Intrusion Detection System using Anomaly Detection Based on outlier Approach

Gurpreet Kour Khalsa[1], Amit Sharma[2]

[1,2]Eternal University, Baru Sahib, India

***Abstract* - An Intrusion Detection System (IDS) is a software application that monitors the system or activities of network for malicious activities and generates reports to the management system. The main focus of Intrusion detection systems (IDS) is to identify the possible incidents, logging information about them and to report such attempts. IDS have become an essential addition to the security infrastructure of every organization. The main goal of an intrusion detection system is to detect the attacks efficiently. Furthermore, it is equally important to detect attacks at a beginning stage in order to reduce their impacts. Computer intrusion is unauthorized access to data, computer or internet service. With greater dependency on computers there is a need of prevention and detection of such frauds in order to prevent malicious activities which can hamper the computing resources. Increase in fraud has resulted in loss of billions of dollars worldwide. In order to detect and prevent frauds several techniques are continually evolved and applied to many systems. This paper proposes intrusion detection system that uses anomaly detection technique based on outlier approach to predict attacks in web applications. The experimental results proved that the proposed approach identifies the anomalies very effectively than any other approaches.**

***Index Terms* - Fraud detection. Intrusion. Fraud prevention. Anomaly detection.**

I.INTRODUCTION

With the advancement in internet, security of network has become the key principle. It is very hard to design completely secure system. A secure computer system protects its data and resources from unauthorized access, tampering, denial of use, fabrication, masquerading. Increased connectivity has provided faster access to data than ever before, but it also provided an access path to the data from virtually anywhere on the network leading to fraud. In general, there is a flow of data from a source to a destination over the channel. The security system should restrict access to information over channel only to those parties that are authorized to have access, according to the security policy in use. The Association of Certified Fraud Examiners (ACFE) defined fraud as “the use of one’s occupation for personal enrichment through the deliberate misuse or application of the employing organization’s resources or assets [1].” It can be defined as an intentional deception or misrepresentation of information which results in some unauthorized benefit. Fraud detection is primarily considered to be a classification problem, but with a vast imbalance in fraudulent to legitimate transactions misclassification is common and can be significantly costly [2].E-commerce applications has become primary target for such frauds. Several security improvement methods are developed for e-commerce applications. There exist some sort of vulnerability in websites which can lead to fraud. Several security improvements are installed in e-commerce web applications but still their exist some sort of vulnerability which can be exploited to commit fraud. Currently, computer systems play very major role in society and its economy. With increase in technology part of data is migrated and distributed in nature. This migration of data has lead to many problems like data duplication, misplacing of data, mismatching of data and missing value. Thus making it difficult to find required data in right time. Computers have become target of malicious threats that turn into intrusions. This is the reason computer security has become vital concern for network. Different security mechanisms are used to enforce the security properties defined in a given security policy. Depending on the type of attack, different means have to be adapted to handle such attack. Intrusions can cause disaster in computer network. Thus, some proactive measures should be taken to deal with intrusions. Network security is any activity designed to protect the usability and integrity of your network and data[3].Network Security targets a variety of threats and stops them from entering or spreading on your network. Thus it manages access to network. Intrusion detection is one of the core areas of computer security whose primary objective is to identify malicious activities in network and protect from attacks. IDS act as a tool which monitors events in a system or network and analyzes them for signs of security attacks. An intrusion prevention system (IPS) scans network traffic to actively block attacks. Cisco Next-Generation IPS(NGIPS) appliances do this by correlating huge amounts of global threat intelligence to not only block malicious activity but also track the progression of suspect files and malware across the network to prevent the spread of outbreaks[4]. An intrusion detection system inspects all inbound and outbound network packets and identifies suspicious patterns that may indicate a network or system attack. The rest of the paper is organized as follows. In Section 2, the literature review is presented. In Section 3, a brief overview of intrusion detection system is presented, the proposed IDS architecture is described, along with the considered attack types and outlier approach. In Section 4,simulation results are presented. Finally, Section5 concludes the paper.

