ISSN 2047-3338

Artificial Intelligence Technique Applied to Intrusion Detection

K. P. Kaliyamurthie[1] and Dr. R. M. Suresh[2]

[1]Department of IT, Bharath University, Chennai, India
[2]CSE, RMD Engineering College, Chennai, Tamil Nadu, India

***Abstract*– Communication network is facilitated with different protocol. Each protocol supported to increase the network performance in a secured manner. In communication process, user's connectivity, violations of policy on access of information are handles through intrusion. Intrusion prevention is the process of performing intrusion detection and attempting to stop detected possible incidents. It focused on identifying possible incidents, logging information about them, attempting to stop them, and reporting them to security administrators. However, organizations use Intrusion detection and prevention system (IDPS) for other purposes, such as identifying problems with security policies, documenting existing threats, and determining individuals from violating security policies. Communication architecture is built up on IP. The Internet Control Message Protocol (ICMP) protocol is tightly integrated with IP. ICMP messages, delivered in IP packets, are used for out-of-band messages related to network operation or mis-operation. In this paper describes about HDLC, ICMP protocol sequences which is used to detect the intrusion on hybrid network and its attributes are recommend the standardized protocol for the intrusion detection process. This standardization protocol compare the previous part of the research of HDLC and ICMP protocols.**

***Index Term*– High Level Data Link Control (HDLC), Internet Control Message Protocol (ICMP) and Network Instruction Detection**

I. INTRODUCTION

THE modern information and communication Technology (ICT) system developed and facilitated many communication enhancement options for the up gradation of our living standards. The computer network is played vital role as a backbone of ICT. Many challenges are managed in this system. According to Peter, Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices.

II. INTRUSION DETECTION SYSTEM (IDS)

In recent years, dramatically increase the amount of data (text; images; audio; etc.) that available electronically on the Internet. Therefore, the intruders had made many successful attempts to bring down high-profile company networks and web services. Many methods have been developed to secure the data over the network and over the Internet. Computer networks are usually protected by anti-virus software, firewall, and encryption, secure network protocols, password protection etc. Since it has been proven that a potential attacker can always find a way to attack a network. These systems are known as Intrusion Detection System (IDS) and are placed inside the protected network, looking for potential threats in network traffic and or audit data recorded by host.

III. REVIEW OF LITERATURE

Initially intruder attempts to break into an information system or performs an action not legally allowed, we refer to this activity as an intrusion (Graham, 2002; see also Jones and Sielken, 2000). Intruders can be divided into two groups, external and internal. The former refers to those who do not have authorized access to the system and who attack by using various penetration techniques. The latter refers to those with access permission who wish to perform unauthorized activities. Intrusion techniques may include exploiting software bugs and system miss configurations, password cracking, sniffing unsecured traffic, or exploiting the design flaw of specific protocols (Graham, 2002). An Intrusion Detection System is a system for detecting intrusions and reporting them accurately to the proper authority. Intrusion Detection Systems are usually specific to the operating system that they operate in and are an important tool in the overall implementation an organization's information security policy (Jones and Sielken, 2000), which reflects an organization's statement by defining the rules and practices to provide security, handle intrusions, and recover from damage caused by security breaches.

There are two generally accepted categories of intrusion detection techniques: misuse detection and anomaly detection. Misuse detection refers to techniques that characterize known methods to penetrate a system. These penetrations are characterized as a 'pattern' or a 'signature' that the IDS looks for. The pattern/signature might be a static string or a set sequence of actions. System responses are based on identified penetrations. Anomaly detection refers to techniques that define and characterize normal or acceptable behaviors of the system (e.g., CPU usage, job execution time, system calls). Behaviors that deviate from the expected normal behavior are considered intrusions (Bezroukov, 2002; see also McHugh, 2001).

IDSs can also be divided into two groups depending on where they look for intrusive behavior: Network- based IDS (NIDS) and Host-based IDS. The former refers to systems that identify intrusions by monitoring traffic through network devices (e.g. Network Interface Card, NIC). A host-based IDS monitors file and process activities related to a software environment associated with a specific host. Some host-based IDSs also listen to network traffic to identify attacks against a host (Bezroukov, 2002; see also McHugh, 2001). There are other emerging techniques. One example is known as a blocking IDS, which combines a host-based IDS with the ability to modify firewall rules (Miller and Shaw, 1996). Another is called a Honeypot, which appears to be a 'target' to an intruder, but is specifically designed to trap an intruder in order to trace down the intruder's location and respond to attack (Bezroukov, 2002). This approached is planned with network based sensors. So the intrusion can be detected based on the observation of protocol and its sequence analysis. Therefore we are going to discuss about various protocol definitions which is used and observed in this research.

