ISSN: 2277-9655
Impact Factor: 4.116

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

REVIEW OF APT ATTACKS: HOW BIG DATA FIGHTES BACK

Rukhsana Ambreen*, Prof. Shyam Dubey, Prof. Shahid Nadeem
*M.Tech(CSE), Nuva College of Engineering and Technology, Nagpur, Maharashtra, India
Professor(CSE), Nuva College of Engineering and Technology, Nagpur, Maharashtra, India
Professor(CSE), Nuva College of Engineering and Technology, Nagpur, Maharashtra, India

DOI: 10.5281/zenodo.154215

ABSTRACT

Now a days, threat of previously unknown cyber-attacks are increasing because existing security systems are not able to detect them. The cyber-attacks had simple purposes of leaking personal information by attacking the PC or to reduce the system. The aim of recent heavy blows. attacks has changed from leaking information and destruction of services to attacking big data systems such as critical infrastructures and state agencies. Previous defense technologies to measure these attacks are based on patterns matching methods which are very limited. Because of this reality, In the occurrence of new and previously unknown attack, detections rate become very low and false negative increases. To keep secure from these unknown attacks, which unable to detected with existing technology, We proposed new model based on big data analysis techniques that can take out information from a variety of sources to detect future attacks. Respecting the model on the basis of the future Advanced Persistent Threat (APT) detection and prevention system execution.

Keywords: Cyber-attacks, Security systems, Intrusion detection, Big Data .

INTRODUCTION

The past leaked personal information. This kind of attack is commonly called APT(Advanced Persistent Threat). APT aims to identified system and analyses susceptibility of the system for a long time. Therefore it is hard to prevent and detect APT than traditional attacks and could result heavy damage. Up to today, detection and protection systems for defending against cyber-attacks were firewalls, intrusion detection systems, intrusion prevention systems, anti-viruses solutions, database encryption, DRM solutions and etc. Furthermore, integrated monitoring technologies for managing system logswere used. These security solutions are developed based on signatures and blacklist. Although, according to various reports, intrusion detection systems and intrusion prevention systems are not capable of defend systems against APT attacks because there are no signatures. Therefore to conquer this issue, security communities are beginning to apply heuristic and data mining technologies to detect previously unknown attacks.

Big data has been a great issue in the IT industry for the last couple of years. It defines huge, shortly created and atypical data in digital environment such as text, music, video, and so on. Big data analysis is a technology that searches useful information such as a relation rule, a hidden value from huge data. Big data analysis uses various existing analysis techniques such as machine-learning, AI, data analysis and etc. Among various techniques, focusing on four techniques – prediction, classification, relation rule, atypical . data-mining. It is means that these techniques are useful to detect unknown new attacks. First, prediction is a technique that predicts the future possibility and trend. Regression analysis is a representative prediction technique. Researchers can predict attack possibilities using regressing analysis. Regressing analysis can predict similar behaviours from collected attack logs. Second, classification is a technique that predicts the group of new attack from huge data. Classification helps security administrator to decide direction of protection and analysis. Most used classification techniques are logistic regression analysis and SVM(Support Vector Machine). In this paper, are not proposing effective parallel processing algorithm for real time analysis. Instead of using pattern matching or log analysis for predicting cyber-attacks, believe that can extract valuable information previously unfound from data and status information collected from various sources by big data analysis. Moreover, to apply and validate various analysis methodologies using big data, need professional software and distributed system. In future works, to implement proposed system and get results using real factors and analysis methodologies.

Due to rapid development of Internet and technology, all the machines are connected to each other either by networked system or through mobile communication. The users are producing more and more data through communication media in the unstructured form which is highly demanding and this management of data is the challenging job. The aim is to gather the unstructured data from all the terminals, processed the data to convert into structured form so that accessing of the data would be easier . For this, always a track is kept on data, that this data or event belongs to which category. Consequently, data is analyzed and processed to convert it into meaningful and right information by using the concept of Big Data Analytics. Big Data Analytics accepts the huge data sets and different data types, both half structured and not predefined like videos files, images, audio, web-pages, texts files or electronics mails etc. and convert it into authentic information. Big data analytics describes the simple algorithm for large amount of data without compromising performance. Analysis algorithm are provided directly to database which go beyond the pack and innovate newer additional sophisticated statistical analysis. Big Data Analytics use number of tools to do the analysis and processing of data in significant way. Hadoop is one of the tools which is aimed to improve the performance of data processing. Hadoop is a software framework for storing and processing Big Data and work under Big Data Analytics. It is an open-source tool build on java platform and main objective to improve the performance in terms of data processing on clusters. Hadoop settlement of multiple concepts and modules like ZOO KEEPER Hadoop DFS, Map decreased, HBASE, PIG, HIVE, SQOOP and to perform the easy and fast processing of huge data]. Hadoop is different from Relational databases and can process the high volume, high velocity and high variety of data to generate value In this paper, proposing that the use of Big Data Analytics for analyzing the enterprise data. We discussed a Enterprise data security is challenging task to implement and calls for strong support in terms of security policy formulation and mechanisms. We plan to take up data collection, pretreatment , integration, map reduce and prediction using machine learning techniques. We are developing security alerts which will provide employees with the ability to view the activity. Events will be filtered down and summarized view will be available to each individual employee.