II. LITERATURE REVIEW

This section deals with the work done by researchers in the field of intrusion detection system. In the early days of computers, very rarely automated tools were used to break into systems. Hackers were very intelligent with their own expertise skills to perform such actions. The present situation is completely different. There are different tools and techniques used to exploit scripts. Intrusion detection concept was introduced in early 1980's after the evolution of internet with surveillance end monitoring the threat [5]. Intrusion detection is the process of monitoring events occurring in a computer system or network and analyzing them for signs of intrusions. The goal of intrusion detection is to monitor the network assets to detect anomalous behavior and misuse in network [6]. Duan et al. [7] have concentrated on identifying compromised machines that are recruited to detect spam zombies. In this approach sequential outgoing messages are scanned using probability ratio test. This method determines whether the security of host is being compromised or not. Kim.Chan.Ae[8], propsed digital forensics techniques to analyze system intrusion incidents to detect anomaly transactions that may occur in the user environment during online financial transactions and the risk point calculation model is proposed by scoring anomaly transaction cases in the detection step by items. Anderson [9] suggested an intrusion detection method to efficiently detect the intrusion. An Intrusion Detection Mechanism using Time- series, Markov chains, and statistics was developed by Denning [10].It states that if there are changes in the normal behavior it is considered as anomalous. Abdullah [11] and co-workers elaborated intrusion detection classification rules using genetic algorithms. Several machine learning techniques are used for intrusion detection. Existing machine learning techniques (Artificial Neural Networks - ANN) for intrusion detection was described by Roshani team [12]. Devikrishna et al [13] used MLP (Multi-Layer Perceptron) architecture for intrusion detection that detects and classifies attacks into six types. In the earlier works on intrusion detection system, Hasina A. Razzak. [14] has proposed a Methodological Approach given form implement Intrusion Detection System in Hybrid Network. This approach resolved ambiguities in network by prevention policy. Several approaches are given for hybrid intrusion detection system. Multiple sensors are placed on host or particular network segment. Sensor positioning is very important as it observes or read the packets, thus captures all the information about the packets. Signature database record enables IDS to have a set of signature, criteria or rules against which flow of packets is compared. These packets in the network are compared with signature using pattern matching algorithm by analyzer. If analyzer finds any match it sends appropriate alert message for known attack. In 1983, SRI International and Dorothy Denning began working on a government project that launched a new effort into intrusion detection system development [15]. Around 1990s the revenues are generated and intrusion detection market has been raised. Real secure is an intrusion detection network developed by ISS. After a year, Cisco recognized the priority for network intrusion detection and purchased the Wheel Group for attaining the security solutions [15]. The government actions like Federal Intrusion Detection Networks (FID Net) were designed under Presidential Decision

Directive 63 is also adding impulse to the IDS [15].To identify various types of network intrusions in network several genetic algorithms can be implemented. The genetic algorithm [16] is applied to achieve set of classification rules from network. Audit data which acts as fitness function is applied to check quality of each rule. Then these rules are used to detect network intrusion or types of attack. In [17] Hoque, Mukit and Bikas presented an implementation of Intrusion Detection System by applying the theory of genetic algorithm to efficiently detect various types of network intrusive activities. In this approach realistic detection rate was obtained. The standard used was KDD99 intrusion detection benchmark dataset was used. Standard deviation equation was also used to measure the efficiency of network. In [18] Zadeh initiated the idea of fuzzy set theory and it was mainly intended mathematically to signify uncertainty and vagueness with formalized logical tools for dealing with the vagueness connected in many real world problems. Fuzzy logic identifies normal and abnormal behavior in computer networks. This concept is also used to reduce fake alarm rate of determining intrusion detection system.

III. INTRUSION DETECTION SYSTEM

Intrusion is defined as any set of actions that attempt to compromise the integrity, confidentiality and availability of resource. Intrusion is possibility of violation of security policy of the system. Security policies are violated because there is exploitation of vulnerability due to intrusion. Unauthorized attempt is made to access or manipulate information. Intrusion detection (ID) is the process of identifying and responding to malicious activities targeted at computing and network resources. Detection of Intrusion attempts to detect the attacks of computer by examining different information records observed in network processes [19] [20]. Thus attacks are needed to be detected very efficiently. Intrusion detection system is application or software which monitors the network in order to detect unauthorized use of information and resources. At times even authorized users can become corrupt. Intrusion Detection System can be considered as a security operation that complements protection, e.g., firewalls [21]. Many of the IDS research studies have been done in order to improve the detection stability and detection precision [22]. Detection stability and detection precision are two key indicators used to evaluate IDS (Intrusion Detection System) [23]. Intrusion detection system automate the process of monitoring events occurring in a computer system or network using software or hardware. IDS tools detect such attacks and alert the proper individual. It issues some warning signal whenever any intruder is trying to misuse the data. Intrusions are caused by attackers accessing the systems from internet, users who want to gain additional privileges. The main functions served by IDS are identifying possible incidents, monitoring information about those incidents, trying to stop them and report such incidents to security administration.

