ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

# Pattern Driven Internal Intrusion Detection and Protection System Using Data Mining and Forensic Techniques

Dipali Vijay Karche, Prof. Amrit Priydarshi

PG Scholar, Department of Computer Engineering, Dattakala Faculty of Engineering, Pune, Maharashtra, India

Professor, Department of Computer Engineering, Dattakala Faculty of Engineering, Pune, Maharashtra, India

**ABSTRACT**: Security has become a subject of greater concern for internet users. The Internal Intrusion Detection and protection system Using Data Mining and Forensic Techniques (IIDPS) detect malicious behaviours launched towards a system at SC level with the help of Data mining and Forensic Technique. By using data mining and forensic techniques IIDPS system collects system call patterns that has repeatedly found several times in users log file . By extending IIDPS Able not only detect insider attacker but also prevent attacker entering into system through email .Increases accuracy and response time.

**KEYWORDS**: Data mining, intrusion detection and protection, Identify user, user log file, Attacker profile.

## I. INTRODUCTION

Currently password-based user authentication has established on the Internet to grant users access to security critical services. Authentication through username and password is the commonly used technique However; many people share their login patterns with coworkers and request these coworkers to assist co-tasks, thereby making the pattern as one of the weakest points of computer security. Insider attackers, the valid users of a system who attack the system internally, are hard to detect since most intrusion detection systems and firewalls identify and isolate malicious behaviors launched from the outside world of the system only. 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. Although OS-level system calls (SCs) are much more helpful in detecting attackers and identifying users processing a large volume of SCs, mining malicious behaviors from them, and identifying possible attackers for an intrusion are still engineering challenges.

Internal Intrusion Detection and Protection System (IIDPS), is used as security tools in this system to creates users' personal profiles to keep track of users' regular habits as their forensic features and determines whether a authorised login of user or not and if not then comparing users current computer usage behaviours with the patterns collected in the user's personal profile.Internal Intrusion Detection and Protection System (IIDPS), which detects behaviours at SC level. The IIDPS uses data mining and forensic techniques to collect system call patterns that have repeatedly occured several times in a user's personal log According to user's forensic features, defined as an SC-pattern frequently appearing in a user's submitted habits , but rarely being used by other users, are find out from the user's computer usage history .IIDPS detect internal attacker but it cannot provide protection from it.
Thus the proposed system in this paper, make use of IIDPS. The IIDPS creates users' personal profiles to keep track of users' usage habits as their forensic features and determines whether a valid login user is the account holder or not by comparing his/her current computer usage behaviors with the patterns collected in the account holder's personal profile. If it is found that it is attacker then system inform to user that his/her account is handled by another person by using email with ip-address and resist that user from further working on that login.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Thus the main objectives of this system are:
1. Identify user is insider attacker or not.
2. Improve accuracy of detection.
3. Improve response time to inform user.
4. Very effectively resist attacker.

The remainder of the paper is organized as follows:
Section 2 describes the previous work related to intrusion detection. Then the implementation details of the proposed framework are provided in Section 3. Experimental evaluation and details of dataset is described in Section 4. Section 5 concludes the paper.

## II. RELATED WORKS

### 2.1 Thwart a Phisher with Trusted Computing [2]

In [2] this paper, author GajekAhmad-Reza Sadeghi, Christian Stubble, and Marcel Winandy presents of compartmentalization for isolating applicationsof different trust level, and a trusted wallet for storing credentials and authenticating sensitive services.

### 2.2. Analyzing Log Files for Postmortem Intrusion Detection [3]

Approach for postmortem intrusion detection, which factors malicious behavior, thus, speeds up the process of locating the execution of an exploit. Intrusion detection mechanism is a classifier, which separates abnormal behavior fromnormal one. This classifier is built upon a method that combinesa hidden Markov model with k-means.

### 2.4 .Alerting Subsystem for a Keystroke-based User Identification System [4]

The use of Syslog protocol, definedas the communication relay of the Alerting Subsystem. Furthermore, to ensure a certain level of thrust and security forthese message transfers between a host and the server, a combined method of securing the communicationchannel, as well as securing the message contained in a Syslogpacket encapsulation.