IV. PROTOCOLS

Protocols are set of rules that governing how data is transferred, compressed and presented over networks. There are many protocols, each one governing the way a certain technology works. A network protocol defines rules and conventions for communication between network devices. Protocols for computer networking all generally use packet switching techniques to send and receive messages in the form of packets. Network protocols include mechanisms for devices to identify and make connections with each other, as well as formatting rules that specify how data is packaged into messages sent and received. Some protocols also support message acknowledgement and data compression designed for reliable and/or high-performance network communication. Hundreds of different computer network protocols have been developed each designed for specific purposes and environments.

In general, The Internet Protocol family contains a set of related functionaries used network protocols. Besides Internet Protocol (IP) itself, higher-level protocols like TCP, UDP, HTTP, and FTP all integrate with IP to provide additional capabilities. Similarly, lower-level Internet Protocols like HDLC and ICMP also co-exist with IP. These higher level protocols interact more closely with applications like Web browsers while lower-level protocols interact with network adapters and other computer hardware. Here we are going to discuss few protocols which is observed over the network. The following part of the paper provides more details about various protocols and its functional services.

Internet Control Message Protocol (ICMP) is an integrated part of the IP suite. ICMP messages, delivered in IP packets, are used for out-of-band messages related to network operation or miss-operation. ICMP packet delivery is unreliable, so hosts can't count on receiving ICMP packets for any network problems. ICMP [1] relies on IP to perform its tasks, and it is an integral part of IP. It differs in purpose from transport protocols such as TCP and UDP in that it is typically not used to send and receive data between end systems. It is usually not used directly by user network applications, with some notable exceptions being the ping tool and trace route High-Level Data Link Control, also know as HDLC, is a bit oriented, switched and non-switched protocol. It is a data link control protocol, and falls within layer 2, the Data Link Layer, of the Open Systems Interface (OSI) model. It has also led to many subsets. Two subsets widely in use are Synchronous Data Link Control (SDLC) and Link Access Procedure-Balanced (LAP-B). Table 1 specifies the commands and Responses of the HDLC protocol.

Table 1: HDLC Commands and Responses

Information Transfer	Information Transfer
Format Commands	Format Responses
I - Information	I - Information
Supervisory Format	Supervisory Format
Commands	Responses
RR - Receive ready	RR - Receive ready
RNR - Receive not ready	RNR - Receive not ready
REJ - Reject	REJ - Reject
SREJ - Selective reject	SREJ - Selective reject
Unnumbered Format	Unnumbered Format
Commands	**Commands**
SNRM - Set Normal Response Mode	UA - Unnumbered Acknowledgment
SARM - Set Asynchronous Response Mode	DM - Disconnected Mode
SABM - Set Asynchronous Balanced Mode	RIM - Request Initialization Mode
DISC - Disconnect	RD - Request Disconnect
SNRME - Set Normal Response Mode Extended	UI - Unnumbered Information
SARME - Set Asynchronous Response Mode Extended	XID - Exchange Identification
SABME - Set Asynchronous Balanced Mode Extended	FRMR - Frame Reject
SIM - Set Initialization Mode	TEST - Test
UI - Unnumbered Information	
XID - Exchange identification	
RSET - Reset	
TEST - Test	

V. EXPERIMENTAL SPECIFICATION

The network consists of wired and wireless with internet, intranet, and extranet using LAN and WAN architectures to provide the services for the students, staff. This network used for file transfer (FTP), Remote access (TELNET), Active Directory Services (DNS), NETBIOS, Print server, IP telephony (Internal),Wireless Fidelity, Bluetooth, VPN, Email (IMAP), SMTP, E-Learning (Web server- HTTP), PING-ICMP, etc services. While providing the above specified services the network response and its Quality of Services varies due to the protocols which is used for the specific service. To reach the maximum service utilization, existing services are observed based on its protocol in and between the networks. There are many protocols running over the network to facilitate various requests and services. In this study we considered few services and its related protocol for the observation and analysis to construct the packet sequence to detect the intrusion.

The Network architecture of hybrid academic network which connect three academic department and four non academic departments. This network provides Teaching-learning and educational management service over 3000 students and the faculties in the campus. This network consists of LAN and the following technological configurations.

This academic network is framed as three clusters to provide the educational services. For the effective administration and maintenance of this network services, the classification and cluster made in the department level. In this study, the academic network structure and its laboratories' setup data communication and transformation architecture is adopted.