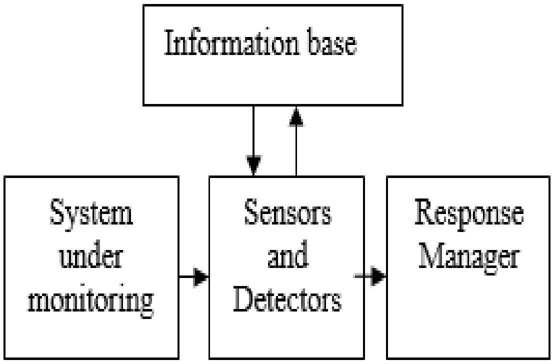

Fig.1. IDS

Intrusion detection system is of two types host based and network based. In host-based system the data source which is used for detection is in form of logs, system calls and passwords. Agents on host identify intrusions by analyzing data sources. Host based IDS can sometimes tell exactly files which are open, commands which intruder ran, system calls executed. It provides detailed information about intrusion, but it is used where broad intrusion detection is not needed or bandwidth is not available for communication. In network-based system data source acts as packets which are flowing in the network. Sensors capture the incoming traffic flow and analyze the individual packets for intrusive activities. This system is much more self-contained as it runs on a dedicated system which is easy to install, some configuration is required, plug it in network location which allows monitoring of traffic.

The approach used for IDS can be classified into two categories: misuse detection and anomaly detection. Misuse detection observes intrusions in form of signature. Signature is a pattern of activity which corresponds to intrusion. IDS identifies intrusion by looking forward for those patterns in data being analyzed. If match is find it sends appropriate alert message for known attack. Misuse approaches include expert systems, model-based reasoning, state transition analysis, and keystroke dynamics monitoring [24]. Very specific attacks can be detected using this system whereas known attacks are detected with reliable low false alarm rate. All different attacks cannot be detected in this system because it is based only on know patterns. Anomaly detection compares current network activities against statistical models for past behavior. In this system for every user a historical profile is created and then large deviation from that profile is used to detect intrusion. Statistical model is not tied with any sort of pattern or prediction. In this approach sometimes legitimate users are also considered anomalous because this scheme gives high rate of false alarm. The anomaly detection has two phases; training and detection. In training phase legitimate traffic and characteristics of normal usage are recorded. The goal of this phase is to define how normal accepted traffic looks and using this traffic dynamic set of rules are created .During detection, the passing traffic is compared against the ruleset created during the training phase, and any deviations from this rule-set will be marked as anomalous. The basic assumption is that difference between anomalous behavior and normal behavior can be expressed quantitatively or qualitatively. Several techniques are used to extract information from an HTTP request which can be later applied to profile the behavior of normal traffic in web applications.

Architecture

The main aim of this paper is to develop an IDS based on anomaly detection model that would be precise, not easily cheated by small variations in patterns, low in false alarms, adaptive and be of real time. The packets are received form internet and then data sets are collected. Then features are extracted from datasets which are further passed to Intrusion Detection System. The feature extracted include source and destination IP addresses, source and destination ports, protocol, flags, number of bytes and number of packets, window time for fast scanning activities. As network activities are processed, the system periodically generates values which are considered as measure of abnormality. Then this data is fed into proposed anomaly detection module which uses outlier approach to assign profile behavior to each network connection. These profiles need to be updated for every audit record. In this step anomalous connections are also analyzed to determine if they are actual attacks or other interesting behavior. Sometime legitimate behavior of network is considered as attack.

Types of Attack

Cross site tracing attack involves use of Trace. Trace method is used to steal legitimate user's credentials. This attack begins when unwanted user visits site hosted by server. The compromised server sends scripting code to victim computer. Trace request is sent to other site recently visited by computer. This site further sends cookies to hacked server and makes data available to attacker.

Probing is an attack in which hacker scans machine or networking devices in order to determine weaknesses or vulnerabilities to exploit the system. This technique is commonly used in data mining e.g. saint, portsweep, mscan, nmap etc.[25].Probe is also considered as an action taken for the purpose of learning something about the state of network.

DOS attack makes service unavailable by flooding it with traffic from different sources. Hacker makes resources or memory too busy to serve legitimate requests and denying users to access to a machine. Apache, smurf, Neptune, ping of death, back, mail bomb, UDP storm etc. are all DoS attacks.

