Aarthi Arun Kumar, Salim Hariri, Guangzhi Qu, Yi Luo and Ferenc Szidarovsky (2008) Multi-Level intrusion detection system. IEEE, Intl. Conf. on Automonic Computing. pp:131-140
23. Zhan Jiuhua (2008) Intrusion detection system based on data mining. IEEE, Workshop on Knowledge Discovery and Data Mining, ISBN:978-0-7695-3090-1. pp:402-405.
24. Zhengbing H, Zhitang Li and Junqi W (2008) A novel network intrusion detection system (NIDS) based on signatures search of data mining. IEEE, Workshop on knowledge discovery and data mining. pp:1-7.