Journal of Network and Information Security
6 (2), December 2018, 20-23
http://www.publishingindia.com/jnis

A Big Data Approach Towards Detection of Insider Attack

Vikar Ansar Shaikh[1*] and Tanuja R. Pattanshetti[2]
[1]Department of Computer Engineering and IT, College of Engineering, Pune, Maharashtra, India.
Email: vikarshaikh18@gmail.com
[2]Professor, Department of Computer Engineering and IT, College of Engineering, Pune, Maharashtra, India. Email: trp.comp@coep.ac.in
*Corresponding Author

Abstract: In a big data system, infrastructure is made up like that large number of information is stored on a server which has all client's data and other data also and that data is used by users, basically they host the data. Information security is considered as a major challenge in such system. From a client's standpoint, the biggest risks in using big data systems is that they have to believe on the service provider of big data system, this system are owns and designed by service provider, user have to store and access that data so that they have lot of risk about it.

Methodology:

This work propose a new system architecture in which insider attacks can be identified by using the repetition of data on different nodes in the system. From all of the attacks, Insider attacks are one of today's most difficult cyber security problem that are not well addressed by commonly employed security solutions. Until several scientific research paper published in domain of insider attack, this paper certify that the field can benefit from the proposed structure, taxonomy and novel categorization of research that contribute to the organization of insider attack incidents and the defense solutions used for them. The target of our order is to systematize learning in insider threat research, while utilizing existing ground theory strategy for rigorous literature review Work process of proposed system categories among some classes that include: 1) Events and datasets, 2) Examination of attackers, 3) Process act, and 4) Defense solutions. Special attention is paid to the definitions and taxonomies of the insider threat; we present an auxiliary scientific classification of insider threat incidents, which is based on existing taxonomies.

Outcome:

This paper will help to improve researcher's work in the domain of insider attack, because it provides following things: 1) Time-to-time an updated and mostly available datasets that can be helpful for testing new detection solutions against different attack, 2) References of existing case studies and architecture of insider's behaviors is used for the purpose of testing defense solutions or expanding their coverage, 3) An exchange of knowledge about current patterns and further research directions that can be used for thinking in the insider risk space.

Keywords: Big data, Hadoop, Internal attack detection, Intrusion detection, Security, Spam, Spark.

I. Introduction

In real life insider attack is most difficult attack models to deal. As indicated by an ongoing overview, 27% of all cybercrime attacks were suspected to be committed by insiders, and 30% of respondents showed that the harm dispensed by insiders was more extreme than the harm caused by outside attackers [9]. Comparable numbers accounted for by [10]: "23% of online crime events were found or known to be happened by insiders," while 45% of the user assumed that the effects were worse than the effect of outsiders. According to some review investigating financial crime [9], they have found that in 29% cases internal attack is the main offender. As per a review directed by Vormetric, Inc. [12], just 11% of respondents felt that their association was not vulnerable against insider attacks, while 89% felt at least rate fairly helpless against insider assaults.

In recent years, famous insider attack events done on the media, for example, the high profile data was leaked such cases involving Edward Snowden, Chelsea Manning, and Daniel Ellsberg (see [10] for a collection of famous insider threat incidents). If we see attack performed by outsiders are difficult than insider attacks, this attack are very easy to perform and do not require advanced level of software or knowledge. Insiders are authorized clients, and they may know the loophole and issues of the deployed systems and business processes so it become easy to them for attack. Impression of internal attack is very easy to hide than outsider attack so that it become difficult to detect the insider attack as compared to outsider.

This is likewise exacerbated by the way that organizations pay a lacking measure of consideration and resources to the detection of malicious insider threats McAfee and Evalueserve 2011; Ponemon Foundation 2013. As per Cole and Ring in 2005 there are a few reasons why the insider issue danger has been to somewhat disregard: associations are unaware of the active threat posed by insiders, it is anything but difficult to be trying to claim ignorance in regards to the presence of this danger, what's more, besides, associations fear bad publicity and negative effect they may suffer if such cases are revealed (e.g., A drop in the share value). Nowadays, there is increase in the number of unintentional insider attack in recent years [10]. In this manner, the inspiration for managing insider danger is high and is probably going to develop. Worries about insider threat in the literature are not new, and there is an impressive collection of information in this wide field. In the most recent decade there have additionally been several attempts to survey this field. However, after studying such works, this paper experience various shortcomings and identified that they need an up-to-date and more detailed survey. For example, some reviews focus exclusively on recognition approaches [9] or they do not have a proper knowledge to categorize the literature [13]. The target of this work is to make vast literature survey for insider attack, while making proper systematic information and research led in this area. We see this as critical for both the researchers that design, experimental defense solutions, and in addition the security experts who look to comprehend the issue and are tasked with choosing or executing appropriate defense solutions for their particular needs.

