International Journal of Research
Available at
https://edupediapublications.org/journals

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 05
April 2017

A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection

Lalu Banoth [1] . M P Sri Kamal Teja[2]. M Saicharan[3]. N Jaya Chandra[4]
Associate Professor, Department of CSE, Guru Nanak Institutions, Ibrahimpatnam, Hyderabad, India[1]
B.Tech Student, Department of Computer Science & Engineering, Guru Nanak Institutions, Hyderabad, India[2]
B.Tech Student, Department of Computer Science & Engineering, Guru Nanak Institutions, Hyderabad, India[3]
B.Tech Student, Department of Computer Science & Engineering, Guru Nanak Institutions, Hyderabad, India[4]

ABSTRACT

Cyber security is that the body of technologies, processes and practices designed to safeguard networks, computers, programs and knowledge from attack, harm or unauthorized access. During a computing context, the term security implies cyber security. This survey paper describes a targeted literature survey of machine learning (ML) and data processing (DM) strategies for cyber analytics in support of intrusion detection. This paper focuses totally on cyber intrusion detection as it applies to wired networks. With a wired network, associate oppose must experience many layers of defense at firewalls and operative systems, or gain physical access to the network. The quality of ML/DM algorithms is addressed, discussion of challenges for victimization ML/DM for cyber security is conferred, and some recommendations on once to use a given methodology area unit provided.

INTRODUCTION

distributed denial-of-service (DDoS), eavesdropping attack, and spear-phishing attack insider attack is one of the most difficult ones to be detected because firewalls and intrusion detection systems (IDSs) usually defend against outside attacks. To authenticate users, currently, most systems check user ID and password as a login pattern. However, attackers may install Trojans to pilfer victims' login patterns or issue a large scale of trials with the assistance of a dictionary to acquire users' passwords. When successful, they may then log in to the system, access users' private files, or modify or destroy system settings. Fortunately, most current host-based security systems and network-based

IDSs can discover a known intrusion in a real-time manner.

EXISTING SYSTEM

When people exploit powerful capabilities and processing power of computer systems, security has been one of the serious problems in the computer domain since attackers very usually try to penetrate computer systems and behave maliciously, e.g., stealing critical data of a company, making the systems out of work or even destroying the systems. Attackers may install Trojans to pilfer victims' login patterns or issue an large scale of trials with the assistance of a dictionary to acquire users' passwords

EXISTINGSYSTEM DISADVANTAGES

- When successful, they may then log in to the system, access users' private files, or modify or destroy system settings.
- Accuracy of detection is low

PROPOSED SYSTEM

In this paper, we propose a security system, named Internal Intrusion Detection and Protection System (IIDPS), which detects malicious behaviors launched toward a system at SC level. The IIDPS uses data mining and forensic profiling techniques to mine system call patterns (SC-patterns) defined as the longest system call sequence (SC-sequence) that has repeatedly appeared several times in a user's log file for the user. The user's forensic features, defined as an SC-pattern frequently appearing in a user's submitted SC-sequences but rarely being used by other users, are retrieved from the user's computer usage history.

PROPOSED SYSTEM ADVANTAGES

- Techniques used for intrusion detection provide effective attack resistance.
- Average detection accuracy is higher

MODULES:

- User Interface Design
- Control Center Initialization Model
- Mining User and Attacker Habits
- Implementing Attack Analyzer
- Attack Graph Model

Modules Description:

User Interface Design

This is the first module of our project. The important role for the cloud user is to move login window to cloud user window. This module has created for the security purpose. In this login page we have to enter login user id and password. It will check username and password is match or not (valid user id and valid password). If we enter any invalid username or password we can't enter into login window to user window it will shows error message. So we are preventing from unauthorized user entering into the login window to user window. It will provide a good security for our project. So server contain user id and password server also check the authentication of the user. It well improves the security and preventing from unauthorized user enters into the network. In our project we are using JSP for creating design. Here we validate the login user and server authentication

CONTROL CENTER INITIALIZATION MODEL

This is the second and important module in our application Manager Interface Design plays an important role for the Manager to move login window to Manager welcome window. Manager will enter the salary details of users and allocating the project to the Team leader.

