IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 46-50

To Detect and Protect TheInternal Intrusion Using Data Mining And Novel Algorithm

Mali Rohini S.[1], UndeShital N.[2]ShedgePrajvaleeta T.[3]Prof.Gholap.P.S.[4]
[1](Student, Computer Engineering, SPCOE, Pune, Maharashtra, India)
[2](Student, Computer Engineering, SPCOE, Pune, Maharashtra, India)
[3](Student,Computer Engineering,SPCOE, Pune, Maharashtra, India)
[4](Assistant Professor, Computer Engineering, SPCOE, Pune, Maharashtra, India)

Abstract: *IDPS uses a local computational grid to detect malicious behaviors in an internal system.Most of the computer system uses user-id and password as login to the system. If any user/admin shares their login id and password to their employee or friends then it will be the weakest point of security. Intrusion detection and protectionin the field of computer science is an important area of research from the last few years. Many approaches of classification have been proposed and they are compared with the existing approaches. In this system we used the techniques like data mining, novel algorithm and forensic technic. These algorithms are shows best results in terms of accuracy detection rate and false rate. In future other machine learning algorithms are used and their results must be compared with the previous approaches.*
Hence,Thesystem designed Intrusion Detection and protection System (IDPS) that implements algorithms for identifying the malicious attacks over a network. The security system, named Internal Intrusion Detection and Protection System (IIDPS), is proposed to detect internal attacks at System Call (SC) level by using data mining,novel algorithm and forensic techniques.

I. Introduction

Now dayComputer security is one of the most serious problemsin the computer domain. Attackers are very easily trying to hack the computer security and behavemaliciously. There are two types of attackers they areInternal attackers and External attackers. Internal attackers are the persons have some access privileges in the network and they are trying to insinuation the security system intentionally or unintentionally. Internal attacks are very difficult to detect in the network system. The External intruders are the outsiders from the network system. Generally, most of the attacks are well-known attacks like pharming attack, distributed denial-of-service (DDoS), eavesdropping attack, and spear-phishing attack, insider attack is most difficult to detect. Ones the internal attacks are detected because firewalls and intrusion detection systems (IDSs) usually defend againstoutside attacks.To identify authorized users, currently, most systems check user ID and password as a login pattern. However, attackers may install Trojans to login patterns or issue a large scale of trials with the assistance of a dictionary toacquireusers-id passwords.When attackers are successfully login to the system, they may access user’s private files or modify or destroy the system settings.

In this paper, we propose a network security system, named Internal Intrusion Detection and Protection System (IIDPS), which detects malicious behaviors of system at system call (SC) level. The IIDPS uses data mining, novel algorithm and forensic techniques to mine system call patterns (SC-patterns) defined as the biggest system call sequence (SC-sequence) it has been repeatedly appeared several times in a user’s log file for the user/admin.

II. Techniques And Algorithms

Internal Intrusion Detectionsystem (IIDS) contains an authentication module and an IIDPS system. The behavioral authenticationusing typing speed is a strong authenticationmethod ituniquely identify the user. The typing speed of a user can be calculated by the total time taken for typing. The typing speed ofaperson should not be same. Different users have different typing behavior so wecan say that it is a strong authentication mechanism.

Forensic Technic:

The adjective forensic comes from the Latin word forensic, meaning“in open court” or “public.” When you describe something as forensic you usually meanthat is has to do with finding evidence to solve a crime relate problems. It has also mean that it used courts or legal system.In forensic science presents a challenging set of many issues that used face recognition. Forensic science is the use of scientific principles and methods that used toanswer questions of interest to a legal system. Forensic science differs from the field like security; in security applications the goal is to prevent incidents from occurring, while in forensic cases typically an incident has already occurred.In case of security or portal scenarios where the administrators have control over the scene and the setup of cameras, in forensics the evidence and surveillance generated is completely uncontrolled by the user of the facial recognition system. Unconstrained lighting conditions, face orientation, and other factors all make the deployment of face recognition systems for surveillance a difficulttask. In our project we used web-camera foe face recognition to detect the authorized and intrusion person.For example, the cameras in places of business are generally pointed at specific locations to spot criminals or employees in camera. Camera are places in that locationinclude entry/exit doors where the opening of the door may allow the contrast of the camera to be overwhelmed or above an employee's head, where the angle will be steep and the camera is more likely to observe the top of the subject's sh than the front of their face. Such conditions lead to the inability to enroll facial images or the worsening of system accuracy rates. Low system accuracy can be disastrous in legal matters. Thus, many forensic organizations have yet to embrace facial recognition as fully as users in the field of security.

