ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

New Approach of an Internal Intrusion Detection and Protection System

Pritee Shendkar, S. M. Sangve

ME Student, Department of Computer Engineering, ZCOER Pune, Savitribai Phule Pune University Pune, India.

Professor, Department of Computer Engineering, ZCOER Pune, Savitribai Phule Pune University Pune, India.

ABSTRACT- There are many computer systems use user IDs andpasswords as the login patterns to validate users. However, mostof the people share their login patterns with co-workers andrequest them to assist co-tasks. This makes the pattern as one ofthe weakest points of computer security. The main problem ofcomputer security is insider attack. Insider attack is one of themost difficult attacks to be detected because firewalls andintrusion detection systems (IDSs) usually defend againstoutside attacks. In an additional research, analyzing theOS-level System Calls (SCs) arehelpful in detecting attackersand identifying users. Therefore, in this paper, a securitysystem, called Internal Intrusion Detection and ProtectionSystem (IIDPS), is designed to find insider attacks. The IIDPScreates user'shabit profiles to keep track of user's habits anddetermines whether the login user is the account holder or notby comparing the current computer usage behaviorsof the userwith the patterns collected in the account holder's habit profile.

KEYWORDS - Data mining, insider attack, system calls (SCs),user'sbehaviors, intrusion detection.

I. INTRODUCTION

We analyze the possibility of using data stream mining for enhancing the security of AMI through an intrusion detection system (IDS), which is a second line of defense after the primary security methods of encryption, authentication, authorization, etc. We propose a realistic and reliable IDS architecture for the whole AMI system, which consists of individual IDSs for three different levels of AMI's components: smart meter, data concentrator, and AMI head end. We also explore the performances of various existing state-of-the-art data stream mining algorithms on a publicly available IDS data set, namely, the KDD Cup 1999 data set. Then, we conduct a feasibility analysis of using these existing data stream mining algorithms, which exhibit varying levels of accuracies, memory requirements, and running times, for the distinct IDSs at AMI's three different components. In this paper we have referred Advanced metering infrastructure (AMI), data stream mining, intrusion detection system (IDS), smart grid (SG).

Data mining (knowledge discovery) is the process ofbringing together data from different perspectives andcritiquing it into useful information. In other words thepractice of examining large pre-existing databases in order toproduce new information is known as data mining. Computersystems have been widely employed to provide users withsimpler and more comfortable lives. However, when peoplemake use of powerful capabilities and processing power ofcomputers, security has been considered as one of the majorand serious problems in the computer domain since attackersvery usually try to break through computer systems andbehave false heartedly, e.g., nicking critical data of acompany, wrecking the systems oreven making the systemsout of work.Generally, among all well-known attacks such asDistributed Denial-of-Service (DDoS), eavesdropping attack,and spear-phishing attack, insider attack is one of the mostdifficult ones to be detected because firewalls and IntrusionDetection Systems (IDSs) usually defend against outsideattacks. To validate users, presently, most systems examineuser ID and password as a login pattern. However, intrudersmay install Trojans to nick victims'login patterns or issue alarge scale of trials with the aidof a dictionary to acquireuser's passwords. When efficacious, they may then log in tothe system, access user's private files, or modify or destroysystem settings. Although OS-level System Calls (SCs) are helpful in detecting attackers and identifying users,processing a large amount of SCs, excavating maliciousbehaviors from them, and finding possible intruders for

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

anintrusion are still implementing challenges.Therefore, in this paper, a security system, called InternalIntrusion Detection and Protection System (IIDPS), whichfinds malicious behaviors launched toward a system at SClevel is designed.

