ISSN 2348 – 7968

Detecting an Unknown Attacks in Big Data Analysis System

Miss. Bharati Bagde, Prof. Parul Jha

[1] WCC, Nagpur/TGPCET,
Nagpur, Maharashtra 440014/Indora, India

[2] IT, Nagpur/TGPCET/,
Nagpur, Maharashtra 441108/Mohgao, India

Abstract

Recently, threat of previously unknown cyber-attacks are increasing because existing security systems are uanble to identify them. Previous cyber-attacks had simple occasions of leaking personal information by attacking the PC or to reduce the system. However, the goal of recent heavy blows. attacks has changed from leaking information and destruction of services to attacking large-scale systems such as critical infrastructures and state agencies. Previous defense technologies to counter these attacks are based on patterns matching methods which are very limited. Because of this reality, In the event of new and previously unknown attack, detections rate become very low and false negative increases. To keep safe from these unknown attacks, which unable to detected with existing technology, We proposed new model based on big data analysis techniques that can extract information from a variety of sources to detect future attacks. Regarding the model on the basis of the future Advanced Persistent Threat (APT) detection and prevention system implementations.

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

1. Introduction

Hacking in the databases and can process the big volume, big velocity and big 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 the document 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.

2. 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.

Large Data Analysis with Hadoop to inspect 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 Outlier 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 outlier detection algorithm shave been designed.for l data. There are two main problems of outlier detection for unconditional data, which are the same measure between unambiguously explicit and direct data objects and the detection efficiency . outlier detection algorithm for unambiguously explicit and direct data Efficient outlier 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 rules are enforced by collaborative UTM and the feedback events of such rules are returned to the security center.

2.1 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.

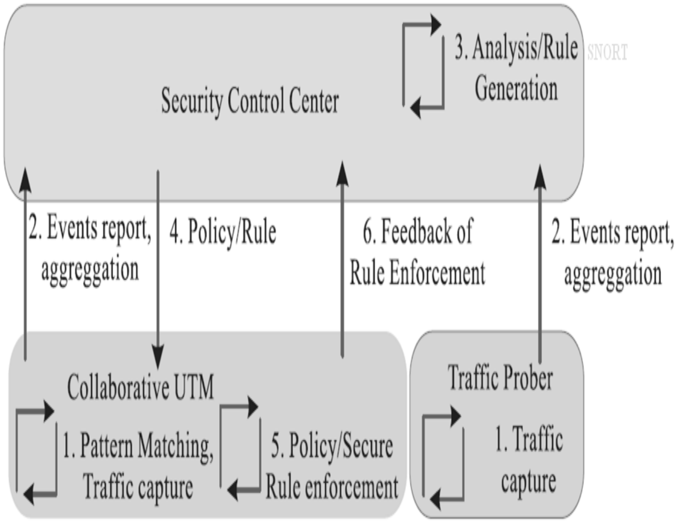
Security Control Center
3. Analysis/Rule Generation
SNORT
2. Events report, aggreggation
4. Policy/Rule
6. Feedback of Rule Enforcement
2. Events report, aggreggation
Collaborative UTM
1. Pattern Matching, Traffic capture
5. Policy/Secure Rule enforcement
Traffic Prober
1. Traffic capture

Fig 1. **System Flow diagram**

3. 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.

4. Conclusions

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.

Acknowledgments

This work was supported by Prop Parul bhanakar Jha Head, Department of Information Technology**,** T.G.P.C.E.T, Nagpur. Prof.Jayant Rohankar, Assistant Professor, Information Technology, T.G.P.C.E.T, Nagpur. and for being a constant source of inspiration. The authors would like to thank the anonymous reviewers for their valuable and constructive comments on improving the paper.

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 cyber attacks,” Gover n-ment Executive, 4 December 2007, www.govexec.com/defense/2007/12/us-british officials target chinese as source of cyber attacks/25874/

[4]. Clayton, Mark; “US Oil Industry Hit by Cyberattacks: Was China Involved?,” The Chritian Science Monitor, 25 January2010, 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/new

[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.