Novel Algorithm:

A novel algorithm of artificial immune system for high-dimensional function numerical optimization.The novelty of this paper lies on how to use a new algorithm of Image Metadata to detect internal intrusion and how to protect it in real time inside the cloud providers. In this paper, we propose a novel algorithm to use features and characteristics technologies of Image Metadata, namely, a novel family of Intrusion Detection Systems (IDS) built on Exchangeable Image File Format (EXIF) of Image Metadata. It can assist in eliminating the anonymity advantage of perpetrator which is considered extremely important for the Digital Investigation to detect the attacker in a real time and to prove the authenticity of the images Metadata against attackers.

III. Internal Intrusion Detection System

Internal Intrusion Detection and Protection System (IIDPS), is used to detect insider attacks at System Call (SC) level by using data mining forensic techniques and novel algorithms. The IIDPS creates users' personal or admin profiles to keep track of users' usage habits as their forensic features and using habit file determines whether a valid login user is the account holder or not by comparing his/her current computer usage behaviors of user with the patterns collected in the account holder's personal profile of user. The results demonstrate that the IIDPS's user identify accuracy is 94.29%, whereas the response time is less than 0.45 s, implying that it can prevent a protected system from insider attacks effectively and efficiently. TheInternal Intrusion Detection and Protection System(IIDPS) uses data processing and identification techniques to mine data supervisor call instruction patterns (SC patterns) outlined because the longest supervisor call instruction sequence that has repeatedly seen many times in a very user's of log file for that user. The user's on basic of habit file outlined as associate SC pattern of times showing in a very user's submitted supervisor call (SC) sequence however seldom getting used by different users, are retrieved from the user's laptop using history. The system must study the SCs generated and also the SC-patterns are created by these commands so the IIDPS will find those attacker behaviors issued by them then forestall the protected system from being.

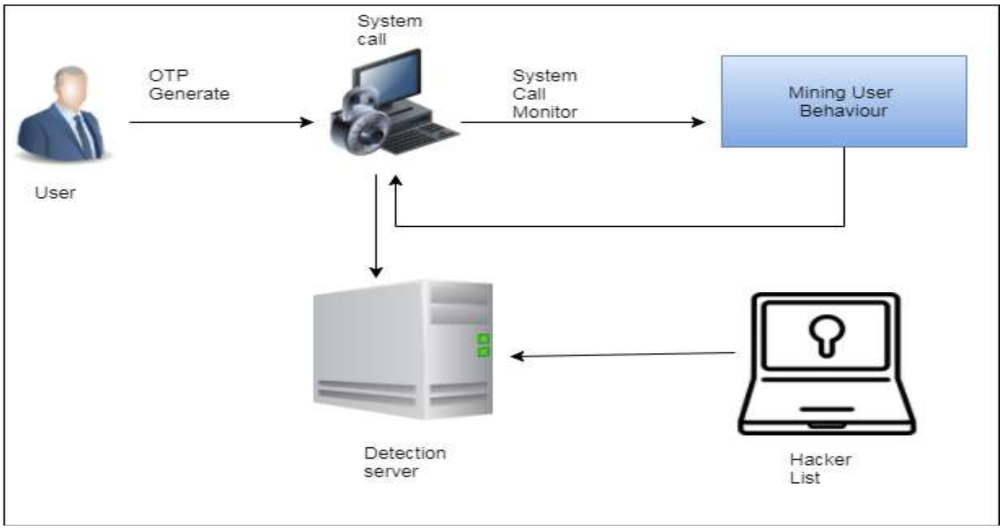

Fig- 1 Basic System architecture

IV. Motivation

- Loses of user private data.
- Financial harm to countries.
- Sharing of political,
- military information by attacker

V. Proposed System

An Internal Intrusion Detection and Protection System (IIDPS), that are detects malicious behaviors of the attacker launched toward a system at SC level. The IIDPS uses data processing and identification techniques to data mine supervisor call instruction patterns (SC patterns) outlined because the longest supervisor call instruction sequence that has repeatedly seen many times in a every user's log file for the user. The user's rhetorical options outlined as associate SC pattern of times showing in a very user's submitted SC sequence however seldom getting used by different users, are retrieved from the user's laptop usage history. The system must study the System Call SCs generated and also the SC-patterns are created by these commands so the An Internal Intrusion Detection and Protection System(IIDPS) will find those malicious behaviors of attacker issued by them the protected system from being attacked.