II. LITERATURE SURVEY

We introduce model-based autonomic security management (ASM) approach to estimate, detect and identify security attacks along with planning a sequence of actions to effectively protect the networked computing system. In the proposed approach, sensors collect system and network parameters and send the data to the forecasters and the intrusion detection systems (IDSes). A multi-objective controller selects the optimal protection method to recover the system based on the signature of attacks. The proposed approach is demonstrated on several case studies including Denial of Service (DoS) attacks, SQL Injection attacks and memory exhaustion attacks. In this paper we have referred Autonomic Computing, Self-Protection, intrusion detection systems, Denial of Service (DoS) attacks, SQL Injection attacks and memory exhaustion attacks.[1]

This model is proposed for such an attack based on network traffic flow. In addition, a distributed mechanism for detecting such attacks is also defined. Specific network topology-based patterns are defined to model normal network traffic flow, and to facilitate differentiation between legitimate traffic packets and anomalous attack traffic packets. The performance of the proposed attack detection scheme is evaluated through simulation experiments, in terms of the size of the sensor resource set required for participation in the detection process for achieving a desired level of attack detection accuracy. In this paper we have referred Attack detection, Network traffic, packets, topology.[2]

We propose a novel approach to limiting pollution attacks by rapidly identifying malicious nodes. Our scheme can fully satisfy the requirements of live streaming systems, and achieves much higher efficiency than previous schemes. Each node in our scheme only needs to perform several hash computations for an incoming block, incurring very small computational latency. The space overhead added to each block is only 20 bytes. The verification information given to each node is independent of the streaming content and thus does not need to be redistributed. The simulation results based on real PP Live channel overlays show that the process of identifying malicious nodes only takes a few seconds even in the presence of a large number of malicious nodes.In this paper we referred Malicious nodes identification, Network coding, Peer-to-peer streaming.[3]

We design a practical traceback framework to identify active compromised mobiles in the mobile Internet environment in this letter. In the proposed framework, we creatively use the IMEI number of mobile hardware as unique marks for the traceback purpose. Two-layer traceback tables are designed to collect global attack information and identify local attacking bots, respectively. Our analysis and simulation demonstrate that the proposed traceback method is effective and feasible, and it can identify every possible attacking mobile in the current mobile Internet environment with single packet marking. In this paper we have referred Traceback, mobile Internet, and attack source.[4]

The Intrusion Detection and Identification System (IDIS), which builds a profile for each user in an intranet to keep track of his/her usage habits as forensic features. In this way the IDIS can identify who the underlying user in the intranet is by comparing the user's current inputs with the features collected in the profiles established for all users. User habits are extracted from their usage histories by using data mining techniques. When an attack is discovered, the IDIS switches the user's inputs to a honey pot not only to isolate the user from the underlying system, but also to collect many more attack features by using the honey pot to enrich attack patterns which will improve performance of future detection. In this paper we have referred Forensic Features, Intrusion Detection, Data Mining, Identifying Users.[5]

We proposed A new logging-based IP trace back approach using data mining techniques IPTraceback is a way to search for sources of damage to the network or host computer. IP Traceback method consists of reactive and proactive methods, and the proactive method induces a serious storage overhead. However, a system capable of solving these

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

problems through cluster-based mass storage, digestible packets and hierarchical collections was designed. It not only performs traceback but also communicates with analysis data of other security systems by using the logging methods. In this paper we have referred IP Traceback, logging-based approach, data mining.[6]

We analyze the possibility of using data stream mining for enhancing the security of AMI through an intrusion detection system (IDS), which is a second line of defense after the primary security methods of encryption, authentication, authorization, etc. We propose a realistic and reliable IDS architecture for the whole AMI system, which consists of individual IDSs for three different levels of AMI's components: smart meter, data concentrator, and AMI head end. We also explore the performances of various existing state-of-the-art data stream mining algorithms on a publicly available IDS data set, namely, the KDD Cup 1999 data set. Then, we conduct a feasibility analysis of using these existing data stream mining algorithms, which exhibit varying levels of accuracies, memory requirements, and running times, for the distinct IDSs at AMI's three different components. In this paper we have referred Advanced metering infrastructure (AMI), data stream mining, intrusion detection system (IDS), smart grid (SG).[7]

We design a method of integration between HTTP GET flooding among DDOS attacks and Map Reduce processing for a fast attack detection in cloud computing environment. This method is possible to ensure the availability of the target system for accurate and reliable detection based on HTTP GET flooding. In experiments, the processing time for performance evaluation compares a pattern detection of attack features with the Snort detection. In this paper we have referred DDoS Attack, HTTP GET Flooding Attack, Web Security, Map Reduce.[8]

III. PROPOSED SYSTEM

The IIDPS includes an SC monitor and filter,aIntrusion detection, three repositories containing user log files, user habit, an attacker habit, userprofiles, an attacker profile, and a Data Extraction from log File.