### 2.5. Malicious Nodes Identification Scheme [5]

Propose a Malicious node Identification Scheme (MIS) that discover and isolates malicious nodes, so that the pollution attack can cause harm to the network only for a short period of time and the subsequent streaming will no longer be influenced.

### 2.6. Network Traffic Analysis and Intrusion Detection using Packet Sniffer [6]

Network Traffic Analysis and Intrusion Detection using Packet Sniffer design Focuses on the basics of packet sniffer and its working, development of the tool on Linux platform and itsuse for Intrusion Detection. It also discusses ways to detect the presence of such software on the network and to handle them in an efficient way.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

## III. PROPOSED SYSTEM

System Frame work has major components, Detection server,Mining Server, Local computational grid and system call monitor and filter and also have three repository systems such as user log file, user profile ,attacker profile.

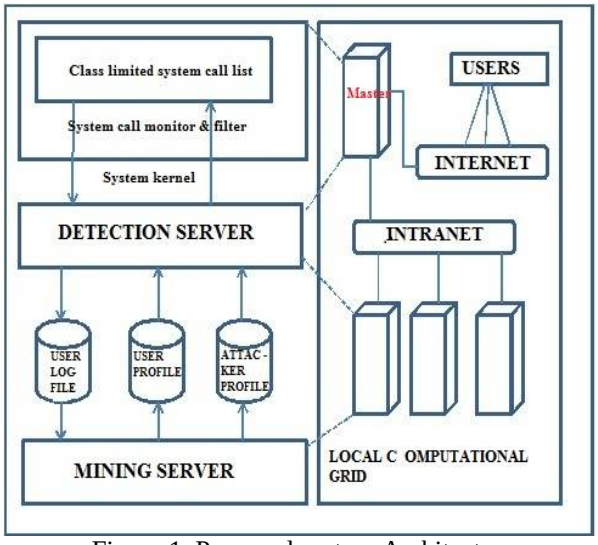


Figure 1. Proposed system Architecture

### 3.1 SC Monitor and Filter:

System call monitor and filter collects system call from system kernel which is in the form of user id, process id and system call. System call s is nothing but the bridge between user applications and services provided by kernel. It also stores the user inputs in the user's log file, which is a file keeping the SCs submitted by the user following their submitted sequence.

In execution of simple commands number of system calls generated hence it's needed to filter that system calls which are repeatedly used. To find out which type system call generated, static model named as frequency-inverse document frequency (TF-IDF) is used.

$$\mathrm{TF}_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{k=h} n_{k,j}}$$

Where *ni, j* is the number of times that *ti* is issued during the execution of *j*, *h* is the number of different SCs generated when *j* is executed, and the denominator sums up the numbers of times that all these SCs are launched. The inverse document frequency (IDF), the measure of the importance of *ti* among all concerned shell commands, is defined as

$$\mathrm{IDF}_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|}$$

where|*D*|, the cardinality of *D*, is the total number of shell commands in the concerned corpus and *{j : ti ∈dj}* is the set of shell commands *dj* , in which each member generates *ti* during its execution. The TF-IDF weight of *ti* generated by *j* is defined as

$$(\text{TF-IDF})_{i,j} = \mathrm{TF}_{i,j} \times \mathrm{IDF}_i.$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

TF-IDF weight as one of the weighting methods in data mining and information retrieval domains increases proportionally to the number of times an SC appears in a user log file.

**3.2 Mining Server**

A mining server extracts SC-sequence generated by a user *u* from *u*'s log file, counts the times that a Specific SC-pattern appears in the file, and stores the result in SC-pattern, appearance counts format in *u*'s habit file. With the help of data mining techniques mining server find out users habits which are stored in user profile. After that compare this user habit with all other users' habit to identify malicious behaviour of attacker. In this process two steps are involved

1. Mining User and Attacker Habits

The IIDPS processes SCs collected in *u*'s log file with a two sliding window, named a log-sliding window (L-window for short), which is used to identify consecutive SCs of size | Sliding window| along their submitted sequence and partition the SCs in the window into *k*-grams where *k* is the number of consecutive SCs, *k* = 2, 3, 4, . . . , |Sliding window | and another is C-window(compared-sliding window) to identify other SC-patterns also in *u*'s log file. This time, *k*' consecutive SCs, preserving their submitted sequence, are extracted from a C-window to generate a total of (| Sliding window| − *k*' + 1) *k*' – grams k'_ = 2, 3, 4, |Sliding window|

*2*. Creating User Profiles and Attacker Profile

Pattern collected in an attacker profile can be classified into user pattern and Attacker-specific Pattern. The latter can be used to identify who the possible attackers are when a protected system is attacked by attacker-specific pattern.

