ISSN No. 0976-5697

Volume 1, No. 4, Nov-Dec 2010

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Novel Approach for Identifying Intruder based on Multiple Approaches

Amit Kumar Sharma*
Research Scholar in (M.Tech)
Gyan Vihar University, Jaipur
amitmatoliya@gmail.com

Mr. Naveen Hemrajani
Vice Principal of Gyan Vihar University,
Jaipur,India
naven_h@yahoo.com

Abstract**:** The paper reviews of research to identify the intrusion and intruder, most computers authenticate user ID and password before users can login these systems. However, danger soon comes if the two items are known to hackers. In this paper, we propose a system that finds who intrude and how they intrude in to the system. The information about the intruder and their saved information in the data warehouse to the other location without knowing to the intruder and in this duration server will find the information about the intruder. In future when intruder again enter in the network then caught and show the information about him and last time you enter in this network. For this technique I can reduce the network attack. And find more type of Network intrusion Detecting System with using data mining Approach. Our experimental results show that the recognition accuracy of students of computer science department is up to 99%.

Keywords**:** Network Intrusion Detection, Data Mining, Honeypots, User Authentication.

I. INTRODUCTION

As the cost of the information processing and Internet accessibility falls, more and more organizations are be-coming vulnerable to a wide variety of cyber threats. According to a recent survey [1] by CERT/CC (Computer Emergency Response Team/Coordination Center), the rate of cyber attacks has been more than doubling every year in recent times. It has become increasingly important to make our information systems, especially those used for critical functions in the military and commercial sectors, resistant to and tolerant of such attacks [2].

The first threat for a computer network system was realized in 1988 when 23-year old Robert Morris launched the first worm, which override over 6000 PCs of the ARPANET network. On February 7th, 2000 the first DoS (Denial of Services) attacks of great volume where launched, targeting the computer systems of large companies like Yahoo!, eBay, Amazon, CNN, ZDnet and Dadet. More details on these attacks can be found at [3]. These threats and others that are likely to appear in the future have lead to the design and development of Intrusion Detection Systems. According to webopedia [4] an intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system.

One of the main challenges in the security management of large-scale high-speed networks is the detection of suspicious anomalies in network traffic patterns due to Distributed Denial of Service (DDoS) attacks or worm propagation [5][6] A secure network must provide the following:

[a] Data confidentiality**:** Data that are being transferred through the network should be accessible only to those that have been properly authorized.

[b] Data integrity**:** Data should maintain their integrity from the moment they are transmitted to the moment they are actually received. No corruption or data loss is accepted either from random events or malicious activity.

[c] Data availability**:** The network should be resilient to Denial of Service attacks.

Intrusion detection techniques using data mining have attracted more interests in recent years. As an important application area of data mining, they aim to meliorate the great burden of analyzing huge volumes of audit data and realizing performance optimization of detection rules. Different researchers propose different algorithms in different categories, from Bayesian approaches [7] to decision trees [8, 9], from rule based models [10] to functions studying [11]. The detection efficiencies therefore are becoming better and better than ever before.

II. EXISTING SYSTEMS (RELATED WORK)

In this section, we present some of the implemented systems that apply data mining techniques in the field of Intrusion Detection.

[a] ISOA (Information Security Officer's Assistant) [12]: ISOA is a system for monitoring security relevant behavior in computer networks. ISOA serves as the central point for real-time collection and analysis of audit information. When an anomalous situation is identified, associated indicators are triggered. ISOA automates analysis of audit trails, allowing indications and warnings of security threats to be generated in a timely manner so that threats can be countered. ISOA allows a single designated workstation to perform automated security monitoring, analysis and warning

[b] Distributed Intrusion Detection System (DIDS) [13]: A risk intrusion detection system that aggregates audit reports from a collection of hosts on a single network. Unique to DIDS is its ability to track a user as he establishes connections across the network.

[c] EMERALD (SRI) [14]: EMERALD is a software-based solution that utilizes lightweight sensors distributed over a network or series of networks for real-time detection of anomalous or suspicious activity. EMERALD sensors monitor activity both on host servers and network traffic streams, and empower system defenders with the capacity to detect and ultimately thwart cyber attacks across large networks By using highly distributed surveillance and response monitors, EMERALD provides a wide range of information security coverage, real-time monitoring and response, protection of informational assets. EMERALD implements an enterprise-wide analysis to correlate the activity reports produced across asset of monitored domains. EMERALD offers protection from network-wide threats such as Internet worm-like attacks, attacks repeated against common network services across domains, or coordinated attacks from multiple domains against a single domain.