LITERATURE SURVEY

Big data analysis system concept for detecting unknown attacks. Unknown cyber-attacks are increasing because existing security systems are not able to detect them. big data analysis techniques that can extract information from a variety of sources to detect future attacks. the event of new and previously unknown attacks, identify rate becomes very low and incorrect negative increases. To keep safe from these unknown attacks Does not detect future Advanced Persistent Threat(APT) detection.

Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data Big data security analytics is used for the growing practice of organization to gather and analyze security data to detect vulnerabilities and intrusions Security and Information Event Monitoring (SIEM) system. The malicious and targeted attacks have become main subject for government, organization or industion Big data analytics is the process of analyzing big data to find hidden patterns, unknown a mutual relationship between two or more things and other helpful information that can be extracted to make better decisions.

Zero Day Attack Signatures Detection Using Honey pot unexpected behavior. Fault distribution studies show that there is a correlation between the number of lines of code and the number of faults. large number of such vulnerabilities. Longest Common Substring (LCS) algorithm on the packet content of a number of connections going to the same services. Zero day attack or computer threat that tries to exploit computer application vulnerabilities that are unknown to others or undisclosed to the software developer. Cloud Model based Out lier Detection Algorithm for unambiguously explicit and direct data numerical data but There will be a large number of unambiguously explicit and direct data in real life. Some out lier detection algorithm shave been designed. for l data. There are two main problems of out lier detection for categorical data, which Are the similarity measure between unambiguously explicit and direct data objects and the detection efficiency . out lier detection algorithm for unambiguously explicit and direct data Efficient out lier detection can help us make good decisions on erroneous data or prevent the negative influence of malicious and faulty behavior. Many data mining techniques try to reduce the influence of outliers or eliminate them entirely.

Cloud Computing-Based Forensic Analysis for involving two or more parties working together Network Security Management System, Internet security issues remain a major challenge with many security concerns such as Internet viruses, trozans, and phishing strike. Botnets, well-developed distributed network strike, consist of a huge numbers of bots that developed huge volumes of spam or launch Distributed Denial of Service (DDoS) attacks on harmed hosts. A distributed security overlay network with a centralized security center leverages a peer-to-peer communication protocol used in the UTMs collaborative module. These new security

ISSN: 2277-9655
Impact Factor: 4.116

rules are enforced by collaborative UTM and the feedback events of such rules are returned to the security center.

PROPOSED SYSTEM

To establish a defense-in-depth intrusion detection framework. For better attack detection, big data incorporates attack graph analytical procedures into the intrusion detection processes. Note that the design of does not intend to improve any of the existing intrusion detection algorithms; indeed, employs a reconfigurable virtual networking approach to detect and counter the attempts to compromise VMs, thus preventing zombie VMs. A cloud system with hundreds of nodes will have huge amount of alerts lifted by Snort. Not all of these alerts can be depends upon, and an effectual mechanism is required to verify if such alerts require to be inscribed. Since Snort can be programmed to develop alerts with CVE id, one proceed towards that work provides is to match if the alert is literally related to some vulnerability being utilized. If so, the existence of that vulnerability in SAG means that the alert is more likely to be a real strike. Thus, the unreal positive rate will be the joint chances of the related between alerts, which will not high the unreal positive rate compared to each individual unreal positive rate. Moreover, cannot keep aside the case of zero day attack where the vulnerability is discovered by the attacker but is not detected by computer security weakness scanner. In such case, the vigilant being real will be related to as false, given that there does not exist correlated node in SAG. Thus, present research does not inscript how to decrease the incorrect negative rate. It is important to note that security weakness scanner should be able to expose most new vulnerabilities and sync with the new vulnerability database to decrease the chance of Zero-day attacks.