VI. Hardware Requirements Specification

There should be required devices to interact with software.

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Ram : 256 Mb.

VII. Software Requirements Specification

- Operating system - Windows XP/7.
- Coding Language - JAVA
- IDE - Eclipse Kepler
- Web server - Apache Tomcat 7.

VIII. IIDE Matrix

The IIDE are improves and increases capabilities that are necessary while it decreases and deletes pointers that are unnecessary used in system. To measure these two aspects in systematic way, we have introduced Learning IDEA Matrix. A solid framework for learning opportunity evaluation and knowledge innovation is required otherwise locating ML opportunity would become more difficult. It is used to create build and practice new paradigmsof machine learning with Knowledge Innovation IDEA Matrix framework is developed. This framework is applied to more than one dozen successful ML projects in different domain to locate ML opportunities. This includes from financial advisories, agriculture, teaching learning, and health care. This section introduces this learning IDEA framework for systemic evaluation of problem to identify and evaluate ML opportunities. This framework focuses on highest leverage point so knowledge building and knowledge flow optimization to locate and evaluate machine learning opportunities. This framework is based on flexibility simplicity and efficientlyare applying the knowledge concepts. IDEA matrix learning is based on learning experiments. Here idea is that one should identify opportunities to evaluate and improve learn ability and effectively the overall performance. Idea matrix helps to identify learning problems. Figure depicts IDEA framework for machine learning and systemic knowledge innovation. It used for different parameters marked by I, D, E, and A depicting impact and need of machine learning.

IX. Flow of System

This system can be used to detect the host intrusion detection where host machine comprises the confidential files. Attackers can attack on host machine that attacks would be detect by the system and updated files can be recovered by system. This system can detect the files modification and also prevent the file modification. If files deleted from the host machine permanently then system can't recovered the files

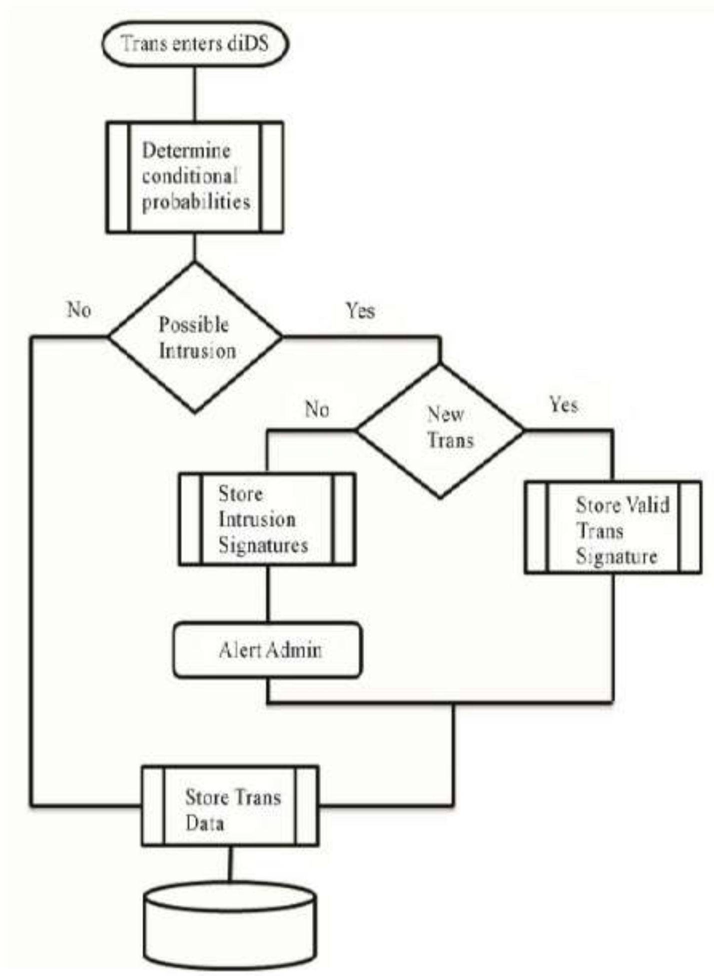

Fig. Flow of IIDE.