Backdoor attacks allow computers to be accessed remotely. They are designed in such a way that intrusion detection systems can be bypassed. Both hardware and software components can allow hackers access through malicious backdoors. Several attack strategies, including port binding, connect-back, and connect availability use can be employed through backdoors.

Feature	Attack
Request Method	Cross Site Tracing (XST)
HTTP Version	Probing
Remote access	Backdoor attack
Pages unavailable	DOS attack

Fig.2.Summary of profiles used by system

Outlier approach

IDS module assigns a degree of outlier to each point which is called local outlier factor (LOF)[26].The degree of being outlier is measured with respect to its neighborhood. The LOF is the average of the ratio of the density of point and density of its neighbor. The density of the neighbors of a given instance plays a key role. In order to consider an instance outlier or non outlier, LOF is computed for each instance. In order to formalize the algorithm to detect density based local outliers following definitions are needed.

1. For each instance x, calculate the k-distance which is the nearest neighborhood denoted by k-distance(x).
2. Next, calculate the reachablity distance of instance x with respect to instance y defined as-
   reach-distance(x,y) = max{k-distance (y), d(x,y)},
   where d(x,y) is the distance between instance x and instance y.
3. Then, calculate local reachability density for each x , inverse of the average reachability distance is based on the MinPts (minimum number of objects) data example x and its nearest neighbors.
4. Calculate LOF to all instance x as an average of the instance x local reachability density ratios and local reachability density of x's MinPts nearest neighbors.

Using this approach outlier value is calculated and thus distance is calculated between extracted features and trained model. The advantage of this scheme is clustering. Group of data instances of the training set are grouped together into clusters using a simple distance-based metric. Clusters are defined as maximal sets of density-connected objects. Once the data is clustered, outlier approach is used to classify these clusters either as anomaly or normal instances. After the clustering process the system can accept the network data instances and thus classify them as possible threats or safe to pass instances, thus detecting possible intrusions. In clustering, threshold value is set and if outlier value is greater then the specified threshold it will generate false alarm. Thus deviation from normal functioning identifies attack. After implementing mathematical models, the system has been analyzed to find its correctness. Then alarm needs to be send. It defines about the attack and reaction of the system. System administrator is informed with all the required data so that particular action should be taken. IDS is administered to monitor the host and to respond the report as an alert.

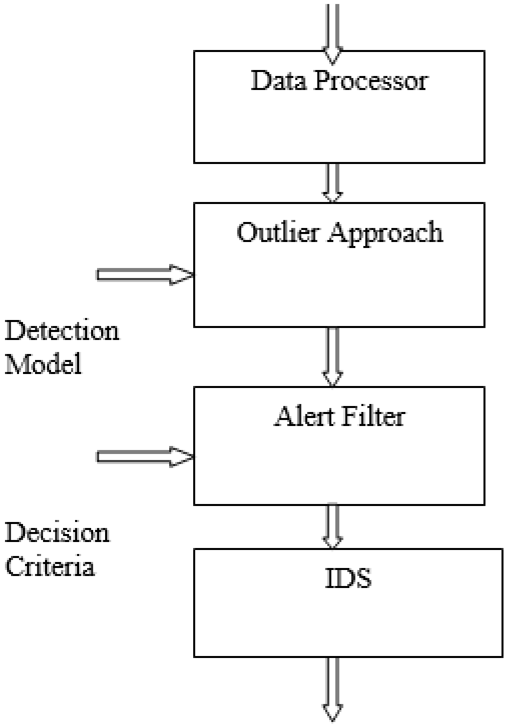

Fig.3.Functionality of Proposed IDS

Thus roles and responsibilities for analyzing and monitoring outcomes of both manual and automatic responses is established.

IV. RESULT AND DISCUSSION

In order to implement the algorithm and measure performance, number of experiments were conducted on datasets. In this paper data processing is carried out on KDD Cup 99. The extracted data is evaluated in two phases training and testing. Hence simulation is prepared with target machines running various services and operating system. Connection is followed by TCP packets with beginning and ending at specified periods. Attacks are categorized into four types: cross site tracing, Denial of service (DOS) and Probing. Training set consist of 129595 connection records and test data consist of 305779 connection records. Following table shows the type of intrusion in datasets.