**Algorithm 1** Creating user, attacker profile and detect intruder

1. G=|log file|-|sliding window|
2. For (i=0; i<=G-1; i++) {
   For (j=i+1; j<=G; j++) {
   Collect all k-grams in current L-window
3. Collect all k'-grams in C-window
4. Compare k and k'-grams
5. If (identified Sc pattern already exist in habit file)
   Count + 1
   Else
   Insert pattern into habit file

Patterns 'similarity weights are calculated to filter out those patterns commonly used by all or most users. Then, the output result is compared with all other users' habit files in the underlying system to further identify *u*'s specific patterns. Finally, the similarity weight is computed to generate *u*p's user profile.

**3.3 Detection Server**

Detection server compare attacker profile with user profile which shows malicious behavior .If there is intrusion detected then notifies to the SC monitor and filter user from the protected system to prevent user from continuously attack.

The detection server captures the patterns sent to the kernel by the underlying user *u* when *u* is executing commands and stores the pattern in the *u*'s log file. After this, the server tries to identify whether *u* is the underlying account holder or not by calculating the similarity score between the newly generated SCs, denoted by *user current input.*

**Algorithm 2:** Detection server detects an internal intruder or an attacker.

1. User current input=ϕ
2...while receiving user input denoted by h {
   User current input=user current input ∪ {h}

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

3. If (user input > sliding window) {
For( j=user current input- sliding window;j>0;j--){
4.C-window=Mid(user current window,j,sliding window)
5.compare k-gram and k'-gram
6.Calculate Sc pattern similarity weights
4.Sort similarity scores for all users
5 if(decisive rate for user profile < threshold1)
system alerts u as user profile
else if(decisive rate for attacker profile< threshold2)
system alert u as attacker profile.

All attackers patterns are also presented in the format of a profile. Given a user profile, we can Determine whether the *user current input* includes attacker-specific attack patterns or not by employing the process similar to that of judging whether *u* is the holder of the account that *u* logs in. After calculating the similarity scores between the corresponding and all users' user profiles, if the decisive rate of the attacker profile is higher, we then suspect that *u* is an attacker and the IIDPS will produce alert message and block that user to prevent him/her from continuously attacking the protected system.

### 3.4 Computational Grid

Detection server and the mining server are run on the local computational grid to support the IIDPS's onlinedetection and mining speeds and increase its detection and mining capability. The computational grid is nothing but the collection of internally connected computers working together as a single integrated computing resource.
The functions of the mining server are accomplished by the grid processors with a hostile that defines computation resources for the mining server. The detection server is implemented on the master node of the computational grid with the hostile defining computation resources for the detection server.

## IV. EXPERIMENTAL RESULTS

### 4.2 Performance Evaluation

For evaluating the performance of the system, the metrics

$$\text{Detection server} = \frac{\text{No of attacker detected}}{\text{No. of times of login}}$$

Number of attacks detection increases with number of times of login of that particular user .Thus, it becomes simpler to identify an attacker.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

