International Journal of Engineering Research in Electronic and Communication Engineering (IJERECE) Vol 3, Issue 5, May 2016

Security System for Internal Attackers Using Data Mining and Forensic Techniques

[1] Ashwini [2] Dr. D Geetha
Dept. of ECE
REVA University, Bangalore India
[1] ashamd601@gmail.com [2] dgeetha@revainstitution.org

***Abstract*— In the current time most boomed technology is network based technology. Based on the current study in network based technology and improved determination on this technology require guarantee dependable process of network based systems. as huge amount of information are saved and transfer from one point to another point while transferring the information it is prone to the attacks so security of the client is an important issue. This paper analysis advanced explanation for the issue of insider assault can find appear in the PC security research literature. In the stage of network based technology, Security is an main problem. As there is the unbelievable raise in the resource and information sharing, the need for security is also raises Intrusion detection system is designed, which monitors the doubtful activity, misuse, unauthorized access etc. Based on the survey on intrusion detection system provides two techniques they are data mining and forensic techniques for finding internal attackers present in system and we also propose two algorithms to overcome inside attack.**

***Keywords*—Intrusion detection, Data mining, Forensic technique System calls (SC).**

I. INTRODUCTION

Currently the majority of the systems deals with authentication that is whenever two persons wants to communication with each other it must be authorized or authenticated. Based on the above criteria the security comes into the above criteria the security comes into picture because it is one of major challenge in the computer network. Attackers are more prone to computer system and behave anonymously .Example accessing the critical data from the client or company , destroying the system by jamming etc [1],[2].

In most of the computer models or present systems while authenticating the models are prone to various types of attacks , it may be insider attacks or outsider attacks. However these attacks try to penetrate into login system and get access to the private data of the individual or company and modify the system settings. In the present system one of the major challenge is how to detect an insider attack that is internal intrusion detection (IDS) because attack packets are frequently issued with fake IPs or attackers may enter a system with legal signing. Intrusion is generally defined as set of actions which breaks the rules of computer system policy of network.

IDS (Intrusion detection system) are the security majors that provide to strengthen the security and privacy of the communication system or it is also defined as detecting and preventing the intrusion in the computer network.

Generally these are classified into two major types detection systems they are (a) Anomaly based detection system: It detects deviation of the patterns from the specified statistical build model. The reason of anomaly might be a mischievous action or different category of intrusion.

The major advantages of anomaly based detection system are

- They are equipped for detecting insider attackers.
- The detection scheme based on habit based profile.

(b) Signature based detection system: This method specifies or matches a specific signature of large database with gathered information this method also known as misuse detection. The major advantages of signature based detection system are

- Often thought to be generously more right at perceiving an intrusion attempt.
- No difficulty of searching reason for alert because of point by point log records.

Time is saved since officials contribute Here in this paper we are majority targeting two techniques they are

1.Data mining: It is a field which s combination of computer science and statistic. Data mining is a process

which discovers patterns in large data set. The main aim of using this technique is extracting information from data set and transfer it into under stable manner.

Advantage: Data mining is known as a solution to managing the analysis of data due to its adaptableness and validity and it is now used widely used for network point [3,4] .

2. Forensic technique: A basic aim of forensic technique used to identify malicious behavior records that programmers introduced in focused on framework. Detecting such malware in a computer network is very hard on the basis that it is difficult to distinguish the motivation behind the parts of executable files.

We starts this paper by giving brief introduction about authentication between two uses while sending massages later on we described IDS and its classifications and at last discuss about data mining and forensic techniques. rest of the part is organized as Section II literature survey. In section III we present the proposed work. Section IV presents, trade of between network security and network forensic. Section V presents the experimental result. section VI gives the conclusion finally.

II. LITERATURE SURVEY

Some researchers have proposed various models which deals with detection of IDS using above two techniques. In ref [5] they proposed effective approach towards IDS by considering four issues they are categorization of data, level of human interaction , lack of labeled information & efficiency of DDoS attacks. In ref [6] they proposed introduce a similarity based method that examines focused on executable folder to recognize a malware exist in a traded off framework. Assigning identical value to the fragments of executable record available in a compromise hard disk against across source documents. We can give various output depend on mismatch of assembly instruction sequences of known hacking equipment with those different executable folder, and suggest several ways to reduce forged data[6]. In [7] presented an intelligent lightweight IDS that utilizes a forensic technique to profile user behaviors and a data mining technique to carry out cooperative attacks. The authors claimed that the system could detect intrusions effectively and efficiently in real time. Following are the sections which describes about majority of attacks and also we mention existing system used in data mining and forensic techniques.