The network architecture constructed with modern technological equipments such as cisco switches (Core Switch) - 4503E, SAN-SWITCH-IBM-2005-16B, cisco-routers-1700, 2800 series; Firewall-CISCO-ASA-5510, cisco IP phones encompass of CISCO-MCS-7800-KQGCY35-Pentium-D-2.80GHz call manager. This also integrated with High end servers' such as HP Proliant-DL380-GB8639NHPS-Xeon 3.4Ghz; IBM-3850-99B5265-Xeon-3.5GHz; DVR-Proline-DVR-UK; SAN SWITCH- A device that routes data between servers and disk arrays in a storage area network. Its' 800 nodes are typically Conduit with UTP CAT-5, CAT-5E,CAT-6 and Fiber Channel switch made up of fiber multimode channels.

The established infrastructure integrated with wireless fidelity of various manufacturers. The network is enhanced with Video conferencing supported for inter and intra conferencing facility. There are many protocols are observed for the intrusion detection process to frame the sequence formation. But in this paper we are going to discuss the common sequence formation of the HDLC and ICMP protocol.

VI. STANDARDIZED 64-BYTEPROTOCOL STRUCTURE

The communication facilitation allows the intrusion attacker to the network. To Monitor and detect the same users, the following sequence are proposed.

1	2	3	4
Frame Info (0-31)			
Frame	Number	Length	Capture Length
Link	Data	Data	Data

From 1-4 bytes (32 bit) Frame Information.

The first byte represented about the frame information. This provides information about when the packets are travelled at that system or device, as well as number, length and capture of the packet.

5	6	7	8	9	10
Destination Address			(32 - 79)		
Broad Cast					
Group Address					
	Multi Cast	Local Address			

The next 48 bit (6 byte) provides the information about the destination. If any of the destinations is not listed with the specified network then that device will be blocked from the attached using GA algorithms.

11	12	13	14	15	16
Source (80 - 127)					
Uni Cast Individual					

The next 48 bit (6 byte) provides the information about the source. If any of the source not listed with the specified network then that device will be blocked from the attached using GA algorithms.

17	18	19	20	21	22	23	24	25	26
Type of Protocol (128-143)		ICMP (144 - 367)							
		Hardware Type		Protocol Type	Hardware Size			Protocol Size	OpCode

This ten byte information provides more details about the protocol type, hardware and related information's.

The following sequence will provide data about the MAC address of the sender as well as target device.

Table 2: 64 byte Protocol Structure

27-30	31-36	37-40	41-46
ARP (144 - 367)			
Mac Address	Sender IP	Target MAC	Target IP

VII. GENETIC ALGORITHM FOR INTRUSION DETECTION SYSTEM IDS

Genetic Algorithm (GA) has been used in different ways in IDSs. The Applied Research Laboratories of the University of Texas at Austin (Sinclair, Pierce, and Matzner 1999) uses different machine learning techniques, such as finite state machine, decision tree, and GA, to generate artificial intelligence rules for IDS. One network connection and its related behavior can be translated to represent a rule to judge whether or not a real-time connection is considered an intrusion. These rules can be modeled as chromosomes inside the population. The population evolves until the evaluation criteria are met. The generated rule set can be used as knowledge inside the IDS for judging whether the network connection and related behaviors are potential intrusions (Sinclair, Pierce, and Matzner 1999). The COAST Laboratory in Purdue University (Crosbie and Spafford, 1995) implemented an IDS using autonomous agents (security sensors) and applied AI techniques to evolve genetic algorithms. Agents are modeled as chromosomes and an internal evaluator is used inside every agent (Crosbie and Spafford, 1995).

A. Intrusion Detection using GA over the Communication Network

Sinclair introduced the concept of GA to detect IDS. As per his concept, Genetic algorithms can be used to evolve simple rules for network traffic (Sinclair, Pierce, and Matzner 1999).These rules are used to differentiate normal network connections from anomalous connections. These anomalous connections refer to events with probability of intrusions.

The rules stored in the rule base are usually in the following form (Sinclair, Pierce, and Matzner 1999):
if { condition } then { act }

For the problems we presented above, the condition usually refers to a match between current network connection and the rules in IDS, such as source and destination IP addresses and port numbers (used in TCP/IP network protocols), duration of the connection, protocol used, etc., indicating the probability of an intrusion. The act field usually refers to an action defined by the security policies within an organization, such as reporting an alert to the system administrator, stopping the connection, logging a message into system audit files, or all of the above. At the end GA is to generate rules that match only the anomalous connections. These rules are tested on historical connections and are used to filter new connections to find suspicious network traffic. To adopt and work for this IDS using GA this proposal is initiated.

