ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Increasing Anomaly Detection Accuracy in Cloud Networks

K.V.Aditya, K.Kiran Kumar, K.Narayana Reddy

Assistant Professor, Dept. of MCA, Narayana Engineering College, Nellore, AP, India

MCA Student, Narayana Engineering College, Nellore, AP, India

MCA Student, Narayana Engineering College, Nellore, AP, India

ABSTRACT: Cloud services are prominent within the private, public and commercial domains. Many of these services are expected to be always on and have a critical nature; therefore, security and resilience are increasingly important aspects. In order to remain resilient, a cloud needs to possess the ability to react not only to known threats, but also to new challenges that target cloud infrastructures. In this paper we introduce and discuss an online cloud anomaly detection approach, comprising dedicated detection components of our cloud resilience architecture. More specifically, we exhibit the applicability of novelty detection under the one-class support Vector Machine(SVM) formulation at the hypervisor level, through the utilization of features gathered at the system and network levels of a cloud node. We demonstrate that our scheme can reach a high detection accuracy of over 90 percent whilst detecting various types of malware and DoS attacks. Furthermore, we evaluate the merits of considering not only system-level data, but also network-level data depending on the attack type. Finally, the paper shows that our approach to detection using dedicated monitoring components per VM is particularly applicable to cloud scenarios and leads to a flexible detection system capable of detecting new malware strains with no prior knowledge of their functionality or their underlying instructions.

KEYWORDS: Security, invasive software, network-level security and protection

I. INTRODUCTION

CLOUD datacenters are beginning to be used for a range of always-on services across private, public and commercial domains. These need to be secure and resilient in the face of challenges that include cyber attacks as well as component failures and mis-configurations. However, clouds have characteristics and intrinsic internal operational structures that impair the use of traditional detection systems. In particular, the range of beneficial properties offered by the cloud, such as service transparency 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 [1].The approach taken in this paper relies on the principles and guidelines provided by an existing resilience framework [2]. The underlying assumption is that in the near future, cloud infrastructures will be increasingly subjected to novel attacks and other anomalies, for which conventional signature based detection systems will be insufficiently equipped and therefore ineffective. Moreover, the majority of current signature-based schemes employ resource intensive deep packet inspection (DPI) that relies heavily on payload information where in many cases this payload can be encrypted, thus extra decryption cost is incurred. Our proposed scheme goes beyond these limitations since its operation does not depend on a-priori attack signatures and it does not consider payload information, but rather depends on per-flow meta-statistics as derived from packet header and volumetric information (i.e., counts of packets, bytes, etc.). Nonetheless, we argue that our scheme can synergistically operate with signature-based approaches on an online basis in scenarios were decryption is feasible and cost-effective. Overall, it is our goal to develop detection techniques that are specifically targeted at the cloud and integrate with the infrastructure itself in order to, not only detect, but also provide resilience through remediation.

The elements presented here form the basis in which different detection techniques can be hosted and further allow

DOI: 10.15680/IJIRCCE.2017. 0504036

the identification and attribution of anomalies. In this paper we discuss the detection of anomalies using a novelty detection approach that employs the one-class Support Vector Machine (SVM) algorithm and demonstrate the effectiveness of detection under different anomaly types. More specifically, we evaluate our approach using malware and

- Experiments carried out in this work are done so in the context of an overall cloud resilience architecture

under the implementation of one-class Support Vector Machines (SVMs). The resulting experimental findings show that anomalies can be effectively detected online, with minimal time cost for reasonably realistic data samples per Virtual Machine (VM), using the one-class SVM approach, with an overall accuracy of greater than 90 percent in most cases.

- Our work is the first to explicitly address the a spectra malware detection in pragmatic cloud-oriented scenarios as performed by cloud providers, such as VM live-migration.
- We provide an online novelty detection implementation that allows the adaptive SVM-specific parameter estimation for providing better detection accuracy benefits.
- This work assesses the VM-based feature selection spectrum (i.e., system, network-based or joint datasets)with respect to the detection performance benefit son two distinct network-wise attacks (malware and DOS) under novelty detection.

II. BACKGROUND AND RELATED WORK

The intrinsic properties of virtualized infrastructures (such as elasticity, dynamic resource allocation, service co-hosting and migration) make clouds attractive as service platforms. Though, at the same time they create a new set of security challenges. These have to be understood in order to better protect such systems and make them more secure. A number of studies have addressed aspects of cloud security from different viewpoints (e.g. the network, hypervisor, guest VM and Operating System) under various approaches derived either from traditional rule-based Intrusion Detection Systems (IDSs) or statistical anomaly detection models. This paper presents a cloud security solution derived from a sub-domain of anomaly detection, viz. novelty detection.

