ISSN 2395-1621

# Review on Data Mining and Forensic Techniques for Internal Intrusion Detection and Protection System

[#1]Mahendra Badole, [#2]Suhas Bhalerao, [#3]Milind Kambale, [#4]Anmol Shinde,
[#5]Prof. Reshma Patil

[#1234]Dept. of Computer Engineering
[#5]Assistant Professor, Dept. of Computer Engineering

K.J Collage of Engineering And Management Research Pune.

**ABSTRACT**

**There are different ways to protect the data as well as the networks from attackers. Firewalls are used to protect passwords as per need. Many times these are not enough. Due to that systems and networks are always under the observation of thread. Intrusion detection system(IDS) detects unwanted activities of computer system, which are comes through the internet. The manipulation may take form of attacks by hackers. But it is observed that most firewalls and IDS commonly try to protect computer system against outsider attacks. This paper focuses survey on different data mining and forensic techniques to detect and protect internal computer system from intrusion using Internal Intrusion Detection and protection system Using Data Mining and Forensic Techniques (IIDPS) to find out insider attacks at SC level with the help of Data mining and Forensic Technique.**

ARTICLE INFO

**Article History**

Received:20 th January 2018

Received in revised form :

20th January 2018

Accepted: 22nd January 2018

**Published online :**

**14th February 2018**

**Key words: Functionality, Identify user, tf-idf, user log file, Attacker profile.**

## I. INTRODUCTION

Today everyone access the network based information .So via networks many attackers enter into system. These attacks are not only outsider but also insider . In outsider attacks the unauthorized users get access to the systems by using different types of attacks In case of insider attacks the authorized users try to compromise the integrity, confidentiality or availability of resources. Intrusion means any set of activities that try to harm the security goals of the information. Various approaches like as encryption, firewalls, virtual private network, etc., But they were not enough to secure the network fully.

Hence, 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 profiling techniques to mine system call patterns that has repeatedly appeared several times in a user's personal profile. 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.

## II. MOTIVATION

Our main goal for Developing this project a security system, named Internal Intrusion Detection and Protection System (IIDPS), which detects malicious behaviors launched toward a system at SC level.Other objective is that, integrate this system with Hadoop.

## III. LITERATURE SURVEY

**"A Model-based Approach to Self-Protection in SCADA Systems" Qian Chen Sherif Abdelwahed 2010.**
Supervisory Control and Data Acquisition (SCADA) systems, which are widely used in monitoring and controlling critical infrastructure sectors, are highly vulnerable to cyber-attacks. Current security solutions can protect SCADA systems from known cyber assaults, but most solutions require human intervention. This paper applies autonomic computing technology to monitor SCADA system performance, and proactively estimate upcoming attacks for a given system model of a physical infrastructure. We also present the feasibility of intrusion

 

detection systems for known and unknown attack detection. A dynamic intrusion response system is designed to evaluate recommended responses, and appropriate responses are executed to influence attack impacts. We used a case study of a water storage tank to develop an attack that modifies Modbus messages transmitted between slaves and masters. Experimental results show that, with little or no human intervention, the proposed approach enhances the security of the SCADA system, reduces protection time delays, and maintains water storage tank performance. In this paper, autonomic computing technology has been used to self-protect the SCADA system from cyber-attacks. This new technology integrates current security solutions so that the system can proactively monitor, estimate, detect, and react to known and unknown attacks with little or no human intervention. It also ensures the SCADA system is accessible 24/7. We applied the proposed approach to enhance the security of a SCADA system, which controls and monitors a water storage tank. Through the experimental result, we validated that the autonomic SCADA system maintained normal infrastructure operations and regulated the water level back to the normal operation region when alarm conditions were changed by attackers. The overhead time for identifying and protecting the SCADA system was short. It cost 22 sample time to regulate the water level back to normal. In the future, we will simulate more sophisticated cyber-attacks to validate the efficiency of the approach. In addition, we will also employ autonomic computing to self-protect the next generation SCADA systems from cyber assaults.

**"The use of computational intelligence in intrusion detection systems: A review " Shelly Xiaonan Wu, Wolfgang Banzhaf 2010.**

Intrusion detection based up on computational intelligence is currently attracting considerable interest from the research community. Characteristics of computational intelligence (CI) systems, such as adaptation, fault tolerance, high computational speed and error resilience in the face of noisy information fit the requirements of building a good intrusion detection model. Here we want to provide an overview of the research progress in applying CI methods to the problem of intrusion detection.

The scope of this review will be on core methods of CI, including artificial neural networks, fuzzy systems, evolutionary computation, artificial immune systems, swarm intelligence, and soft computing. The research contributions in each field are systematically summarized and compared, allowing us to clearly define existing research challenges, and to highlight promising new research directions. The findings of this review should provide useful insights into the current IDS literature and be a good source for anyone who is interested in the application of CI approaches to IDSs or related fi el ds.

**"Safe Side Effects Commitment for OS-Level Virtualization" Zhiyong Shan Xin Wang Tzi-cker Chiueh 2011.**

A common application of virtual machines (VM) is to use and then throw away, basically treating a VM like a completely isolated and disposable entity. The disadvantage of this approach is that if there is no malicious activity, the user has to re-do all of the work in her actual workspace since there is no easy way to commit (i.e., merge) only the benign updates within the VM back to the host environment. In this work, we develop a VM commitment system called Secom to automatically eliminate malicious state changes when merging the contents of an OS-level VM to the host. Secom consists of three steps: grouping state changes into clusters, distinguishing between benign and malicious clusters, and committing benign clusters.