B. Functional Process of Genetic Algorithm

The process usually begins with randomly generated population of chromosomes, which represent all possible solution of a problem that are considered candidate solutions. Different positions of each chromosome are encoded as bits, characters or numbers. These positions could be referred to as genes.

An evolution function is used to calculate the goodness of each chromosome according to the desired solution; this function is known as “fitness function”. During evaluation, two basic operators, cross over and mutation are used to simulate the natural reproduction and mutation of species. The selection of chromosomes for survival and combination is biased towards the fittest chromosomes.

- *Create a random initial state:* An initial state is created from a random selection of solutions. From the available packets, few packets are verified with the type and the registered devices across the network.
- *Evaluate Fitness:* The protocol type and the values which is registered in the list are verified. The list of values and the incoming protocol type value is differs the fitness of the protocol device is evaluated.
- *Crossover:* The combinational value of the list , type of service are offered, request IP and the responds are verified before providing service to the requested
- *Next generation:* If the value which is obtained and the value in the list are differ the packets are recheck the signal and verify the service. Otherwise the identified packet device is declared as a Intruder.

C. Decision Tree

Decision Tree induction is one of the classification algorithms in Data mining. The classification algorithm is inductively learned to construct a model from the reclassified data set (Briement et al., 1984). In our approach we have used the J48 algorithm for decision tree to audit data. The decision tree is partitioned into two classes of normal and Attack. Our object is to separate normal and attack pattern according to the attack class. This process is repeated for all the classes. The classifier was constructed using the training data and testing data.

The GA approach work with two modules with different stage. In the training state, GA used in an offline environment. In the ID stage the generated rules are used to classify incoming network connection in the real-time environment. The attributes are protocol, source and destination port IP and attack-name. It generates initial population, sets the default parameter and audit the data. The initial population is evolved by the number of generation. Quality of the rules is initially calculated, then a number of best-fit rules are selected and finally GA operators are applied to all generations.

D. J48 Algorithm

J48 implements Quinlans c4.5 algorithm for generating a pruned or unpruned C4.5 decision tree. The decision tree generated by J48 can be used for classification. J48 builds decision tree from a set of leveled training data using the concept of information entropy. It uses the fact that each attribute of the data can be used to make a decision by splitting the data into smaller subsets. J48 examines the normalized information gain that results from choosing an attribute for splitting the data. To make the decision, the attribute with the highest normalized information gain is used. Then the algorithm recurs on the smaller subsets.

The splitting procedure stops if all instances in a subset belong to the same class. Then a leaf node is created in the decision tree telling to choose that class. In this case J48 creates a decision node higher up in the tree using the expected value of the class .It handle both continuous and discrete attributes, training data with missing attribute value and attribute with differing cost. Further, it provides an option for pruning trees after creation.

VIII. PERFORMANCE EVOLUTION

We used real-time collected protocol converted file for detection data set .This set defined the features between normal connections from attacks. We used three type of protocol structure in the hybrid academic network. The result summary displayed in Table 3.

In the observation of the analysis, the intrusion is detected in all the protocol. But the identification is not met all the levels. In the HDLC protocol identified and detected 2.25 % of the total packets. At the same time it performed 21.65 % of total intuition and the 5.5% of the HDLC protocol packets. Similarly the number of packets detected in the ICMP is 233 in an average. This 0.44 % of the total packets is 4.35% of the intuition and 4.64% of the captured ICMP packets. As per the result the HDLC is performed well. The data are presented in the Table 3 and Table 4.

IX. CONCLUSION

This paper proposed standardized 64-byte structure is easy to capture the intuition from the network. All the required information from the source and the sender as well as sender and target device are captured in this structure. This is not affected the data transformation process but this can be integrated to the monitor the network. In this process, the detection is full. The maximum percentage is 30% while we are using the HDLC protocol. This observation identified that all the inductions are not occurs during these protocol .If this protocol IDS is extended to other protocol all the intuition are possible to identified and prevent from the attackers. The Genetic algorithm and the j48 tree DSS process supported to increase the speed of detection. This process planned to integrate with the real time network dynamic data processing and online process.