Nowadays big data systems are largely used in various government and enterprise area such as software, finance, retail and medical services. The most frequent use-cases of big data are information retrieval of complex, unstructured information; and ongoing information examination [1]. Be that as it may, alongside its quick market development, the huge information incline likewise has a lot of difficulties and dangers. In a period where extracting information from data is authorized to all, clients are justifiably more distrustful to let service providers host their data far from them. This, alongside the ongoing increment in the quantity of cyber-attacks elevate the importance of security in big data. Despite the fact that protection and security are touted to be essential issues in the big data world, the solutions focus only on utilizing big data systems for efficient security in different areas.

This is unsatisfactory in the big data world where the income depends on the proficient administration of client information. As of late, two unapproved indirect accesses were found in Juniper Systems, firewalls that may have given attackers access to highly classified information. Some essential facts about this specific hack are: (a) It comes at the cost of bargaining national security (b) It demonstrates that even a major network security organization is vulnerable to attacks (c) It is trusted that these indirect accesses were left unfamiliar for almost 3 years; and (d) It was reported that the attackers could have deleted the security logs [8]. To fight the evolving scope of attacks and attackers, it is essential to apply conventional security techniques in new blends or format. Security in popular big data systems such as Hadoop [4] and Spark [5] is provided by authentication through Kerberos [6], Access Control Lists (ACL), observing log and data encryption (to some extent).

But for an insider, especially a traitor, circumventing these mechanisms is not difficult [7]. Traditional security techniques are vital yet not adequate for big data systems. Big data security has some special difficulties concerning the both applications and data. For example, current big data security platforms focus on giving fine-grained security through broad analysis of stored data. But such model unauthorized used the user data in the hands of applications and service provider. This led to the rise of differential privacy that aims to protect sensitive client's information while supporting data analytics. Another such security worry that has been only sometimes addressed in the big data world is insider attacks. Insider assaults are winding up more typical and are viewed as the hardest attacks to identify [2]. There does not exist much in the literature on solutions for insider attacks in general [3]. Existing insider detection strategies focus on client profiling and access control. For these strategies are applicable in the big data world, it is expected that arrangement is an uncommon occasion. Despite the assumption holds true in most cases, the genuine disadvantage with existing insider detection strategies is their inability to be applied in distributed compute environments. To the best of our insight, there is no vigorous answer for identifying or averting insider threats within big data infrastructures. Be that as it may, it is significant to address the issue of insider attacks in big data systems for three main reasons: (a) traitor within the suppliers' association will have the capacity to go around the security system set up (b) sensitivity of client information stored in the system is expanding by the day; and (c) there is no accord or far reaching concurrence on all around characterized security models in the big data community.

In Section 2 Survey Approach of this paper is given, while in Section 3 Survey Scope is given. In Section 4 Contribution given to this work is explained. Section 5 is related to Existing survey on current subject and in last Section 6 Defense solution is explained.

II. Survey Approach

Our primary target of this work is to address the recognized gaps and combine the information contained in existing surveys, by including a more exhaustive and up-to-date literature set, emphasizing the audit and unification of scientific classifications, and utilizing a reexamined and refreshed the list of sources of the literature. Overall, our example set contained 322 works, and 108 of them were filtered out based on our incorporation and prohibition criteria. Note that given the vast amount of work in the field, the objective of this work was not to exhaustively cover the majority literature, but rather to choose a

sufficiently vast literature set to survey the cutting edge and to propose a sensible classification of it.