A. Classification of Attacks

A PC structure is a heterogeneous, distributed PC framework, which faces a couple attacks [8]. An attack is confirmation of threat, the malicious action wants to find and misuse the system vulnerability. Vulnerability is some poor ordinary for the system setting up conditions for the danger to rise. The PC system is impacted by the dynamic part a subject (customer) that begins the inquiry for the thing (resource) get to and utilize. Access is collaboration between the subject and the object in the middle of which they exchange information. Event contains the attack and the PC system response to it. Attack can ignore to perform the normal focus for a couple reasons, however and still, after all that there exists believability that the system ends up being weaker. Classes of attack consist of passive attack, active attacks, insider attack, outsider attack, phishing attack and hijack attack these attacks are explained below

1. Passive Attack

An inactive assault is a system assault, in which a framework is observed and infrequently filtered for open ports and vulnerabilities. The reason for existing is exclusively to pick up data about the objective and no information is changed on the objective.

2. Active Attack

An active attack in computer security categorized by the attacker attempting to crack the framework. During this attack, the intruder will bring data into the system and also potentially change data inside the structure.

3. Distributed Attack

It requires to enemy set up code, like a Trojan horse or back-door program, to a “trustworthy” module or software that will be spread to several company and users Distribution attacks focus on the mischievous change of hardware or software at the company or during distribution.

4. Insider Attack

An insider assault is a mischievous assault executed on a system or PC framework by a man with approved framework access.

5. Outsider Attack

Assaults executed by enemies that don't have admittance to direct access to any of the approved hubs in the system. In any case, the enemy may have entry to the physical medium, especially in the event that we are managing remote systems. Thus attacks, for example, replay messages and listening in fall into this characterization. Be that as it may, adapting to this assault is genuinely simple by utilizing conventional security systems, for example, encryption and digital signature.

6. Phishing Attack

In phishing assault the programmer makes a fake site that looks accurately like a known site. The phishing part of the attack is that the programmer then sends an email message trying to trap the customer by clicking on created fake site. Exactly when the customer attempts to sign on with their record information, the developer record the username and secret key and after that tries that information on the actual site.

7. Hijack attack

The most widely recognized strategy for session hijacking is called IP spoofing, when an aggressor utilizes source routed IP packets to embed orders into a moving message between two hubs on a system and masking itself as one of the validated clients.

B. Existing technique used in Data mining

Some of the most important data mining techniques for intrusion detection systems are generic algorithm, fuzzy logic and support vector machine(SVM) are explained in below section.

a)Generic Algorithm: It was presented in the field of computational science. These calculations fit in with the bigger class of Evolutionary Algorithms (EA). They create answers for advancement issues utilizing systems enlivened by normal development, for example, legacy, determination, change and hybrid. From that point forward, they have been connected in different fields with exceptionally encouraging results. In interruption identification, the Genetic Algorithm (GA) is connected to determine an arrangement of characterization guidelines from the system review information. The support confidence system is used as a wellness capacity to judge the nature of every principle. Noteworthy properties of GA are its robust against commotion and self learning abilities. advantage of GA technique is used to report in case of IDS at high detection rate and low false positive rate [9].

b)Fuzzy Logic: It is very suitable for using on IDS because there is no clear limit between anomaly and normal event . The fuzzy logic branch of the system is responsible for dealing with the error of the input data and managing the large number of input parameter. In this approach the data is ordered on the premise of different statistical metrics. These bits of information are connected with fuzzy basis guidelines to characterize them as typical or harmful. There are different other fuzzy data mining systems to concentrate designs that speak to typical conduct for intrusion recognition that describe an collection of change in the current data mining calculations keeping in mind the end goal to expand the effectiveness and accuracy [9,10].

c) Support vector machine: SVM is broadly useful to field of pattern detection and also used in intrusion detection system for detecting malicious code changes in system. It separate information focuses into two classes +1 and - 1 utilizing hyper plane in light of the fact that it is binary grouping classifier. +1 speaks to typical information and -1 for suspicious information. Hyper plane can be communicated as: w. x +b=0 Where w={w1,w2,.....,wn} are weight vector for n properties A={A1,A2,.....,An}, x={x1,x2,.....,xn} are attribute qualities and b is a scalar. The principle objective of SVM is to locate a straight ideal hyper plane so that the edge of partition between the two classes is amplified. The SVM utilizes a segment of the information to prepare the framework [10,11].