[d] The MINDS System [15]: The Minnesota Intrusion Detection System (MINDS) uses data mining techniques to automatically detect attacks against computer networks and systems. While the long-term objective of MINDS is to address all aspects of intrusion detection, the system currently focuses on two specific issues: – An unsupervised anomaly detection technique that assigns a score to each network connection that reflects how anomalous the connection is, and – An association pattern analysis that summarizes those network connections that are ranked highly anomalous by the anomaly detection module. Experimental results on live network traffic at the University of Minnesota show that the applied anomaly detection techniques are very promising and are successful in automatically detecting several novel intrusions that could not be identified using popular signature-based tools such as SNORT. Furthermore, given the very high volume of connections observed per unit time, association pattern based summarization of novel attacks is quite useful in enabling a security analyst to understand and characterize emerging threats.

[e] The IDDM project [16]: The IDDM (Intrusion Detection using Data Mining) project is a project that uses data mining techniques in order to describe the data on a network and analyze them for further deviation in observed traffic. IDDM utilizes meta-mining to achieve its goals. The goal is to track and understand changes in the network traffic over time. IDDM uses association rules in order to observe network traffic. Two different snapshots of the association rules -created in two different timestamps- are compared in order to see which rules have remained the same, have been changed, been added and which have been eliminated. The system uses agents that apply association rule mining on raw network packets.

[f] The IDIS (Intrusion Detection and Identification System) [17] framework consists of intelligent monitor, detection server and mining server. Intelligent monitor collects input commands from underlying user and transfers the command sequences to detection server which compares these commands with attack patterns real-time to discover attacks. In IDIS, attack patterns are represented by a reverse tree, a tree of which commands are organized in the reverse order of their arrival from the root. If matched, detection server notifies intelligent monitor to disconnect the session established for the user. Mining server analyzes log data with data mining techniques to identify user habits. The IDIS can discriminate who an underlying user is in a concerned intranet by comparing the user's current inputs with all others' habits.

A. Problem in Existing System

In IDIS When patterns is matched with the detection server, then it will disconnect the session established for the user, some time we need the information about the hacker, But we cannot find more about the Hacker.

III. SYSTEM FRAMWORK (PROPOSED WORK)

In this paper we will tell the new concept to remove network intrusion or indentify the network intruder and the technique used to enter in the network. In this system the intelligent monitor [18] collects input commands from underlying user and transfers the command sequences to detection server which compares these commands with attack patterns real-time to discover attacks. It will then find it previous data and then ask some question from his previous record. If the answer is right then user will be able to get authenticated and enter into the server. Otherwise the intruder will be sent to the other (honey pots) dummy server in between server will find the information about the hacker, Intruder, and store the information in the database. According to the information we easily find the location and caught the hacker.

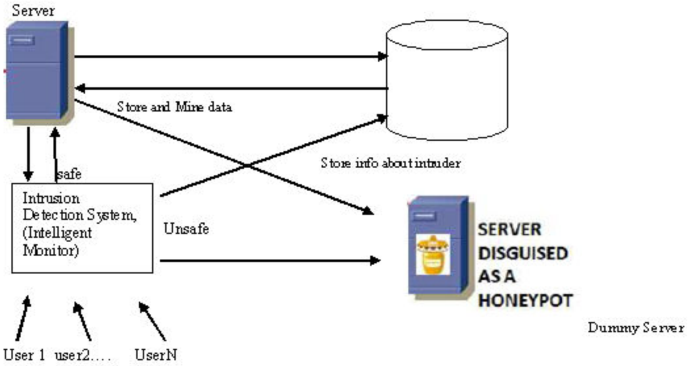
Server
Store and Mine data
Store info about intruder
safe
Intrusion Detection System, (Intelligent Monitor)
Unsafe
SERVER DISGUISED AS A HONEYPOT
Dummy Server
User 1
user2....
UserN

Figure 1

A. Intelligent Monitor

As an extended portion of an operating system, intelligent monitor comprises input interceptor and system call filter. When a user submits a command, at least one system call will be generated. A system call generated by shell in executing a command is compared with sensitive call table, a table holding all sensitive calls. Once matched, the system call will be transferred to system call filter to check to see whether the call is safe or not. Unsafe system calls will be retained for a further analysis. A safe one will be sent to system kernel to perform its corresponding service. Besides, we divide users into groups according to their occupations. Each group G has its corresponding inhibited commands, named class-limited command list (G) [19]. Intelligent monitor denies a user's input command immediately if the command is in the user's class-limited command list. After the IDIS starts up, input interceptor sends a command, request cmd list to request monitored command list which consists of last commands of all attack patterns. As an input command is in the monitored command list, this command will be held by input interceptor until detection server replies with "safe" or "unsafe"

B. Mining Server

Mining server extracts commands that a user has habitually used from his/her log file, counts the frequency each command appears in the log file, and stores the result in the user's habit file. After that, users' habit files are mutually compared with each other to identify common and user specific command sequences, with which user profiles [20,21] can be then created.