III. Survey Scope

The extent of our review depends on the following criteria: a) The articles included in this study were chosen dependent on a widespread view of the insider threat problem, ranging from definitions and scientific categorizations of insider threat, and examination and demonstrating of attackers, to theoretical defense solutions and their verification of ideas; b) We concentrated on studies for which the insider threat issue was the primary subject. We did, however, include several examples of masquerade detection approaches, which are identified with the data fraud issue but at the same time are viewed as a component of the insider threat issue; The more established form of an investigation was typically superseded by the more up to date one, aside from in cases in which the new version contained less points of interest or on the other hand focused on another idea. After specifying the scope of the study, we connected an iterative procedure to develop a literature categorization depend on a grounds hypothesis for rigorous literature review [11], Which serves as guidelines on an investigation and introduction of discoveries in a specific field of Survey of Insider Threat Scientific categorizations, Examination, Modeling, and Countermeasures 3 explore. While preparing the arrangement of papers, we recognized a few unique ideas, proposed a work process based classification, and reworked the list of sources according to it. Note that our fundamental classifications are not intended to be disjoint, as there are works that address aspects of different proposed classes. However, we trust that our order offers valuable measurements with which to characterize works in the literature, while at the same time enabling researchers to identify the relevant related work.

IV. Contributions

In total, this work shows a novel insider threat overview that means to be exhaustive, yet concise and simple to access by scientists searching for new roads to investigate or to find out about the subject. The primary contributions of this study can be summarized as follows: a) To the best of our knowledge, this is the main work that efficiently classifies heterogeneous insider threat considers and subsequently empowers readers to get a panoptic view on this disparate subject, including, an audit of datasets, contextual analyses, investigation and demonstrating of the attackers, simulations, calculated and practical defense solutions, and best practices; b) We review existing scientific categorizations of the insider threat issue, and depend on them, we propose a commonsense and bound together scientific categorization that can be utilized to order: 1) an insider threat incident, or 2) specialization / inclusion of a defense solution; c) We aggregate total data about freely accessible datasets that can be used for testing a detection solution and examination with different works included in this review that have previously utilized the datasets; d) We recognize open inquiries and difficulties in insider threat detection, prevention, and mitigation, while proposing further research directions.

V. Existing Surveys

In this section we provide a brief summary of existing surveys involving insider threats, and afterward we depict how our proposed scientific classification varies from them. According to knowledge about the target system [3] made the types of malicious insiders into two parts as: traitors and masqueraders. The creators looked into the literature on insider detection works and separated the works into three kinds of methodologies: a) “host-based client profiling approaches,” b) “network level methodologies,” and c) “integrated methodologies.” System level and host-based profiling may have, according to the authors, a high possibility of distinguishing traits, while host-based client profiling might be effective in recognizing masqueraders. As well as, the malicious actions of insiders occur at the time of application and business process level is summited by the author. This paper led a study containing an overview of three types of methods to deal with insider attack detection: 1) “sociological, psychological, organizational,” 2) “socio-technical,” and 3) “technical.” The creators stressed that effective insider threat identification methods require a combination of various methodologies [13]. Classified related works into six classes: 1) Psychological and social theories 2) Anomaly-based methodologies 3) Honeypot-based methodologies 4) Graph-based methodologies 5) Game theory methodologies 6) Motivating studies. The creators expected to reinterpret standards and consequences of the methodologies included in their review.

Definition:

Insiders:

Pfleeger [14] *et al.* in 2010 defines insider as a person who has authorized access to an organization’s computers and networks. The report from RAND Corp. Aimed at database security, Garfinkel [15] *et al.* in 2002 defines an insider as a subject of the database who has personal knowledge of the information in the confidential field.

VI. Defense Solutions

This section briefly describes defense solutions for the insider threat issues, Means-based categorization of mitigation / prevention methodologies is presented, and this is followed by intrusion-detection-derived classification of insider threat assessment and detection research. A detailed overview of this category is depicted in Fig. 1.