C. Existing technique used in Forensic Technique

Similarity Algorithm

In this section it uses to find malicious executable files for computer forensic[6] and this algorithm is divided into two part they are

i) similarity for found fragments

ii) Order of same sections

i) similarity for found fragments

In this part explained how to find similarity for two same fragment between two executable documents. On the off chance that we identify n comparative pieces from target program then there exist n parts in source program despite the fact that they can be copied. we explain below

$M=\{a_1, a_2, a_3.... a_n\}$, $S =\{ a_1', a_2', a_3'.... a_n'\}$ also $\forall a_i \epsilon M$ there must be $a_i' \epsilon S$ which means same to a_i and each couple a_i and a_i' has its individual similarity. There is another difficulty for calculating similarity for finding same fragments. Every similar value for couple a_i and a_i' do not have similar weight because length of each fragment entirely different. Occasionally the length of one fragment shorter than 10 bytes and for longer it should larger than 100 bytes. In this pair case do not have same weight for finding the total similarity detected fragments. Hence weight of every fragment depends on their length.

Similarity of compare blocks(CBs) equation is given by

$$CBs\ (m, s)= \sum_{i=1}^{n}(w_i \times sim(a_i,a_i'))$$

$$\text{where } w_i= \frac{\min(size(ai),size(ai'))\times sim(ai,ai')}{\sum_{j=1}^{n}(\min(size(a),size(aj'))\times sim(aj,aj')}$$

$$\text{where } a_i'= \arg\max(sim(a_i,a_i')),\text{for}(k=1,2,...n)$$

ii) Order of same sections

When likeness has been identified by taking the length of blocks into thought, we need to calculated similarity based on detected similar fragments. luckily, it not a innovative problem. We use the similar technique that has been used for contrasting two string. For this purpose make use of the Damerau-Levenshtein edit distance[12] for sequence matching that is given as

$$\text{Sequence matching} = 1- \frac{dist(m,s)}{max\ (length(m),length(s))}$$

Next analyze two probabilities for the similarity among frequencies of two sets of assembly instructions and their direct in the series. If there are a few sections that assure the above two conditions, At this time consider them similar with additional assurance. As the two conditions are not dependent, we are capable of get the final similarity value by multiplying the results similarity of CBs and sequence matching. Thus, it calculate the similarity between two sequences as follows

Similarity (m, s) = Similarity of CBs(m, s)×Sequence matching (m, s)

III. PROPOSED WORK

1. System Framework

System framework as shown in fig 1, consist of four modules they are classified as system call (SC) monitor and filter, detection server, mining server and local computational grid. The system call monitor and filter fixed into the kernel of the PC and it gathered those SCs, yield to the kernel. These SCs are stored in the protected manner for example uid, pid these represent as user id and process id respectively in SCs pattern.

Those SCs are stored or collected in the use log file. The mining sever uses to check log data with data mining technique to detect the uses pc usage habits as use behavior instructions, which are stored in user profile. The detection server module is used to compare the behavior of users pattern with those collected SCs in the attacker profile. then this comparison is used to detect mischievous behavior and detect the who is inside aggressor in real time.

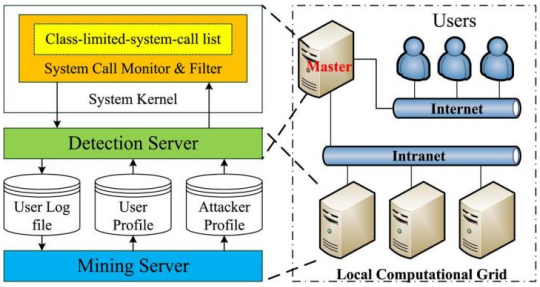

Fig 1. System Architecture

When attackers are find out, the detection module send notice to SC monitor and filter to separate the user from the secured system. Both detection and mining sever run on computational grid to speed up the detection and mining server modules. If any user login into the system by using another authorized person's sign in pattern, the IIDPS detect the attacker who is damaging the system internally by computing the identity score between users present input and behavior pattern in different uses user profile.

2. Algorithms

In this paper we proposed two algorithm for detect internal intrusion in a system. By using those two algorithm we are computing similar score between current users input and behavior pattern stored in different users user profile. the two algorithm used to overcome the challenges faced by above algorithms we are proposing following algorithm

A. Algorithm for generating up's user habitat file

Step 01:G=|log file|-|sliding window| , where sliding window =|L window|=|C window|

Step 02:For (i=0;i≤G-1;i++) {

Step 03:For (j=i+1;j≤G-1;j++) {

Step04:For (each of $\sum\_(k = 2)\text{^}|sliding\ window|$ (|sliding window|-k+1) k grams in current L window{

Step05:For (each of $\sum\_(k' = 2)\text{^}|sliding\ window|$ (|sliding window|- k′+1) k′ grams in current C window{

Step 06: contrast k and k′ grams with longest common subsequence algorithm.