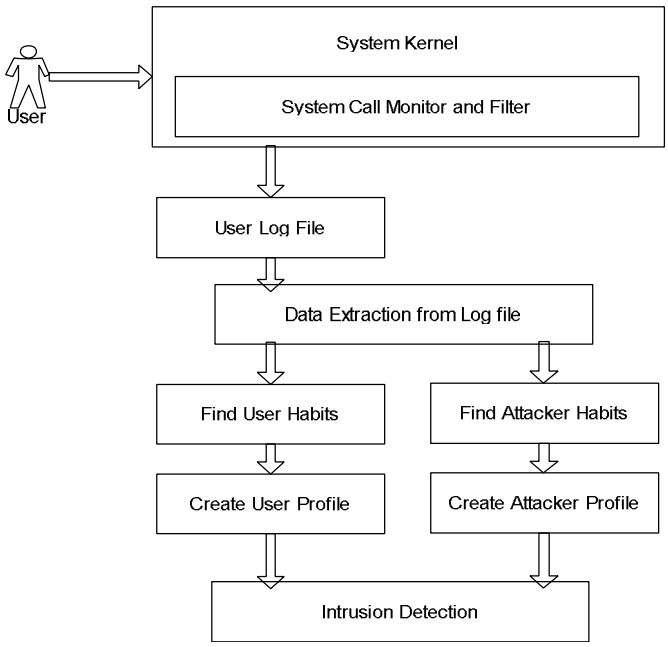


Fig.1. System Architecture

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

System call monitoring and filtering

The system call monitor & filter, collects the System Calls(SCs) i.e. user‘s actions and stores the SCs in the protectedsystem. The user inputs are stored in the user‘s log file i.e., afile used to keep the SCs submitted by the user.F or example,when a user changes his/her password, up to 2916 SCs will begeneratedby capitulating a ―passwd‖ command to a Linuxoperating system, including open(), close(), read(), write(),etc. So, it is hard for a system to keep track of all SCs at thesame time. As a result, we need to sifter out some frequentlyused safe SCs.To find out what SCs are essentialones,the statistic model ofterm frequency-inverse document frequency (TF-IDF) is usedto examine the importance of captured SCs collected in a userlog file.The term frequency (TF) is employed to calculate theweight of the frequency of an SC produced. The inversedocument frequency (IDF),measures the significance of ti i.e.,execution time of commandamong all concerned user actionsrecorded.

Finally the TF-IDFis generated by(TF-IDF)i,j = TFi,j × IDFi ……. (1)

Data Extraction from log File:

A mining server extracts thesystem calls from the user‘s logfile and counts the number of times that a SC-pattern appearsin this file and stores it in the user‘s habit file. A habitfile isanSC-pattern‘sappearance count that is substituted by itscorresponding similarity weight. Then theSC-pattern‘ssimilarity weights are considered to sifter out thoseSC-patterns usually used by all or most users. An attackpattern or a signaturemay be an attacker-specific pattern or apattern usually used by attackers.The IIDPS processes SCs collected in u‘s log file with asliding window, called a log-sliding window (L-window),which is used to detect consecutive SCs of size |Slidingwindow| along their collected sequence and partition the SCsin the window into k-grams where ‗k‘ is the number ofconsecutive SCs. In addition, another sliding window of thesame size known as compared-sliding window (C-window) isemployed to find other SC-patterns also in u‘s log file. Thistime, ‗k‘‘ consecutive SCs, stabilizing their submittedsequence, are mined from a C-window to produce a total of(|Sliding window|−k' + 1) k'−grams.

Total time of (k-gram, k‘-gram) comparison, isgiven by,

$$\text{Ttotal} = \frac{(x-y+1)(x-y)}{2} * \frac{y(y-1)}{2} * \frac{y(y-1)}{2}$$

$$\sim = \frac{1}{9}(1-y)^2\ (y)^4 \ldots..(2)$$

Where x = |SC-sequence| and y = |Sliding window|.

