MALWARE DETECTION WITH ADAPTIVE ANOMALY DETECTION

Prof. P. Arunadevi
Assistant Professor

Ms. Anjali Vijayakumar
UG Scholar

Mr. K. M. Jiju
UG Scholar

Ms. K. Nithya
UG Scholar

Department of Computer Science and Engineering,
RVS College of Engineering and Technology,
Coimbatore, Tamilnadu, India

Abstract—Cloud computing is a growing industry which provide instant access to data anytime and anywhere. The downside is there is no standardization of security features and entrusting our data with third parties. The hackers find way with security holes to steal and disrupt easily. Malware is of high importance to be detected. The adaptive anomaly algorithm with clustering and classification overcome shortcomings in previously devised algorithms.

Keywords—Cloud Computing; Malware; Anomaly Detection

I. INTRODUCTION

Cloud computing is the sharing of computer processing resources and data to computers and other devices instant access over the internet. It is of pay for use basis, agility, high productivity and elasticity. It has three service models: Infrastructure as a service(IaaS), Software as a Servics(SaaS), Platform as a Service(PaaS).According to NIST, Infrastructure as a service provisions processing, networks, storage and other fundamental computing resources where the consumer is able to deploy nd run arbitrary software, which can include operating systems and applications. The consumer does not manage the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components. Software as a service [1][2] able consumer to use provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface such as a web browser. In platform as a service, the consumer can deploy onto the cloud their own applications created using programming languages, libraries, services, and tools supported by provider.

The data centers are used for a range of always-on services across private, public and commercial domains. These need to be secure and flexible in the face of challenges that include cyber attacks as well as component failures. However, clouds have characteristics and internal operational structures that impair the use of traditional detection systems. In particular, the range of advantageous properties offered by the cloud, such as service opacity and elasticity, introduce a number of vulnerabilities which are the outcome of its underlying virtualized nature. Moreover, an indirect problem lies with the cloud's external dependency on IP networks, where their resilience and security has been extensively studied, but nevertheless remains an issue [3][4]. The infrastructure level contain elements cloud nodes, which are hardware servers that run a hypervisor in order to host a number of Virtual Machines and network infrastructure elements that provide the connectivity within the cloud and connectivity to external service users. The infrastructure as a service allows customers to install and administer their own choice of OS. Focusing on this type of system, we develop and test adaptive anomaly detection algorithm which uses clustering and classification to detect malware with prior history and those without.

II. PROBLEM DEFINITION

The Novelty Detection Approach employs one class support vector machine algorithm which helps to identify DoS(Denial of Service) attacks with high detection accuracy but low when detecting malware on its own.

The novelty detection approach use decision function which returns a classifier to signify malware detection. If there is no malware, the classifier is a true negative(1). If there is, it returns false negative(-1).

III. LITERATURE SREVEY

3.1 Level Network Resilience

Traffic analysis and anomaly detection have been widely used to characterize network utilization as well as to identify abnormal network traffic such as malicious attacks. However, so far, techniques for traffic analysis and anomaly detection have been passed out independently, relying on mechanisms and algorithms either in edge or in core networks alone. In this paper we suggest the notion of multi-level network resilience, in order to provide a more strong traffic analysis and anomaly detection architecture, combining mechanisms and algorithms operating in a coordinated fashion both in the edge and in the core networks. This work is provoked by the potential complementarities between the research being developed at IIT Madras and Lancaster University. In this paper we describe the existing [5] work being developed at IIT Madras and Lancaster on traffic analysis and anomaly detection, and outline the principles of multilevel resilience architecture.

3.2 Malware analysis in cloud computing

The consumption of cloud computing environments is increasingly common, and we are implicitly dependent on them for many services. However, their dependence on virtualized computer and network infrastructures introduces risks related to system flexibility. In particular, the virtualized natural history of the cloud has not yet been methodically studied with respect to security issues as well as vulnerabilities and appropriate anomaly detection. This paper proposes an approach for the [6] examination and analysis of malware in virtualized environments. We carry out an study, on a system and network-wide scale, and additional pinpoint some system and network features specifically by study the model of the Kelihos malware.