Step 07: if(the identified sc pattern already exists in the habit file)

Step 08:Increase the count of the sc pattern by one:

Step 09:else

Step 10: Insert the sc pattern into the habit file with count=1;}}}}

B. Algorithm for detect inside attackers

Here generally we are detecting an attacker. Input given is users current input SCs that is NSC_u and all users profile.
output we get is whether user is an internal intruder or known attacker.
Step 01:$NCS_u = \Phi$;

Step 02:while(receiving u's input sc,denoted by h){

Step 03: $NCS_u = NCS_u u\{h\}$;

Step04:if(|NCS_u |>|sliding window|){

Step 05:L window=right(NCS_u ,|sliding window|);

Step 06:for(j=| NCS_u -|sliding window|:j>0;j--){

Step 07:C window=Mid(NCS_u ,j,|slidding window);

Step 08:compare k and k grams by using the comparison logic employed in algorithm 1 to generate $NHF_{u:;}$

Step 09: for each user g,1≤g≤N

Step 10:calculate the similarity score sim(u,j) between NCS and g's user profile.

Step 11:if ((|NCS_u | mod paragraph size)==0){

Step 12: sort similarity scores for all users;

Step 13: if ((the decisive rate of u's user profile<$threshold_1$) or (the decisive rate of attackers profile>$threshold_2$));
Step 14:Alert system manager that u is a suspected attacker, rather than u himself/herself;}}}}
This algorithm is used to detect an internal intruder.

IV. TRADE OF BETWEEN NETWORK SECURITY AND NETWORK FORENSIC

System security ensures the framework against assault while system legal sciences does not. System security items search for possible hurtful practices related with different assaults and screen the system full day. Network forensics is post mortem enquiry of the attack Din many cases. It is case confined and is begun after crime notification specifically addressing a exact attack.

Network forensics guarantees that the attacker invests additional time and strength to cover his tracks making the assault very expensive. Network criminals are also careful to avoid prosecution for their illegitimate activities. This acts as a deterrent and diminishes network crime rate, thus enhancing security. Network forensics also can start enquiry in real time provided resources are accessible to handle the traffic and investigate it.