2.1 Virtualization& Cloud Technologies

In [3], [8], [9] the specific security threats and challenges introduced into clouds through the use of core virtualization technologies are discussed. Despite the end-user benefit gained by virtualization it also comes with a range of threats that include: exploits to security holes on virtual machines(e.g. root kit attacks on virtual machines [10]); mutated cloud-specific Internet-based attacks that aim to compromise cloud networks (e.g. malware [3], [11]; and DDOS attacks on cloud services [11]). According to [12] black hat hackers have already identified the potential of the cloud since the instantiation, maintenance and continued operation of bonnets seems to be much more effective under a cloud paradigm. In parallel, co-residence as a security concern has been explored in [10] and is the result of VMs belonging to different customers being hosted on the same cloud node. It was revealed that the outcome of co-residence is to enable shared memory attacks that, at their most benign, are capable of leaking sensitive information, and at their most destructive are capable of taking control of the entire node. Moreover, the aspect of VM migration is also a possible enabler of malicious side effects in cases where infected VMs are migrated around the cloud to a number of nodes. The cause of migration could be as a result of the provider's load balancing policy, but as an unwanted side-effect the result is to place malware in contact with a larger number of potential targets throughout the cloud infrastructure. Additionally, automation is becoming an increasingly integral part of computer system configuration through the

2.2 Malware & Detection Methods

One of the biggest challenges within the development of resilient and secure cloud-oriented mechanisms is related to the adequate identification and detection of malware. This is due to the fact that, in the majority of cases, malware is the first point of initiation for large-scale Distributed Denial of Service (DoS) attacks, phishing and email spamming [3], [8], mainly through the deployment of beware. Current methods of detecting attacks on cloud infrastructures or the VMs resident within them do not sufficiently address cloud specific issues. Despite the huge efforts employed in past studies regarding the behavior of certain types of malware in the Internet [13], [14], so far little has been done to tackle malware presence in clouds. In particular, the studies in [15], [16] aimed to adjust the performance of traditional

Intrusion Detection Systems under signature based techniques that employ Deep Packet Inspection on network packets. Moreover, work in [17], [18] studied system- related features on monitored VMs by employing Virtual Machine Introspection (VMI) methods in order to detect threats on a given VM's Operating System. Nevertheless, despite the important lessons learned from these studies they do not develop an overall online detection strategy that considers real-time measurement samples from each VM. Further, these approaches are purely signature-based, and as such are not in a position to provide a robust scheme for any future threats posed by novel malware strains due to their simplistic rule-based nature. Each solution to detection is performed in an isolated manner and neglects to consider the unique topology of the cloud, which is at its heart a network of interconnected nodes, each with their own isolated execution environments. If a detection system is to perform effectively within a cloud it is required to possess the capability of communicating detected faults and challenges across the whole infrastructure, especially if it is to perform as part of a larger, autonomous and self-organizing, cloud resilience system.

2.3 Anomaly Detection in Clouds

In [20] an anomaly detection technique to detect intrusions at different layers of the cloud was proposed. However, the technique appears to lack the flexibility required by dynamic cloud environments. It is also not sufficiently demonstrated how such techniques can be operationally applied. In [21] the authors propose a multi-level approach, which provides fast detection of anomalies discovered in the system logs of each guest OS. One of its disadvantageous is the apparent lack of scalability since it requires increasingly more resources under high system workload. Further, it is designed to classify text based log data, which may not manifest the effects of malware. The work in [19] provided a novel prototype that enabled an online spatio-temporal anomaly detection scheme in a cloud scenario. Thus, the authors were able to initially formulate and further implement a wavelet-based multi-scale anomaly detection system. The system relies on measured cloud performance metrics (e.g. CPU utilization, memory) gathered by multiple components (e.g. hardware, software, system) within the examined institution-wide cloud environment. The resulting experimental outcomes were quite promising since the proposed approach reached a 93:3 percent of sensitivity on detecting anomalous events with only just a 6:1 percent of the reported events to be false alarms. The only study that has some similarities to what we propose in this paper is the approach by Pannu et al. in [22]. In particular, the authors in [22] instrumented an online adaptive anomaly detection (AAD) framework that was able to detect failures through the analysis of execution and runtime metrics using the traditional two-class Support Vector Machine algorithm. Under a real experimentation, over a 362-node cloud computing environment in a university campus, the produced results were extremely promising since they exhibited the efficiency of the proposed scheme, which reached an overall of over 87 percent of anomaly detection sensitivity. However, the main issue raised by this study was that the formulation of the two-class SVM algorithm suffered from the data imbalance problem [23], which affected the training phase, and consequently led to several mis-classifications of newly tested anomalies.

III. PROPOSED METHODOLOGY

The cloud test bed used in this work is based on KVM hypervisors under Linux (which in turn use Qemu for hardware emulation). The test bed comprises two compute nodes, one of which also acts as the storage server for VM images, and a separate controller server. The management software is Virtual Machine Manager (sometimes referred to as virt-manager), which interfaces with libvirt daemons on the compute nodes. Cloud orchestration software (such as Open Stack) is not deemed necessary for our particular experiments since we are concerned solely with direct data acquisition from VMs and not the interaction of the detection system with management software. However, the tools used in this work are compatible with any cloud orchestration software that uses either Xen or KVM as a hypervisor and the approach we take here could therefore be applied to such an environment. In general, our test bed is capable of many of the functions associated with cloud computing such as flexible provisioning of VMs, cloning and snapshot ting VM images, and offline and online7 migration.