Dataset	Cross Site tracing	Probing	DOS
Train	87280	3107	39148
Test	50594	3177	251758

Fig.4. Intrusion distribution in dataset

In training phase dataset extracted is used then filtered from any malicious attack to be considered as a legitimate traffic. Training phase learns characteristics from requests and profiles are stored in xml files to get switched to detection mode. To graphically represent the anomaly counts that are obtained via anomaly detection system, charts were generated. These charts represent the execution time to calculate the distance and outlier values to determine the anomalies in system. As anomaly detection rate depends on outlier value. The distance calculation is dependent on execution time. The given fig 5. gives the overview of execution time and dataset. The execution time of outlier detection is less than all other approaches.

The fig.6. gives overview of outlier values and dataset. Anomaly detection rate is dependent on outlier values. Higher the outlier value more the dataset is assumed as intrusion detection.

Than Using outlier detection method , distance and outlier values are calculated. Outlier value increases if distance between extracted and trained model increases. The results are shown in the figure 7.

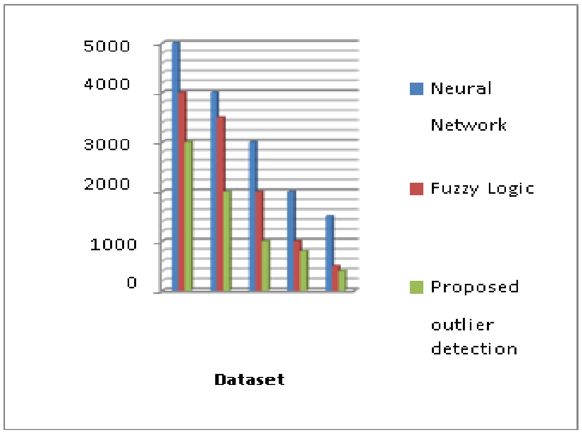

Fig.5.Dataset versus Execution time

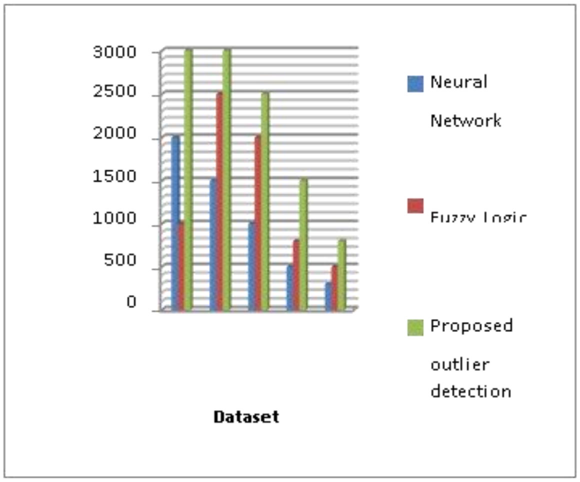

Fig.6.Outlier value versus dataset

The ability of an Intrusion Detection System can be measured based on performance metrics. The metrics defined in this area can be analysis of compromise, operational performance impact and timeliness. In analysis of compromise the extent of damage and compromise due to intrusions can be reported. Operational performance impact can calculate the negative impact due to the operation of IDS can be calculated in the form of processing power. Timeliness gives the average time between an intrusion's occurrence and its reporting.

ID	Distance	Outlier Value
1.	2.2	4
2.	4.5	7
3.	3.4	6
4.	5.5	10
5.	2.4	5

Fig.7. Distance and Outlier Value

V. CONCLUSION & FUTURE WORK

In this paper we have presented the details of approach called outlier detection to detect intrusions in the network. The overall objective is detecting attacks and threats against computer systems and it giving satisfactory results compared to other approaches. The performance of proposed IDS is better than that of the considered neural network-based and fuzzy-based benchmark schemes. Improvement is very important as there is requirement for such algorithms to build reliable intrusion detection system. Further the proposed work can be used to improve the efficiency of Intrusion Detection System. The proposed work can be possibly used for various distance computation function between the trained model and testing data model. Thus, our research work can be considered to improve the efficiency of Intrusion Detection System in a better manner.

REFERENCE

[1] Investigating Fraudulent Acfi, http://www.uhsa.uh.edu/samiAM/01C04.hhll, 2000.

[2] E.Duman and MH. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search" Expert Systems with Applications, 38, 13057-63.

[3] What is network security. Cisco technical report, https://www.cisco.com/c/en/us/products/security/what-is- network-security.html

[4] Next Generation Intrusion Prevention System. https://www.cisco.com/c/en/us/products/security/ngips/index.html

[5] S.A. Asmaa, G.Sharad ," Importance of Intrusion Detection System (IDS)", International Journal of Scientific & Engineering Research, Vol. 2, Issue 1, Jan 2011.