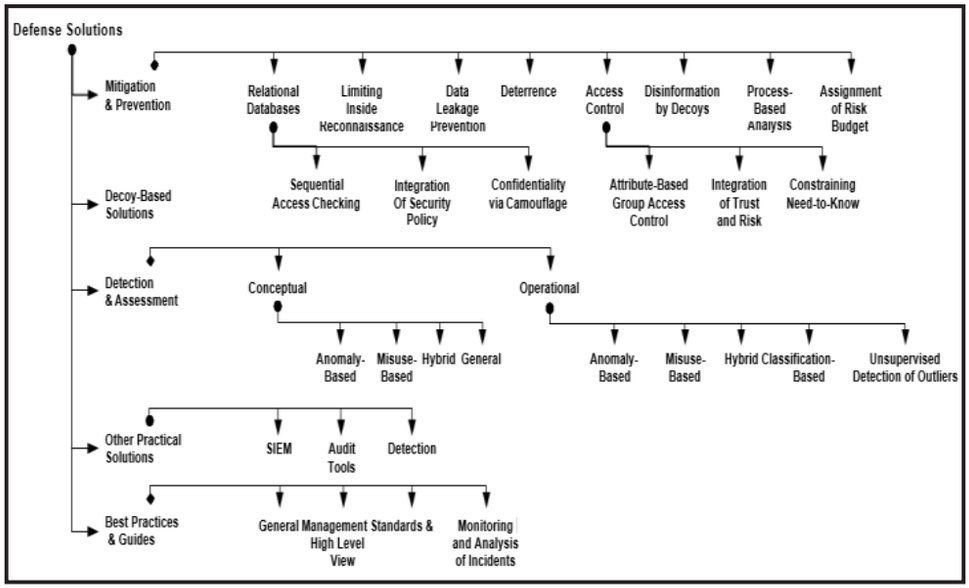

Fig. 1: Defense Solution

VII. Conclusion

From the study of this paper, it defines that insider attack is more harmful because of it is very easy to hide, insider have all access for the system so that they do not need any software and advance knowledge for attack as compared to outsider attack. There are different types of defense solution available if that solution is utilized insider attack can be avoided.

References

[1] S. Aditham, and N. Ranganathan, "A system architecture for the detection of insider attacks in big data systems," *IEEE Transactions on Dependable and Secure Computing*, 2017.

[2] Vormetric. "2015 Insider Threat Report," Vormetric, Inc., 1st September 2015. Web. 1st January 2016.

[3] M. B. Salem, S. Hershkop, and S. J. Stolfo, "A survey of insider attack detection research," *Insider Attack and Cyber Security*, pp. 69-90, Springer, Boston, MA, 2008.

[4] T. White, *Hadoop: The Definitive Guide*, O'Reilly Media, Inc., 2012.

[5] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," *Hot Cloud 10*, pp. 10-10, 2010.

[6] B. C. Neuman, and T. Ts'o, "Kerberos: An authentication service for computer networks," *IEEE Communications Magazine*, vol. 32, no. 9, pp. 33-38, 1994.

[7] S. Aditham, and N. Ranganathan, "A novel framework for mitigating insider attacks in big data systems," *2015 IEEE International Conference on Big Data (Big Data),* IEEE, 2015.

[8] S. Khandelwal, "Juniper firewalls with screenOS back doored since 2012," *The Hacker News*, 18th December 2015.

[9] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa, "Insight into insiders: A survey of insider threat taxonomies, analysis, modeling, and countermeasures," 2018. Available: https://arxiv.org/abs/1805.01612

[10] D. Cappelli, A. P. Moore, R. F. Trzeciak, and T. J. Shimeall, "Common sense guide to prevention and detection of insider threats 3rd ed. - version 3.1," Published by CERT, Software Engineering Institute, Carnegie Mellon University, 2009. Available: http://www.cert.org

[11] J. F. Wolfswinkel, E. Furtmueller, and C. P. M. Wilderom, "Using grounded theory as a method for rigorously reviewing literature," *European Journal of Information Systems*, vol. 22, no. 1, pp. 45-55, 2013.

[12] H. Poll, and A. Kellett, "Vormetric Insider Threat Report," 2015.

[13] A. Azaria, A. Richardson, S. Kraus, and V. S. Subrahmanian, "Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 2, pp. 135-155, 2014.

[14] S. L. Pfleeger, J. B. Predd, J, Hunker, and C. Bulford, "Insiders behaving badly: Addressing bad actors and their actions," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 169-179, 2010.

[15] R. Garfinkel, R. Gopal, and P. Goes, "Privacy protection of binary confidential data against deterministic, stochastic, and insider threat," *Management Science*, vol. 48, no. 6, pp. 749-764, 2002.