Intrusion Detection:

Itdetects an internal intruder or an attacker. Thedetection server records the SCs submitted by the underlyinguser u andstores the SCs in the u‘s log file. Then, the servertries to identify whether u is the valid account holder or not bycomputing the similarity score between the newly generatedSCs, denoted by NSC, in theu‘s current inputs (in u‘s logfile)andtheusagehabits,i.e.,forensicsignatures(also knownasbehavior patterns),storedinu‘suserprofiletoverifyu.Thecurrent habit file NHF is established by |SCSu|-|slidingwindow| pair-wised. The Detection accuracy is theaccuracy of finding out the user‘s identity.

IV. MATHEMATICAL MODEL & DESIGN

It is hard for a system to monitor all SCs at the same time, particularly when many users are running their programs. As a result, we need to filter out some commonly used safe SCs. In the information retrieval domain, the relationship between a term and a document is similar to that between an SC t_i and the command, e.g., j, which generatest_i. The term frequency (TF) employed to measure the weight of the frequency of an SC produced by j is defined as

$$\text{TF}_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{k=h} n_{k,j}} \qquad (1)$$

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Where, $n_{i,j}$ is the number of times that t_i is issued during the execution of j, h is the number of different SCs generated when j is executed, and the denominator $\sum_{k=1}^{k=h} n_{k,j}$ sums up the numbers of times that all these SCs are launched.

The inverse document frequency (IDF), the measure of the importance of t_i among all concerned shell commands, is defined as,

$$\mathrm{IDF}_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \qquad (2)$$

where, $|D|$, the cardinality of D, is the total number of shell commands in the concerned corpus and $\{j: t_i \in d_j\}$ is the set of shell commands d_j , in which each member generates t_i during its execution.

The TF-IDF weight of t_i generated by j is defined as,

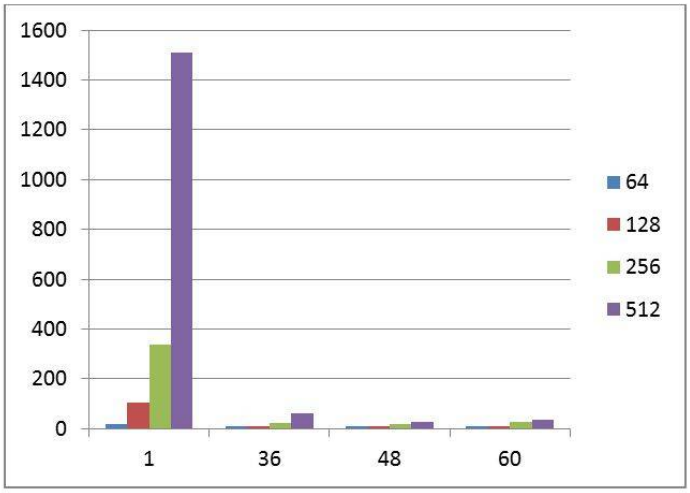
$$(\mathrm{TF\text{-}IDF})_{i,j} = \mathrm{TF}_{i,j} \times \mathrm{IDF}_i. \qquad (3)$$

V. EXPERIMENTAL RESULT

a. Performance Measures Used

1) Measure response time of generating users' habit file.

The speedup S of a parallel system, defined as is one of the parameters used to measure the performance of the parallel system with n processors. The total time required by a program to accomplish a job is 1, f is the sequential part consumed by a single processor, and the time spent by every processor to accomplish its assigned task, is the response time of the n processors.



ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Table III Compute response time of generating user's habit file using parallel computing.

No. of processing cores	64	128	256	512
1	16.34	103.53	335.02	1510.7
36	3.45	10.11	22.94	60.45
48	2.43	9.79	15.92	28.19
60	4.61	8.88	25.98	35.14

b. Detect insider attacks at SC level by using data mining and forensic techniques

The experimental results show that the IIDPS can effectively resist several aforementioned attacks. The outcome extends the features of, confirming that data mining and forensic techniques used for intrusion detection provide effective attack resistance. To verify the feasibility and accuracy of the IIDPS, three experiments were performed. The first defined the decisive rate threshold between the user profile established for u and each of other users' user profiles. The second studied the accuracy for the online detection server when NCSu was submitted by u. The third compared the IIDPS with several state-of-the-art host based IDSs (HIDSs).