3.3 Flexibility and survivability in communication networks

The Internet has become necessary to all aspects of modern life, and thus the penalty of network disruption has become increasingly severe. It is broadly recognized that the Internet is not suitably resilient and dependable, and that significant research, development, and engineering is necessary to improve the situation. This paper provides an architectural framework for flexibility and survivability in communication networks and provides a survey of the disciplines that resilience encompasses, along with considerable past failures of the network infrastructure. A resilience approach is presented to detect, defend against, and remediate challenges, a rest of principles for designing flexible networks is presented, and techniques are described to analyze network flexibility

3.4 Towards a Distributed, Self-organising Approach to Malware Detection in Cloud Computing

Cloud computing is an increasingly fashionable platform for both industry and consumers. The cloud presents a number of single security issues, such as a high level of sharing and system homogeneity, which need special consideration. In this paper we introduce a flexibility architecture consisting of a collection of self-organising flexibility managers distributed within the infrastructure of a cloud. More specifically we exemplify the applicability of our proposed architecture under the situation of malware detection. We describe our multi-layered solution at the hypervisor level of the cloud nodes and judge how malware detection can be distributed to each node.

3.5 On the Analysis of the Zeus Botnet Crimeware Toolkit

In this paper, we present our reverse engineering consequences for the Zeus crimeware toolkit which is one of the current and powerful crimeware tools that emerged in the Internet alternative community to control botnets. Zeus has reportedly polluted over 3.6 million computers in the United States. Our study aims at uncovering the various obfuscation levels and detaching the light on the resulting code. Accordingly, we clarify the bot building and installation/infection processes. In addition, we detail a technique to extract the encryption key from the malware dual and use that to decrypt the network communications and the botnet configuration information. The reverse engineering insights, jointly with network traffic analysis, allocate for a better accepting of the technologies and behaviors of such modern HTTP botnet crimeware toolkits and opens an opportunity to insert falsified information into the botnet communications which can be used to insult this crimeware toolkit.

3.6 Data security in the world of cloud computing

Today, we have the skill to utilize scalable, distributed computing environments within the limits of the Internet, a practice known as cloud computing. In this new world of computing, users are generally required to believe the underlying premise of trust. Within the cloud computing world, the virtual surroundings lets users access computing power that exceeds that contained within their own physical worlds. Typically, users will recognize neither the correct location of their data nor the other sources of the data jointly stored with theirs. The data you can find in a cloud ranges from free source, which has smallest security concerns, to private data containing extremely sensitive information. Does using a cloud environment improve the business entities of their responsibility to ensure that proper security dealings are in place for both their data and applications, or do they split joint responsibility with service providers? The answers to this and other questions lie inside the area of yet-to-be-written law. As with most technological advances, regulators are naturally in a "catch-up" mode to recognize policy, governance, and law. Cloud computing presents an addition of problems heretofore experienced with the Internet. To guarantee that such decisions are informed and suitable for the cloud computing environment, the business itself should establish coherent and effective policy and domination to identify and implement appropriate security methods.

IV. PROPOSED SYSTEM

4.1 Methodology

The system architecture shows how a user and service provider interact in a real environment. Virtualization of a computer is a great way of accessing resources without the need for upfront investment in physical resources. The virtualized infrastructure is accessed from anywhere via a web browser. KVM is a Linux kernel module and relies on other parts of Linux kernel for managing the guest systems. The work is done on KVM hypervisor. The average organization has over 540 unique user enabled third party, OAuth–connected cloud applications active within their environment. When user grants sccess to an application with their Google credentials that allows the application to access their corporate Google account. Oftentimes these apps have excessive access scopes, including permissions to view, edit, delete data. There is also potential to open the door to potentially malicious apps, providing hackers with unfettered access to user's environment. We tend to use adaptive anomaly detection using classification and clustering along with one class support vector machine algorithm to detect malware better. Classification is the problem of identifying to which of set of categories a new observation belongs, on basis of a training set of data containing instances whose category membership is known. An example would be assigning spam classes or diagnosis to a given patient as described by observed characteristics of the patient.

An algorithm which implements classification, especially in a concrete implementation is known as a Classifier. The decision function returns a classifier when it doesn't detect malware, it returns true negative (1). And when it detects, it returns false negative (-1).

The decision function is

$$f(x)=\Sigma\alpha_1 k(x_1x_i)-\rho$$

where x is a feature vector which contains all the features of virtual machine and k() is a kernel function and, α_1 is a multiplier. In this algorithm, the second part is clustering, which is assumed to be done in subspace models. It allows clustering of rows and coluns of a matrix.