**"Autonomous Fault Detection in Self-Healing Systems using Restricted Boltzmann Machines" Chris Schneider Adam Barker Simon Dobson 2014.**

Autonomously detecting and recovering from faults is one approach for reducing the operational complexity and costs associated with managing computing environments. We present a novel methodology for autonomously generating investigation leads that help identify systems faults, and extends our previous work in this area by leveraging Restricted Boltzmann Machines (RBMs) and contrastive divergence learning to analyse changes in historical feature data. This allows us to heuristically identify the root cause of a fault, and demonstrate an improvement to the state of the art by showing feature data can be predicted heuristically beyond a single instance to include entire sequences of information.

The operational costs of large-scale computing environments are continuing to increase. In order to address this problem, self-managing systems are being developed that reduce the supervisory needs of computing environments.

Self-healing systems are one such example, and operate by autonomously detecting then recovering from faults. Although there have been numerous advances in both of these aspects, most self-healing systems continue to require periodic human oversight. This constraint poses challenges for the continued reduction of costs, and restricts self-healing recovery strategies to reactive approaches.

**"A study of secured Design of smart meter with Energy Efficient in Smart grid" M.Asan Nainar, G.Dharani Devi 2015.**

Smart grid replaces analog mechanical meters with digital meters that record usage in real state of affairs. The power grid has been converted into an essential in the recent world. A smart grid is the integration of information and communications technology into electric transmission and distribution networks. Nowadays, the electricity make available manufacturing is grappling with an exceptional array of issues, period from a one requirement gap to getting higher expenses. In addition these and more forces are motivating the necessitate to pertaining the trade. Hence, it makes, is motivating the necessitate for a smart grid. With the increase in unit cost of electricity, there is a need for utilities to replace and renew aging transmission and distribution infrastructure with a pressure of using the assets wisely. Human errors and deliberate errors can be lowered by using smart instruments like smart meters. Smart grid can improve outage management performance by responding faster to repair equipment before it fails unexpectedly. The smart grid can improve load factors and

 

reduce system losses. We can integrate renewable energy projects into the grid.

## IV. PROPOSED SYSTEM

In proposed system, the system propose a security system, named Internal Intrusion Detection and Protection System (IIDPS), which detects malicious behaviors launched toward a system at SC level. In the IIDPS, the SCs collected in the class-limited-SC list, as a key component of the SC monitor and filter, are the SCs prohibited to be used by different groups/classes of users in the underlying system. 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 outcome extends the features, confirming that data mining and forensic techniques used for intrusion detection provide effective attack resistance. IIDPS can detect those malicious behaviors issued by them and then prevent the protected system from being attacked.

The scientific method of gathering and examining information about the past which is then used in a court of law. The word forensic comes from the Latin term forēnsis, meaning "of or before the forum." The history of the term originates from Roman times, during which a criminal charge meant presenting the case before a group of public individuals in the forum. Both the person accused of the crime and the accuser would give speeches based on their sides of the story. The case would be decided in favor of the individual with the best argument and delivery. This origin is the source of the two modern usages of the word forensic – as a form of legal evidence and as a category of public presentation. In modern use, the term forensics in the place of forensic science can be considered correct, as the term forensic is effectively a synonym for legal or related to courts. However, the term is now so closely associated with the scientific field that many dictionaries include the meaning that equates the word forensics with forensic science.

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD),an interdisciplinary subfield of computer science is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system.

Neither the data collection, data preparation, nor result interpretation and reporting are part of the data mining step, but do belong to the overall KDD process as additional steps. The contributions of this paper are identify a user's forensic features by analyzing the corresponding SCs to enhance the accuracy of attack detection able to port the IIDPS to a parallel system to further shorten its detection response time and effectively resist insider attack.

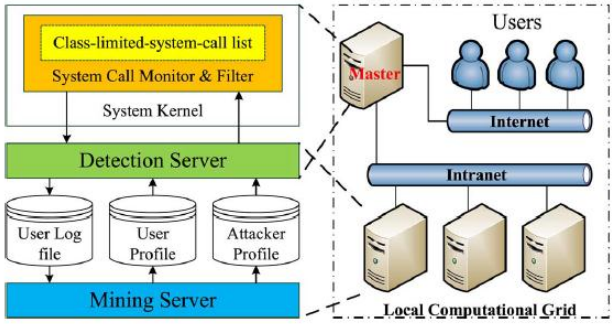


Fig:- System Framework

**Advantages:**

It identify a user's forensic features by analyzing the corresponding SCs to enhance the accuracy of attack detection

It able to port the IIDPS to a parallel system to further shorten its detection response time.

It effectively resist insider attack.

The IIDPS can detect those malicious behaviors issued by them and then prevent the protected system from being attacked.

The mining user profiles by using an unsupervised cluster approach can also improve the performance of the mining process

## V. CONCLUSION

This paper focuses on survey of techniques for data mining and forensic to internal intrusion detection and protection. IIDPS system enables data mining and forensic technique to identify system call , creating user profile and isolated from attacker profile to protect user from internal attack.

## REFRENCES

[1] S. Gajek, A. Sadeghi, C. Stuble, and M. Winandy, "Compartmented se-curity 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 commit-ment for OS-level virtualization," inProc. 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 network-coding-based peer-to-peer streaming," in Proc. IEEE INFOCOM, San Diego, CA, USA, 2010, pp. 1–5.

[10] Jonathon T. Giffin, Somesh Jha, and Barton P. Miller "Automated Discovery of Mimicry Attacks", 2006.

 