V. EXPERIMENTAL RESULTS

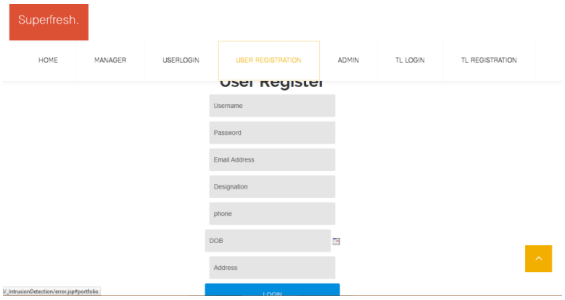

Fig 2 : Registration of user

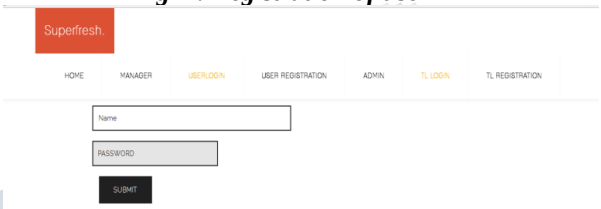

Fig 3: User login

Fig 4: Manager profile

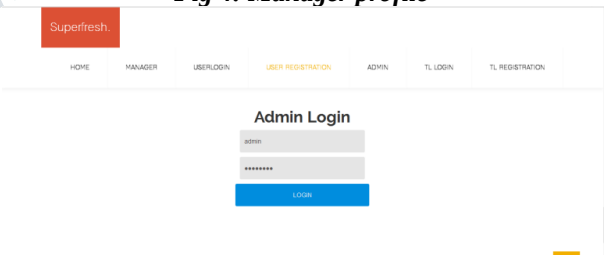

Fig 5: Admin login

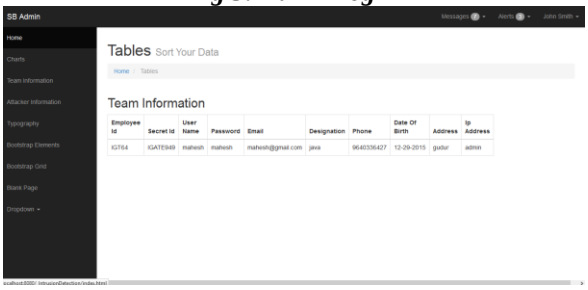

Fig 6: Information of login

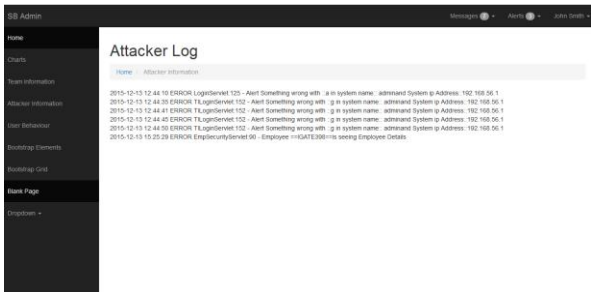

Fig 7: Attacker login information

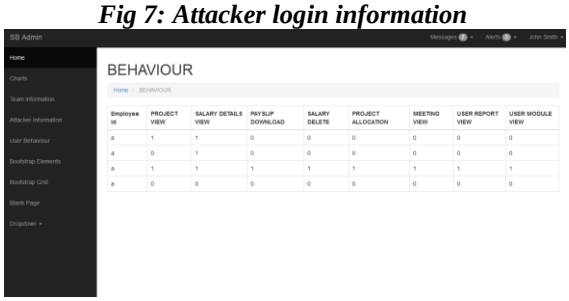

Fig 8: Comparison of user and attacker profile

VI. CONCLUSION

In this Paper we started giving brief introduction about network based system ,attacks and analyzed different techniques of data mining and forensic techniques. With the growing demand of security in the present era data mining is one the technique which is used to the detect the attackers and attackers behavior stored in data base and forensic technique is used to calculate similar weight of files to detect a malware which is present in the system. As Attacks are one of major challenge in the society so in this we are overcoming these attacks by proposed two algorithms.

REFERENCES

[1] Fang-Yie Leu, Kun-Lin Tsai, *Member, IEEE*, Yi-Ting Hsiao, and Chao-Tung Yang," An Internal Intrusion Detection and Protection System by Using Data Mining and Forensic Techniques", IEEE system journal 2015.

[2] Shikha Agrawal, Jitendra Agrawal, " Survey on Anomaly Detection using Data Mining stored in different Techniques", Procedia Computer Science vol. 60, pp. 708 – 713, Aug. 2015.

[3] J.S. Shanthini and S. Rajalakshmi, " Data Mining Technique for Efficient Intrusion Detection System: A Survey ",International Journal On Engineering Technology and Sciences (IJETS) Vol. 2, Issue 6, pp. 2349-3968, Nov. 2015.

[4] Sonam Chourse and Vineet Richhariya, " Survey Paper on Intrusion Detection using Data Mining Techniques", International Journal of Emerging Technology and Advanced Engineering (IJETAE) Vol. 4, Issue 8, pp. 2250-2459, Aug 2014.

[5] Nadiammai G. V., Hemalathain M.,"Effective approach toward Intrusion Detection System using data mining techniques", Cairo University, Elsevier, Egyptian Informatics Journal, 2014, pp. 37-50.

[6] Jun-Hyung Park, Minsoo Kim, Bong-Nam Noh, James B D Joshi, " A Similarity based Technique for Detecting Malicious Executable files for Computer Forensics ", IEEE 2006.

[7] 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.

[8] Jonathon T. Giffin, Somesh Jha, and Barton P. Miller"Automated Discovery of Mimicry Attacks"

[9] Kaur N.. " Survey paper on Data Mining Technique of Intrusion Detection ", International Journal of Science. Engineering and Technology Research, vol.2, Issue 4, pp. 799-804, 2013.

[10] Tang D. H., Cao Z.," Machine Learning - based Intrusion Detection Algorithm", Journal of Computational Information Systems, vol. 5, Issue 6, pp 1825-1831, 2009.

[11] Vaishali B Kosamkar and Sangita S Chaudhari, "Data Mining Algorithms for Intrusion Detection System: An Overview ", International Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS), 2012.

[12] E. Mays, F.J. Damerau, R.L. Mercer. Context-based spelling correction. In *information Proceeding and Management*, 27(5):517-522, 1991.

[13] Natarajan Meghanathan, Sumanth Reddy Allam and Loretta A. Moore, " Tools and Techniques for Network Forensic ", International Journal of Network Security & Its Applications (IJNSA), Vol .1, No.1,April 2009.