Given a set of
{\displaystyle m}
rows in
{\displaystyle n}
columns(i.e., an
{\displaystyle m\times n}
matrix)
the biclustering algorithm generates biclustera - a subset of rows which exhibit similar behaviour across a subset of columns or vice versa.

Account Authentication

The consumers must create account to access the database if a new user. If username and password match, they can login to the service.

Service Activation

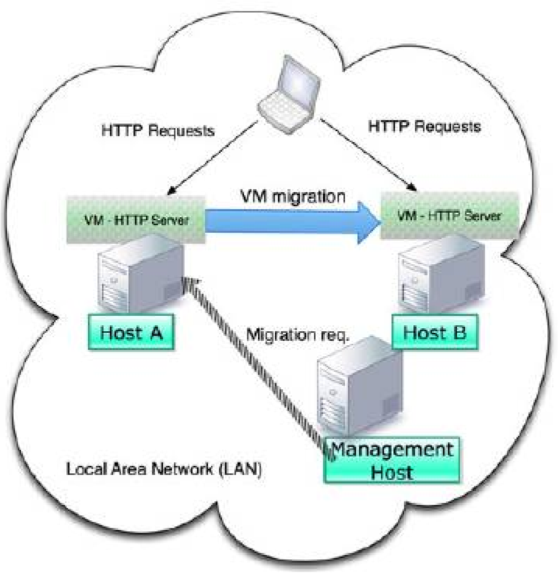
The service is registered after proper authentication. The service is activated with submission of required information. The payment details among other is displayed

Malware Detection

The malware is detected by combining novelty detection and, classification and clustering based adaptive anomaly detection approach.

V. EXPERIMENT & RESULT

The clean Virtual Machine (VM) is created from a known to be clean disk. The VM is monitored for a period of 10 minutes in what we refer to as the normal phase. The malware used, Zeus sample which runs for 15 minute is more than enough to gauge algorithm efficiency. Malware is injected and a further 10 minutes of monitoring follows. This is anomalous phase. The output is a vector *y* with an *n* dimension equal to the *m* dimension of the input matrix, which gives a single value of *y* that lies between {-1,1} for each snapshot vector *x*. The detection performance of a classifier can be assesses by determining the difference between class it produces for a given input and class it should produce. If there is no malware, the classifier is a true negative(1). If there is, it returns false negative(-1).



		Actual Class	
		1	−1
Predicted Class	1	TN	FN
	−1	FP	TP

Fig 1 : Proposed System Architecture

VI. CONCLUSION

The Novelty Detection Approach employs one class support vector machine algorithm which helps to identify DoS(Denial of Service) attacks with high detection accuracy but low when detecting malware on its own. We propose to combine this algorithm with clustering and classification based adaptive anomaly detection which detect failures at runtime without prior history and can learn from failure events at runtime under less computational cost. It gives better accuracy and resilience.

References

[1] N. Gruschka and M. Jensen, “Attack surfaces: A taxonomy for attacks on cloud services,” in Proc. IEEE 3rd Int. Conf. Cloud Comput., Jul. 2010, pp. 276–279.

[2] Y. Chen, V. Paxson, and R. H. Katz. (2010, Jan.). Whats new about cloud computing security?. EECS Department, Univ. of California. Berkeley, Tech. Rep. UCB/EECS-2010-5. [Online]. Available:http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-5.html

[3] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee,“Bothunter: Detecting malware infection through ids-driven dialog correlation,” in Proc. 16th USENIX Security Symp. USENIX Security Symp., Berkeley, CA, USA, 2007, pp. 12:1–12:16.

[4] M. Bailey, J. Oberheide, J. Andersen, Z. Mao, F. Jahanian, and J. Nazario, “Automated classification and analysis of internet malware,” in Proc. 10th Int. Conf. Recent Adv. Intrusion Detection, 2007, vol. 4637, pp. 178–197.

[5] C. Mazzariello, R. Bifulco, and R. Canonico, “Integrating a network ids into an open source cloud computing environment,” in Proc. 6th Int. Conf. Inf. Assurance Security, Aug. 2010, pp. 265–270.

[6] S. Roschke, F. Cheng, and C. Meinel, “Intrusion detection in the cloud,” in Proc. 8th IEEE Int. Conf. Dependable, Autonomic Secure Comput., Dec. 2009, pp. 729–734.