MINING USER AND ATTACKER HABITS:

An insider attacker may log in to a system by using another user's login ID and password and do something maliciously. However, attackers may install Trojans to pilfer victims' login patterns or issue a large scale of trials with the assistance of a dictionary to acquire users' passwords. When successful, they may then log in to the system, access users' private files, or modify or destroy system settings. Fortunately, most current host-based security systems and network-based IDSs can discover a known intrusion in a real-time manner. However, it is very difficult to identify who the attacker is because attack packets are often issued with forged IPs or attackers may enter a system with valid login patterns. ,.

IMPLEMENTING ATTACK ANALYZER

Internal Intrusion Detection and Protection System (IIDPS), which detects malicious behaviors launched toward a system at SC level. The IIDPS uses data mining and forensic profiling techniques to mine system call patterns (SC-patterns) defined as the longest system call sequence (SC-sequence) that has repeatedly appeared several times in a user's log file for the user. The user's forensic features, defined as an

SC-pattern frequently appearing in a user's submitted SC-sequences but rarely being used by other users, are retrieved from the user's computer usage history.

ATTACK GRAPH MODEL

This is the fifth module of our project in this with the advent of web applications, An attack graph is a modeling tool to illustrate all possible multi-stage, multi-host attack paths that are crucial to understand threats and then to decide appropriate countermeasures . In an attack graph, each node represents either precondition or consequence of an exploit. The actions are not necessarily an active attack since normal protocol interactions can also be used for attacks. Attack graph is helpful in identifying potential threats, possible attacks and known vulnerabilities in a cloud system. Since the attack graph provides details of all known vulnerabilities in the system and the connectivity information, we get a whole picture of current security situation of the system where we can predict the possible threats and attacks by correlating detected events or activities.

DESIGN ENGINEERING

Design Engineering deals with the various UML [Unified Modeling language] diagrams for the implementation of project. Design is a meaningful engineering representation of a thing that is to be built. Software design is a process through which the requirements are translated into representation of the software. Design is the place where quality is rendered in software engineering. Design is the means to accurately translate customer requirements into finished product.

DEVELOPMENT TOOLS

This chapter is about the software language and the tools used in the development of the project. The platform used here is JAVA.

FEATURES OF JAVA

Java is a programming language originally developed by James Gosling at Sun Microsystemsand released in 1995 as a core component of Sun Microsystems' Java platform. The language derives much of its syntax from C and C++ but has a simpler object model and fewer low-level facilities. Java applications are typically compiled to bytecode that can run on any Java Virtual Machine (JVM) regardless of computer architecture. Java is general-purpose, concurrent, class-based, and object-oriented, and is specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere".

Java is considered by many as one of the most influential programming languages of the 20th century, and is widely used from application software to web applicationsThe java framework is a new platform independent that simplifies application development internet.Java technology's versatility, efficiency, platform portability, and security make it the ideal technology for network computing. From laptops to datacenters, game consoles to scientific supercomputers, cell phones to the Internet, Java is everywhere!

Servlets

Earlier in client- server computing, each application had its own client program and it worked as a user interface and need to be installed on each user's personal computer. Most web applications use HTML/XHTML that are mostly supported by all the browsers and web pages are displayed to the client as static documents.

A web page can merely displays static content and it also lets the user navigate through the content, but a web application provides a more interactive experience.

Any computer running Servlets or JSP needs to have a container. A container is nothing but a piece of software responsible for loading, executing and unloading the Servlets and JSP. While servlets can be used to extend the functionality of any Java-enabled server.

They are mostly used to extend web servers, and are efficient replacement for CGI scripts. CGI was one of the earliest and most prominent server side dynamic content solutions, so before going forward it is very important to know the difference between CGI and the Servlets.