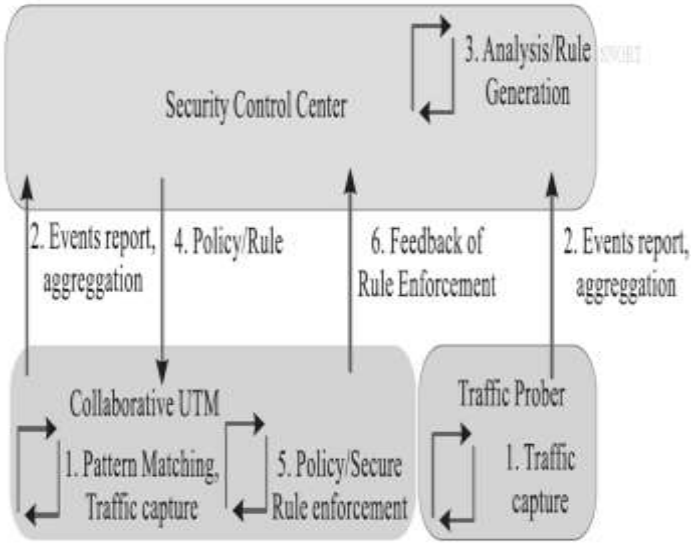

Fig 1. System Flow diagram

ISSN: 2277-9655
Impact Factor: 4.116

TABLE 1
COMPARISON BETWEEN MISUSE DETECTION AND ANOMALY DETECTION

Signature – Based(Misuse Detection)	Behaviour–Based(Anomaly Detection)
Advantages	**Advantages**
-Higher Detection rate, Accuracy for known behaviors. -Simplest and effective method. -Low False alarm rate.	-can examine unknown and more complicated intrusions. - Rate of Missing report is low. -Detect new and unforeseen vulnerabilities.
Disadvantages	**Disadvantages**
- It can detect only known attacks. - Needs a regular update of the rules which are used. - Often no differentiation between an attack attempt and a successful attack. - Rate of Missing report is high.	- Needs to be trained and tuned model carefully, otherwise it tends to false – positives -low detection rate and high false alarm rate. - It can't identify new attacks because intrusion detection depends upon latest model.

CONCLUSION

A multi-phase distributed weak security detection, quantification, and countermeasure selection mechanism called Bigdata, which is built on attack based analytical models and network-based countermeasures. The proposed framework used to maximum advantages Open Flow network programming APIs to build a monitor and control plane over distributed programmable virtual switches in order to significantly improve attack detection and mitigate attack consequences.

REFERENCES

1. Marquand, Robert; Ben Arnoldy; "China Emerges as Leader in Cyberwarfare," The Christian Science Monitor, 14 September 2007, www.csmonitor.com/2007/0914/p01s01 -woap.html
2. Rain; "Analysis of the 2007 Cyber Attacks Against Estonia from the Information Warfare Perspec tive," Proceedings of the 7th European Conferences on Information WR-fare, Plymouth, 2008
3. Brewin, Bob; "U.S., British officials target Chinese as Source of cybe rattacks," Gover n-ment Executive, 4 December 2007, www.govexec.com/defense/2007/12/us-british officials target chinese as source of cyberattacks/25874/
4. Clayton, Mark; "US Oil Industry Hit by Cyberattacks: Was China Involved?," The Chri tian Science Monitor, 25 January 2010, www.csmonitor.com/USA/2010/0125/US
5. Samuel, Henry; "Chip and Pin scam 'Has Netted Millions From British shoppers'," The Telegraph, 10 October 2008, www.telegraph.co.uk/news/uknews
6. Drummond, David; "A New Approach to China," Google Blog, 12 January 2010, http://googleblog.blogspot.com/2010/01/ne
7. A.K.Sood, R.J. Enbody "Targeted Cyber attack: A superset of advanced persistent threats" Security & Privacy, IEEE Volume 11 Issue 1, pages 54-61, Jan-Feb, 2013.
8. Apache Hadoop Project http://hadoop.apache.org
9. "Hadoop Tutorial from Yahoo!", Module 7: Managing HadoopCluster.http://developer.yahoo.com/hadoop/tutorial/module7.html#machines
10. K. Shvachko, H. Kuang, S. Radia and R. Chansler, "The Hadoop distributed file system", in poc. The 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MMST) 2010.
11. F. Cuppens and A. Mige, Alert correlation in a cooperative intrusion detection framework, in Proc. IEEE Symposium on Security and Privacy, Berkeley, California, USA, 2002, pp. 205-215.
12. A. Hofmann, I. Dedinski, B. Sick, and H. de Meer, A novelty driven approach to intrusion alert correlation based on distributed hash tables, in Proc. 2007 IEEE International Conference on Communications (ICC), Glasgow, Scotland, 2007, pp. 71-78.

ISSN: 2277-9655
Impact Factor: 4.116

13. B. Mu, X. Chen, and Z. Chen, A collaborative network security management system in metropolitan are network, in Proc. the 3rd International Conference on Communications and Mobile Computing (CMC), Qingdao, China, 2011, pp. 45-50
14. X. Chen, B. Mu, and Z. Chen, NetSecu: A collaborative network security platform for in-network security, in Proc. the 3rd International Conference on Communications and Mobile Computing (CMC), Qingdao, China, 2011, pp. 59-64.
15. W. H. Allen, Computer forensics, IEEE Security & Privacy, vol. 3, no. 4, pp. 59-62, 2005
16. M. A. Caloyannides, N. Memon, and W. Venema, Digital forensics, IEEE Security & Privacy, vol. 7, no. 2, pp. 1617, 2009.
17. F. Raynal, Y. Berthier, P. Biondi, and D. Kaminsky, Honeypot forensics part I: Analyzing the network, IEEE Security & Privacy, vol. 2, no. 4, pp. 72-78, 2004.
18. F. Raynal, Y. Berthier, P. Biondi, and D. Kaminsky, Honeypot forensics part II: Analyzing the compromised host, IEEE Security & Privacy, vol. 2, no. 5, pp. 77-80, 2004.F. Deng, A. Luo, Y. Zhang, Z. Chen, X. Peng, X.