X. Algorithms for Internal Intrusion Detection and Protection System

Algorithm 1: The algorithm for generating a user file
Input: u's log file where u is user of the underlying system
Output: u's habit file

1. G=|log file|-|Sliding window|/*|Sliding Window|=|L-window|=|C-window|*/
2. for (i=0 ;i<G-1;i++){
3. for (j=i+1;j<G;j++){
4. for (each of)$\sum_{k=2}^{|Sliding\ window\ |}$ (|Sliding window|-k+1)k-grams in current L-window){
5. for (each of)$\sum_{k'=2}^{|Sliding\ window\ |}$ (|Sliding window|-k'+1)k-grams in current C-window){
6. Compare the k-gram and k'-gram with longest common subsequence algorithm;
7. If(the identified SC –pattern already exit in the habit file)
8. increase the count of the SC –pattern by one;
9. else
10. insert the SC-pattern into habit file with count =1;}}}}

Algorithm 2: Detecting an internal intruder or an attacker
Input: user u's current input SCs, I .e, NSC,(each time only one SC input),
And all user's user profiles
Output: u is suspected as an internal intruder or a known attacker

1. NCS=∅;
2. While (receiving u's input SC, denoted by h){
3. NCS= NCS U{h};
4. If (|NCS |>|Sliding window|){

5. L-window=Right(NCS, |Sliding window|);/*Right (x,y) retrievals the last L-window of y from x*/
6. For (j=| NCS |-|Sliding window|;j>0;j-){
7. For (j=| NCS |-|Sliding window| of size z beginning at the position of y from x)
8. Compare k-gram and k’-gram by using comparing logic employed in algorithm 1 to generate NHF ;o
9. For (each user g,1<g<N)
10. Calculate the similarity score Sum(u,j) between NCS and g’s user profile by invoking Eq.(8);
11. If ((|NCS |mod paragraph size)==0){

/* paragraph size =30, meaning we judge whether u is an attacker or the account holder for every 30 input SCs*/

12. Sort similarity scores for all users ;
13. If((the decisive rate of u’s user profile <threshold1)or((the decisive rate of u’s user profile <threshold2)){

/*threshold1 is prefund lower bound of average decisive Rate of user u’s user profile, while threshold is predefine Upper bond of average decisive Rate of attacker profile*.

14. Alert system manager that u is a suspected attacker rather than u himself/herself;}}}

XI. Conclusion

The IIDPS (Internal Intrusion Detection and Protection System) employs data processing and rhetorical techniques to spot the user activity patterns for a user. The time that a habitual behavior pattern seems within the user’s log file is counted, the foremost usually used patterns are filtered out, and then a user’s profile is established. By distinctive a user’s behavior patterns as his/her pc usage habits from the user’s current input, the IIDPS resist suspected attackers. The long run work of business executive attack detection analysis is going to be concerning aggregation the important knowledge so as to check general solutions and models. It’s onerous to gather knowledge from traditional users in many alternative environments. It’s particularly onerous to amass real knowledge from a masker or traitor whereas activity their malicious actions. Whether or not such knowledge were obtainable, it's a lot of possible

References

[1]. S. Gage, A. Sadeghi, C. Stable, and M. Wigand, “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 against phishing 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 to self-protection in computing system,” in Proc. ACM Cloud Autonomic Comput. 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 a dynamic 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 differentiation based 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 commitment for OS-level virtualization,” in Proc. ACM Int. Conf. Autonomic Comput., 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 DDoS attack 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 nodes identification scheme in networkcoding-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 exhaustion attacks in wireless sensor networks,” Comput. Commun., vol. 34, no. 3, pp. 468–484, Mar. 2011.

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

[11]. K. A. Garcia, R. Monroy, L. A. Trejo, and C. MexPerera, “Analyzing log files for postmortem intrusion detection,” IEEE Trans. Syst., Man, Cybern., Part C: Appl. Rev., vol. 42, no. 6, pp. 1690–1704, Nov. 2012.

[12]. M. A. Qadeer, M. Zahid, A. Iqbal, and M. R. Siddiqui, “Network traffic analysis and intrusion detection using packet sniffer,” in Proc. Int. Conf. Commun. Softw. Netw., Singapore, 2010, pp. 313–317.

[13]. S. O’Shaughnessy and G. Gray, “Development and evaluation of a data set generator tool for generating synthetic log files containing computer attack signatures,” Int. J. Ambient Comput. Intell., vol. 3, no. 2, pp. 64–76, Apr. 2011.