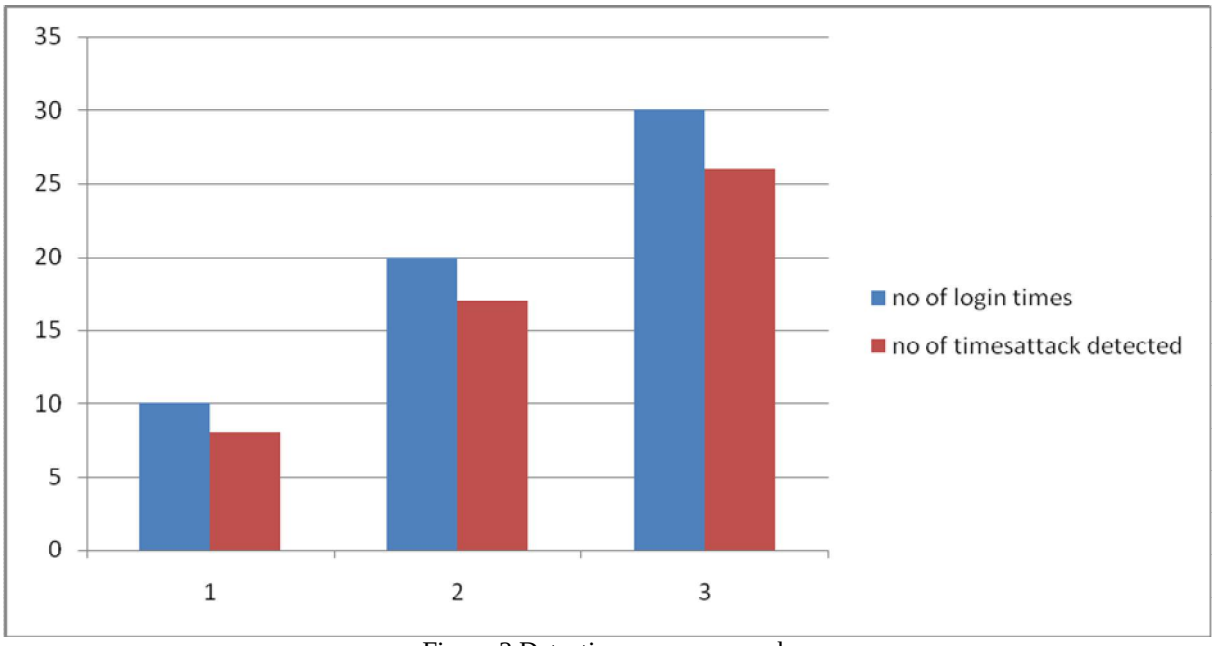


Figure 2.Detection accuracy graph

**4.3 Outcome**

We tested performance of this system by using dataset In the IIDPS, all collected data are analyzed by a data mining tool in which predictability and predictiveness are two parameters utilized to evaluate s weights the obtained results are as in Table 2.

TABLE: Predict Rating

| Commands | Frequency | Predictability | Predicitiveness |
|---|---|---|---|
| Write | 1 | 0.125 | 0.1 |
| Read | 1 | 0.125 | 0.1 |
| Modified | 2 | 0.25 | 0.2 |
| Delete | 2 | 0.25 | 1.0 |
| Encryption | 1 | 0.125 | 1.0 |
| Decryption | 1 | 0.125 | 0.1 |

All attackers' patterns are also presented in the format of a profile. After calculating the similarity scores between the corresponding user *and* all users' user profiles, if the decisive rate of the attacker profile is higher we then suspect that user *is* an attacker.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