[6] I.Paul, "The Evolution of Intrusion Detection Systems", Tetrad Digital Integrity, LLC.

[7] Z. Duan, P.Chen, F. Sanchez, Y. Dong, M. Stephenson and J. M. Barker," Detecting spam zombies by monitoring outgoing messages", IEEE Trans. Dependable and Secure Computing, 9(2):198–210, Apr 2012.

[8] H.D. Lee, H.W. Park, and S. Kim, "Fraud and financial crime detection using malware forensics", Springer Multimedia Tools Appl, pp 479-496, 2013.

[9] K.Shah, N.Dave, S.Chavan, S.Mukherjee, A.Abraham ands. Sanyal, "Adaptive neuro-fuzzy intrusion detection system", IEEE International Conference on Information Technology: Coding and Computing (ITCC'04), vol. 1. USA: IEEE Computer Society;2004; 70–74.

[10] D. Anderson, T. Frivold and A. Valdes,"Next-generation intrusion detection expert system (NIDES): A summary Technical Report", Computer Science Laboratory, SRI International, SRI–CSL–95–07, May 1995.

[11] B. Abdullah, I. Abd-algafar G.I. Salama and A. Abd-alhafez ,"Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System" Proceedings of 13th International Conference on Aerospace Sciences and Aviation Technology (ASAT-13), Military Technical College, Cairo, Egypt, 2009;1-5.

[12] G. Roshani , C. Vaidya, and M. Raghuwanshi, "Survey. Learning Techniques for Intrusion Detection System (IDS)", International Journal of Advance Foundation and Research in Computer (IJAFRC), Feb 2014. ISSN 2348 – 4853, 2014;1(2).

[13] K.S. Devikrishna, and B.B . Ramakrishna,"An Artificial Neural Network based Intrusion Detection System and Classification of Attacks", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622, Jul-Aug 2013

[14] A. Hasina, A. Razzak, Karim, S.S. Handa, M.V. M .Ramana" A Methodical Approach to Implement Intrusion Detection System in Hybrid Network" IJESC, 2017, Volume 7 Issue No.3.

[15] Paul Innella- "The Evolution of Intrusion Detection Systems-Tetrad Digital Integrity", LLC. International Journal of Security, Privacy and Trust Management (IJSPTM) ,Vol 4, No 1, February 2015

[16] J. Gomez & D. Dasgupta, "Evolving Fuzzy Classifiers for Intrusion Detection", IEEE Proceedings of the IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY. (2002)

[17] S.H. Mohammad, M.Abdul & M.B. Abu Naser "An Implementation of Intrusion Detection System using Genetic Algorithm", International Journal of Network Security and Its Applications (IJNSA),Vol.4, No.2, March, pp. 109-120, 2012.

[18] L.A. Zadeh, "Fuzzy Sets", Information and Control, Vol.8, pp. 338-353, (1965) IDS

[19] J.P. Anderson, "Computer security threat monitoring and surveillance" Technical Report, Fort Washington, PA, USA.,1980;911.

[20] C. Endorf, E. Schultz, and J. Mellander, "Intrusion detection and prevention" California: Mc Graw-Hill, (2004).

[21] S.A. Asmaa, G.Sharad ," Importance of Intrusion Detection System (IDS)", International Journal of Scientific & Engineering Research, Vol. 2, Issue 1, Jan 2011.

[22] A. Patcha and J.M. Park "An overview of anomaly detection techniques: Existing solutions and latest technological trends" Computer Networks, 51(12), 3448–3470, 2007.

[23] L. D. S. Silva, A. C.Santos, T. D .Mancilha, J. D. Silva, and A. Montes, "Detecting attack signatures in the real network traffic with ANNIDA", Expert Systems with Applications, 34(4), 2326–2333 , 2008.

[24] S. E. Smaha and J. Winslow, "Misuse detection tools", In Computer Security Journal IO(I), pages 39 - 49, Spring 1994

[25] B. Markus , P.K. Hans, T. Ng . Raymond, and J.Sander, "Identifying density based local outliers" In Proceedings of the ACM SIGMOD Conference, Dallas, TX, 2000.

[26] S.H. Mohammad, M.Abdul & M.B. Abu Naser "An Implementation of Intrusion Detection

System using Genetic Algorithm", International Journal of Network Security and Its Applications (IJNSA),Vol.4, No.2, March, pp. 109-120, 2012.