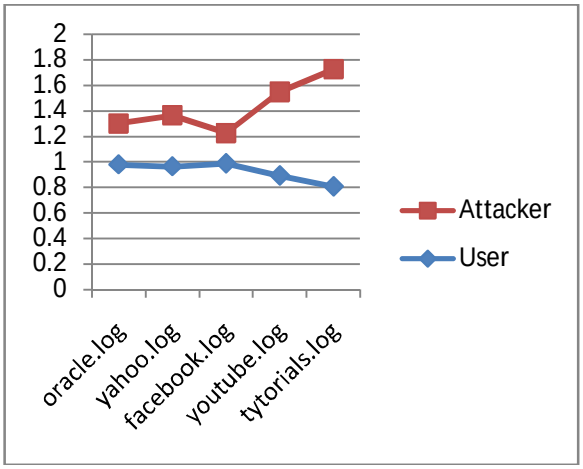


Table II Twelve Users' Decisive Rates Obtained By Cross Comparing Users' Log Files (25% of Test Data)

log file	User	Attacker
oracle.log	0.9788	0.3211
yahoo.log	0.9634	0.4003
facebook.log	0.9875	0.2387
youtube.log	0.8924	0.6541
tytorials.log	0.8073	0.9153

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

c. Comparison with similar systems

Goals	Existing System%	Proposed System%
Accuracy	Accuracy is less.	The average detection accuracy is 94.29% when the decisive rate threshold is 0.9,
Response time	Response time is more.	Response time is less than 0.45 seconds.
Intrusion attacks	Hard to detect since most intrusion detection systems.	To detect insider attacks at SC level by using data mining and forensic techniques.
Security	It provides less security due to intrusion attacks	It provides more security by detecting intrusion attacks.

VI. CONCLUSION

In this paper, an approach that utilizes data mining andforensic techniques to identify the representative SC-patternsfor a user is proposed. The time that a frequent SC- patternappears in the user‘s log file is counted, the most frequentlyused SC-patterns are filtered out, and then a user‘s profile isestablished. By identifying a user‘s SC-patterns as his/hercomputer usage habits from the user‘s current input SCs, theIIDPS resists suspected intruders.

REFERENCES

1] Q. Wang, L. Vu, K. Nahrstedt, and H. Khurana, “MIS: Malicious nodes identification scheme in network-coding-based peer-to-peer streaming,” in Proc. IEEE INFOCOM, San Diego, CA, USA, 2010, pp. 1–5.
2] S. Yu, K. Sood, and Y. Xiang, “An effective and feasible trace back scheme in mobile internet environment,” IEEE Commun.Lett., vol. 18, no. 11, pp. 1911–1914, Nov. 2014.
3] F. Y. Leu, K.W. Hu, and F. C. Jiang “Intrusion detection and identification system using data mining and forensic techniques,” Adv. Inf. Comput. Security, vol. 4752, pp. 137–152, 2007.
4] H. S. Kang and S. R. Kim, “A new logging-based IP trace back approach using data mining techniques,” J. Internet Serv. Inf. Security, vol. 3, no. 3/4, pp. 72–80, Nov. 2013.
5] 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.
6] J. Choi, C. Choi, B. Ko, D. Choi, and P. Kim, “Detectingweb based DDoS attack using MapReduce operations incloud computing environment”, J. Internet Serv. Inf.Security, vol. 3, no. 3/4, pp. 28–37, Nov. 2013.
7] Q. Chen, S. Abdelwahed, and A. Erradi, “A model-basedapproach to self-protection in computing system”,inProc. ACM Cloud Autonomic Comput. Conf., Miami, FL,USA, 2013, pp. 1–10.
8] Z. A. Baig, “Pattern recognition for detecting distributednode exhaustion attacks in wireless sensor networks”,Comput. Commun., vol. 34, no. 3, Mar. 2011.
9] Z. Shan, X. Wang, T. Chiueh, and X. Meng, “Safe sideeffects commitment for OS-level virtualization”, in Proc.ACM Int. Conf. Autonomic Comput., Karlsruhe,Germany, 2011, pp. 111–120