APPLICATION

- Health Record Management
- Credit card application

FUTURE ENHANCEMENT

For future work, further study will be done by improving IIDPS’s performance and investigatingThird-party shell command.This paper presents intelligent lightweight IDS, which used the forensics technique to profile the user behavior in order to automate the maintenance of user profile, data mining technique to find out the cooperative attack, and watermark technique to trace back the hackers or intruders. The goal of the system is to detect the intrusion real-time, effectively and efficiently.

CONCLUSION:

In this paper, we have proposed an approach that employs data mining and forensic techniques to identify the representative SC-patterns for a user. The time that a habitual SC patternappears in the user's log file is counted, the most commonly used SC-patterns are filtered out, and then a user's profile is established. By identifying a user's SC-patterns as his/her computer usage habits from the user's current input SCs, the IIDPS resists suspected attackers.

Reference:

1. S. Gajek, A. Sadeghi, C. Stuble, and M. Winandy, "Compartmented security for browsers—Or how to thwart a phisher with trusted computing,"in *Proc. IEEE Int. Conf. Avail., Rel. Security*, Vienna, Austria, Apr. 2007,pp. 120–127.

2. C. Yue and H. Wang, "BogusBiter: A transparent protection againstphishing attacks," *ACM Trans. Int. Technol.*, vol. 10, no. 2, pp. 1–31,May 2010.

3. Q. Chen, S. Abdelwahed, and A. Erradi, "A model-based approach toself-protection in computing system," in *Proc. ACM Cloud AutonomicComput. Conf.*, Miami, FL, USA, 2013, pp. 1–10.

4. F. Y. Leu, M. C. Li, J. C. Lin, and C. T. Yang, "Detection workload in adynamic grid-based intrusion detection environment," *J. Parallel Distrib.Comput.*, vol. 68, no. 4, pp. 427–442, Apr. 2008.

5. H. Lu, B. Zhao, X. Wang, and J. Su, "DiffSig: Resource differentiationbased malware behavioral concise signature generation," *Inf. Commun.Technol.*, vol.7804, pp.271–284, 2013.

6. Z. Shan, X. Wang, T. Chiueh, and X. Meng, "Safe side effects commitmentfor OS-level virtualization," in *Proc. ACM Int. Conf. AutonomicComput.*, Karlsruhe, Germany, 2011, pp. 111–120.

7. M. K. Rogers and K. Seigfried, "The future of computer forensics:A needs analysis survey," *Comput. Security*, vol. 23, no. 1, pp.12–16,Feb. 2004.

8. J. Choi, C. Choi, B. Ko, D. Choi, and P. Kim, "Detecting web based DDoSattack using MapReduce operations in cloud computing environment,"*J. Internet Serv. Inf. Security*, vol. 3, no. 3/4, pp. 28–37, Nov. 2013.

9. Q. Wang, L. Vu, K. Nahrstedt, and H. Khurana, "MIS: Malicious nodesidentification scheme in network-

coding-based peer-to-peer streaming,”in*Proc. IEEE INFOCOM*, San Diego, CA, USA, 2010, pp. 1–5.

10 . Z. A. Baig, “Pattern recognition for detecting distributed node exhaustionattacks in wireless sensor networks,” *Comput. Commun.*, vol. 34, no. 3,pp. 468–484, Mar. 2011.

11. H. S. Kang and S. R. Kim, “A new logging-based IP traceback approachusing data mining techniques,” *J. Internet Serv. Inf. Security*, vol. 3,no. 3/4, pp. 72–80, Nov. 2013.

12. K. A. Garcia, R. Monroy, L. A. Trejo, and C. Mex-Perera, “Analyzinglog files for postmortem intrusion detection,” *IEEE Trans. Syst., Man,Cybern., Part C: Appl. Rev.*, vol. 42, no. 6, pp. 1690–1704, Nov. 2012.