C. What is Honeypots?

Honeypots are a new technology with enormous potential for the Information Technology community. The first concepts regarding them were introduced by various icons in Information Security, such as those defined by Cliff Stoll in the book "The Cuckoo's Egg" (2002) and the works of Bill Cheswick, documented in the book "An Evening with Berferd" (1997). Since then, those concepts have been in continuous evolution, developing in an accelerated way and becoming a powerful security tool now a day (Riebach, Rathgeb & Tödtmann, 2005).

Honeypots are, in their most basic form, fake information severs strategically-positioned in a test network, which are fed with false information disguised as files of classified nature. In turn, these servers are initially configured in a way that is difficult, but not impossible, to break into them by an attacker; exposing them deliberately and making them highly attractive for a hacker in search of a target (Spitzner, 2002). Finally, the server is loaded with monitoring and tracking tools so every step and trace of activity left by a hacker can be recorded in a log, indicating those traces of activity in a detailed way.

The main functions of a Honeypot are (Pouget & Holz, 2005):

- To divert the attention of the attacker from the real network, in a way that the main information resources are not compromised

- To capture new viruses or worms for future study

- To build attacker profiles in order to identify their preferred attack methods, similar to criminal profiles used by law enforcement agencies in order to identify a criminal's modus operandi

- To identify new vulnerabilities and risks of various operating systems, environments and programs which are not thoroughly identified at the moment?

In a more advanced context, a group of Honeypots becomes a Honeypots, thus providing a tool that spans a wide group of possible threats which gives a systems administrator more information for study. Moreover, it makes the attack more fascinating for the attacker due to the fact that Honeypots can increase the possibilities, targets and methods of attack.

D. Mining User Stored Previous Data and Attack Patterns

A log file consists of many sessions. Each comprises commands a user submitted within the period of time between its login and the corresponding logout. Given a user's log file, the IDIS processes the commands with a sliding window of size 10, named Log-sliding window (L-window in short), to partition the commands along their submitted sequence into k-grams where k is the number of consecutive commands, k =2, 3, 4....10. Besides, another sliding window of 10 commands, named Compared-sliding window (C-window in short), is also deployed on another concerned session. This time, k' consecutive commands, preserving their submitted sequence, are extracted from C-window generating a total of (10 – k' + 1) k'-grams, k'=2, 3, 4...10. Mining server invokes algorithm 1 to compare each of.

$$\sum_{K=2}^{10} (10-k+1) \quad \ldots\ldots \quad \sum_{K=2}^{10} (10-k'+1)\, k'$$

By using the longest common sequence algorithm. After that, C-window shifts one command right. The procedure repeats until the last session of the log file is involved. Then Log-window shifts one input command right. The whole procedure repeats until last ten or all (if less than ten) commands of the last second session are encountered by the L-window.

IV. CONCLUSION

In this article, we bring up an approach to find out users' habits by deploying data mining and forensic techniques. To identify the representative C-sequences for a user, the frequency that a habitual command sequence appears in the user's log file is counted and its discrimination score is calculated so that the user's profile can be established. By comparing a user's current input commands with all others' profiles, the IDIS can identify who the user is. The accuracy is high enough to make the IDIS be a valuable auxiliary subsystem in a closed environment to assist the identification of an internal hacker. Of course, a new user whose user profile has not been established will not be a candidate to be identified. Meanwhile, a user's input commands are compared with the common reverse tree in which all commands of an attack pattern are organized in their reverse order so as to real time detect whether underlying inputs are an attack or not. Employing the common reverse tree can lightweight IDIS and lower the load of detection server. Moreover, accurately and completely collecting user behaviors on much more basic operations, such as system calls instead of commands, is much more helpful in detecting hackers and identifying a user. Such will also help us to collect intrusion behaviors in a system that employs GUI interface. However, how to process and mine such a huge volume of data may be the first challenge. Several papers have addressed this topic [15, 24]. But many systems have not been implemented, and many did not describe their implementation. Additionally, to detect an attack and respond real time, we need a fast algorithm and a distributed computing environment to speedup data processing since the time complexity of algorithm 1 is high. Cluster and/or Grid computing should be the candidates. Besides, a mathematical analysis on the IDIE's behaviors so as to build its formal performance and cost models is interesting. Those are our future research topics.

V. REFERENCES

[1] Successful Real-Time Security Monitoring, Riptech Inc. white paper, September 2001.