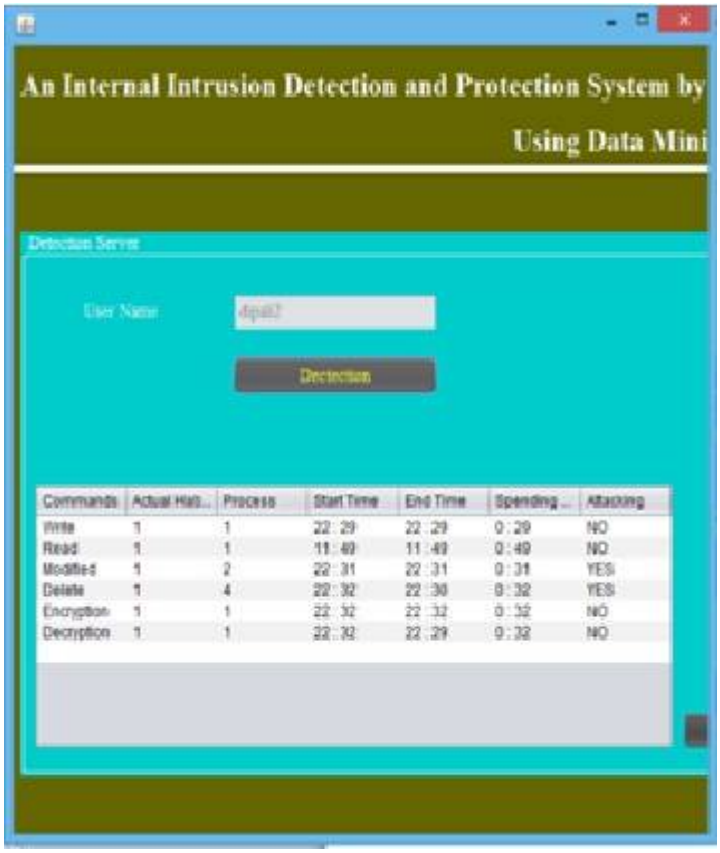


Figure 3.Result summary of detection attacker

The IIDPS will produce an alert message to user register email and block that user to prevent him/her from Continuously attacking the protected system. Following figure 4 shows file operation where attack actually occurs.


Figure 4.Attacker graph

## V. CONCLUSION

Thus it seems that, proposed system employs data mining and forensic techniques to identify the representative SC-patterns for a user. The time that a habitual pattern appears in the user's log file is counted, the most commonly used

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

patterns are filtered out, and then a user's profile is established. By identifying a user's patterns as his/her computer usage habits from the user's current input, the IIDPS resists suspected attackers.

## ACKNOWLEDGMENT

We would like to express our gratitude towards Dattakala Facuty of Engineering.for providing encouraging environment which help us for carrying this research work.

## REFERENCES

[1]Fang-YieLeu, Kun-Lin Tsai, Yi-Ting Hsiao, and Chao-Tung Yang,'' An Internal Intrusion Detection and Protection System by Using Data Mining and Forensic Techniques",IEEE Int. Conf. Avail., Rel. Security, Taiwan,pp 1932-8184,2015
[2] 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.
[3]K. A. Garcia, R. Monroy, L. A. Trejo, and C. Mex-Perera, "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.
[4] S. C. Arseni, E. C. Popovici, L. A. Stancu, O. G. Guta, and S. V. Halunga,"Securing an alerting subsystem for a keystroke-based user identification system," in *Proc. Int. Conf. Commun.*, Bucharest, Romania, 2014,pp. 1–4.
[5]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.
[6] 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.
[7] B. Sayed, I. Traore, I. Woungang, and M. S. Obaidat, "Biometric authentication using mouse gesture dynamics," *IEEE Syst. J.*, vol. 7, no. 2,pp. 262–274, Jun. 2013.
[8] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Data-streambased intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study," *IEEE Syst. J.*, vol. 9, no. 1, pp. 1–14, Jan. 2014.
[9] S. Yu, K. Sood, and Y. Xiang, "An effective and feasible traceback scheme in mobile internet environment," *IEEE Commun.Lett.*, vol. 18, no. 11,pp. 1911–1914, Nov. 2014.
[10] 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.
[11] S. X. Wu and W. Banzhaf, "The use of computational intelligence inintrusion detection systems: A review," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 1–35, Jan. 2010.
[12] F. Y. Leu, K.W. Hu, and F. C. Jiang "Intrusion detection and identification system using data mining and forensic techniques," *Adv. Inf. Comput.er Security*, vol. 4752, pp. 137–152, 2007.
[13]M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Data-streambased intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study," *IEEE Syst. J.*, vol. 9, no. 1, pp. 1–14,Jan. 2014.
[14]Z. B. Hu, J. Su, and V. P. Shirochin "An intelligent lightweight intrusion detection system with forensics technique," in *Proc. IEEE Workshop Intell.Data Acquisition Adv. Comput. Syst.: Technol. Appl.*, Dortmund,
Germany, 2007, pp. 647–651.
[15]R. J. Roger and M. W. Geatz, *Data Mining: A Tutorial-Based Primer*.Reading, MA, USA: Addison-Wesley, 2002.
Ms DIPALI VIJAY KARCHE, has received B.E degree in Information technology from SVPM,Savitribai Phule Pune University, Pune. Currently pursing Master of Computer Engineering in Dattakala Facuty of Engineering,Bhigwan,pune,Maharashtra,India,
Prof. Amrit Priyadarshi,PHD persuing,Department of Computer Engineering,Dattakala Faculty of Engineering,Pune,amritpriyadarshi@gmail.com