Table 3: Result summary of ICMP Protocol

S. No	% of ICMP Ide vs Total	% of ICMP Ide vs Inti	% of Ide ICMP vs ICMP
1	1.01	7.95	8.73
2	0.22	2.75	2.80
3	0.34	4.33	3.50
4	0.26	2.44	4.01
5	0.33	1.95	6.36
6	0.62	4.24	6.70
7	0.48	6.38	5.00
8	0.32	4.94	2.64
9	0.48	4.40	3.89
1	0.34	4.11	2.77
Min	0.22	1.95	2.64
Max	1.01	7.95	8.73
A	0.44	4.35	4.64

Table 4: Result summary of HDLC Protocol

S. No	#packets	% of Inti Pac	% of HDLC Ide vs Total	% of HDLC Ide vs Inti	% of Ide HDL
1	52489	12.67	2.91	22.93	5.44
2	52489	7.82	2.42	30.95	4.54
3	49676	7.81	1.11	14.15	2.34
4	49676	10.57	1.79	16.96	3.80
5	51305	16.77	4.37	26.07	8.89
6	51305	14.58	3.44	23.62	6.99
7	49781	7.59	1.78	23.39	3.34
8	57756	6.45	1.43	22.10	3.03
9	63383	10.84	1.34	12.35	3.47
10	54761	8.18	1.96	23.94	3.84
Min		6.45	1.11	12.35	2.34
Max		16.77	4.37	30.95	8.89
Avr		10.33	2.25	21.65	4.57

REFERENCES

[1] A. Abraham and C. Grosan “Evolving Intrusion Detection Systems”, Studies in computational intelligence (SCI), 2006.

[2] A. Abraham, Sandhya peddabachigai, “Modeling intrusion detection system using hybrid intelligent systems”, journal network and computer applications, 2007.

[3] Bridges, Susan, and Rayford B. Vaughn. 2000. “Intrusion Detection via Fuzzy Data Mining.” In Proceedings of 12th Annual Canadian Information Technology Security Symposium. 109-122. Ottawa, Canada.

[4] Crosbie, Mark, and Gene Spafford. 1995. “Applying Genetic Programming to Intrusion Detection.” In Proceedings of 1995 AAAI Fall Symposium on Genetic Programming, pp. 1-8 Cambridge, Massachusetts.

[5] David C. Plummer (1982-11). "RFC 826, An Ethernet Address Resolution Protocol -- or -- Converting Network Protocol Addresses to 48.bit Ethernet Address for

Transmission on Ethernet Hardware". Internet Engineering Task Force, Network Working Group. http://tools.ietf.org/html/rfc826.

[6] http://csrc.ncsl.nist.gov/publications/nistpubs/800-94/SP800-94.pdf Guide to Intrusion Detection and Prevention Systems (IDPS), NIST CSRC special publication SP 800-94, released 02/2007.

[7] Jones, Anita. K. and Robert. S. Sielken. 2000. "Computer System Intrusion Detection: A Survey." Technical Report.Department of Computer Science, University of Virginia, Charlottesville, Virginia.

[8] Li, Wei. 2002. "The integration of security sensors into the Intelligent Intrusion Detection System (IIDS) in a cluster environment." Master's Project Report. Department of Computer Science, Mississippi State University.

[9] McHugh, John, 2001. "Intrusion and Intrusion Detection." Technical Report. CERT Coordination Center, Software Engineering Institute, Carnegie Mellon University.

[10] Paxson, Vern. 1998. "Bro: A System for Detecting Network Intruders in Real-time." In Proceedings of 7th USENIX Security Symposium, pp. 31-51. San Antonio, Texas.

[11] Pohlheim, Hartmut. 30 Oct. 2003. "Genetic and Evolutionary Algorithms: Principles, Methods and Algorithms."Genetic and Evolutionary Algorithm Toolbox. Hartmut Pohlheim.

[12] R. Bace and P. Mell, "Intrusion Detection Systems", National Institute of standards and Technology NIST", 2001.

[13] Robert Graham. URL: http://www.robertgraham.com/pubs/network- intrusion-detection.html (30 Oct. 2003).

[14] Roesch, Martin. Nov. 7-12, 1999. "Snort - Lightweight Intrusion Detection for Networks." In Proceedings of 13th Systems Administration Conf. (LISA '99), pp. 229-238.Seattle, Washington.

[15] S. Kumar and E.H. Spaffort," a Software Architecture to support misuse Intrusion Detection", Technical Report, 1995.

[16] Sinclair, chris, Lynpierce and Sara Matznel, 1999."An application of Machine learning to network intrusion detection", In proceedings of 1999 Annual Computer Society applications Conference. pp. 371-377.

[17] Sinclair, chris, Lynpierce and Sara Matznel, 1999."An application of Machine learning to network intrusion detection", In proceedings of 1999 Annual Computer Society Applications Conference. pp. 371-377.

[18] Softpanorama: Open Source Software Educational Society. Nikolai Bezroukov. URL: http://www.softpanorama.org/Security/intrusion_detection.shtml (30 Oct. 2003).

[19] W. Li. "Using Genetic Algorithm for network Intrusion Detection", Proceedings of the United States Department of Energy Cyber Security Group", 2004.