[2] Data Mining for Network Intrusion Detection Paul Dokas, Levent Ertoz, Vipin Kumar, Aleksandar Lazarevic, Jaideep Srivastava, PangNing Tan Computer Science Department, 200 Union Street SE, 4-192, EE/CSC Building University of Minnesota, Minneapolis, MN 55455, USA{dokas, ertoz, kumar, aleks, srivasta, ptan}@cs.umn.edu

[3] EBay, Amazon, CNN, ZDnet and Dadet http://www.securityfocus.com/news/2445

[4] Intrusion Detection Systems. According to webopedia http://www.webopedia.com

[5] Christos Douligeris, Aikaterini Mitrokotsa, "DDoS attacks and defense mechanisms: classification and state-of-the-art" ,Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 44, Issue 5 , pp: 643 - 666, 2004.

[6] Z. Chen, L. Gao, K. Kwiat, Modeling the spread of active worms, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Vol. 3, pp. 1890 1900, 2003.

[7] B. Schneier, "CardSystems Exposes 40 Million Identities, "http://www.schneier.com/blog/archives/2005/06/cardsystems_exp.html

[8] A.P. Mitchell, "40 million," http://www.theinternetpatrol.com/cardsystems-compromisesdata-of-40-million-mastercard-and-visa-cardholders

[9] Y. Sheng, V. V. Phoha, and S. M. Rovnyak, "A Parallel Decision Tree-Based Method for User Authentication Based on Keystroke Patterns," IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, vol. 35, no. 4, August 2005, pp.826-833.

[10] M. Ray, P. Meenen, and R. Adhami, "A Novel Approach to Fingerprint Pore Extraction," IEEE Southeastern Symposium on System Theory, March 2005, pp.205-208.

[11] John, G.H., Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers. In: Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence (1995)

[12] Winkler, J. R., Landry, L. C., "Intrusion and anomaly detection, ISOA update", In Proceedings of the 15th National Computer Security Conference, pages 272-281, Oct. 1992.

[13] Winkler, J. R., Landry, L. C., "Intrusion and anomaly detection, ISOA update", In Proceedings of the 15th National Computer Security Conference, pages 272-281, Oct. 1992.

[14] Snapp, S. R., Smaha, S. E., Grance, T., Teal, D. M., "The DIDS (Distributed Intrusion Detection System) Prototype", In Proceedings of the USENIX Summer 1992 Technical Conference, pages 227-233, June 1992.

[15] Porras, A. and Neumann, P. G.,"EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances", In Proceedings of the National Information Systems Security Conference, October 1997.

[16] Levent Ertoz and Eric Eilertson and Aleksandar Lazarevic and Pang-Ning Tan and Vipin Kumar and Jaideep Srivastava and Paul Dokas, "MINDS - Minnesota Intrusion Detection System", Next Generation Data Mining, MIT Press, 2004.

[17] A Real-Time Intrusion Detection System using Data Mining Technique Fang-Yie Leu* Department of Computer Science and Information Engineering, Tunghai University, Taiwan leufy@thu.edu.tw And Kai-Wei Hu Department of Computer Science and Information Engineering, Tunghai University, Taiwan

[18] The intelligent monitor http://en.wikipedia.org/wiki/Snort_%28software%29

[19] F. Dridi, B. Muschall, and G. Pernul, "Administration of an RBAC system," International Conference on System Sciences, 2004, pp.1-6. http://csdl2.computer.org/comp/proceedings/hicss/2004/20 56/07/205670187b.pdf

[20] Y. Okazaki, I. Sato and S. Goto, "A new Intrusion Detection Method Based on Process Profiling,". Sympo sium on Applications and the Internet, Feb. 2002, pp. 82 - 90.

[21] J.E. Dickerson and J.A. Dickerson, "Fuzzy Network Profiling for Intrusion Detection," International Conferen ce of the North American on Fuzzy Information Process ing Society, July 2000, pp. 301 -306.

[22] A. Leuski, "Evaluating Document Clustering of Interactive Information Retrieval," ACM CIKM, Nov. 2001, pp. 33-40.

[23] V.S. Verykios et al., "State-of-the-Art in Privacy Preserving Data Mining," SIGMOD Record, vol.33, no.1, March 2004, pp.50-57.

[24] K. Lu, Z. Chen, Z. Jin, and Jichang Guo, "An Adaptive Real-Time Intrusion Detection System Using Sequence of System Call, " IEEE Electrical and Computer Engineering, May/mai 2003, pp.789-792.

ABOUT AUTHOR

Amit Kumar Sharma received his Msc degree in Computer Science from M.D.S University, Ajmer Rajasthan, India, year 2008 and is presently pursuing his M.Tech in Software Engineering at the School of Engineering, Suresh Gyan Vihar University, Jaipur, India. His current research interests include information security, intrusion detection identify the intruder. He is being guided by Mr. Naveen Hemrajani towards the completion of his thesis. Mr. Naveen Hemrajani, is Vice Principal and Research Scholar of Gyan Vihar University, Jaipur.