3.1 Data Collection & Feature Extraction

The data collection and analysis tools installed on each compute node in the described test bed include libVMI8 and Volatility9 for real-time Virtual Machine Introspection, tcpdump10 and CAIDA's CoralReef11 for packet capturing and network flow summarization. Overall, the data acquisition, feature extraction and anomaly detection performed by both the SAE and NAE components of our resilience architecture are achieved through custom software that operates on VMs in real-time at the hypervisor level of the cloud node.

At the network level the NAE gathers data through tcpdump, which separates packets into 8 second time bins. Features are then extracted using the CAIDA Coral Reef suite of tools, which provides the capability to generate statistics.

3.2 One-Class SVM

The core of our online detection methodology within the SAE and NAE lies with the implementation of the supervised one-class SVM algorithm, which is an extension of traditional two-class SVM, and was proposed by Scholkoph et al. in [35]. In practice, the one-class SVM formulation handles cases using unlabelled data (i.e., novelty detection), the main goal of which is to produce a decision function that is able to return a class vector y given an input matrix x based on the distribution of a training dataset. The class y is a binary class where one outcome is the known class, which in our case is the normal VM behavior, and the other is the novel class, which represents any testing instances that are unknown to the classifier. If we let x=(x1,x2,.....,xn) represent a feature vector, which contains all of the VM-related features described earlier.

3.3 SAE & NAE One-Class SVM Tuning

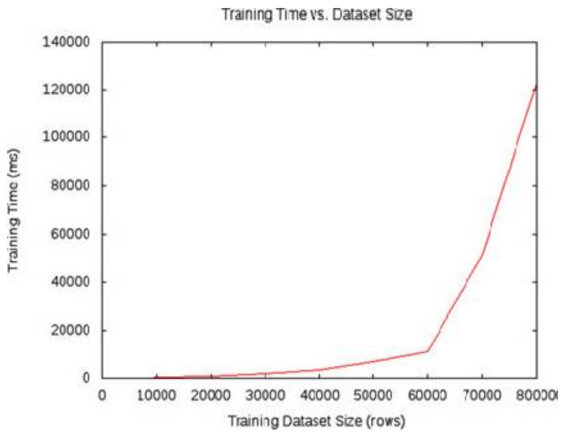
Prior to the training process, the SAE & NAE engines automatically transform the initial gathered dataset by scaling them towards a Gaussian distribution. This is due to a requirement of the RBF kernel that the data be centered on zero and have unit variance. Thus the tuning process embedded in the SAE and NAE removes the mean from each feature and divides the feature vector by the standard deviation. The training process subsequently involves passing the scaled training dataset as an input to the one-class SVM algorithm, which produces a decision function that is able to classify new feature vectors. In general, the training process is determined by four factors: the size and content of the training dataset and the two parameters n and g. The training dataset size is determined by the length of time over which VM monitoring is conducted, after which it is possible to select subsets of the available data resulting in a refinement of training data and a reduction in dataset size if required. Dataset content is determined by the behavior of the processes in the VM and is not accurately controllable, hence the only influence that can be imposed on the data is by varying the applications and the loads on each of them.

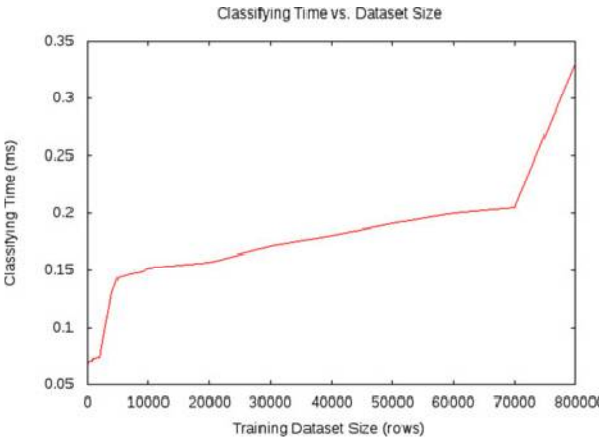
3.4 Classification Performance Metrics

The detection performance of the classifier can be assessed by determining the difference between the class it produces for a given input and the class it should produce. For example, if a sample of data contains no anomalies due to a malware strain, and the classifier produces an output of 1 for that data point, it is a correct classification. In order to quantify the classification performance we consult a confusion matrix that describes all possible outcomes of a prediction and has the form: In our experiments a “positive” outcome is one in which the detector detects an anomaly, i.e., produces a class of _1.From this we can conclude that a True Positive (TP) is possible when the classifier produces a _1 during malware execution, otherwise it is treated as a False Positive (FP).

IV. EXPERIMENTAL RESULTS

The experiments we present in this section test the detection aspects of the System and Network Analysis Engines. Given the fact that both engines perform online anomaly detection under the one-class SVM formulation we initially present our results related to the computational cost of the online training and testing of the algorithm, since





they affects the overall response of the real-time detection process. We subsequently present our assessment on detecting the Kelihos and Zeus malware strains as well as the DDoS attacks. In addition, we further present a comparison between the detection accuracy obtained when using a joint dataset (i.e. composed of both system and network features) with a feature set that strictly considers network-based features. The experiments that focus on the SAE functionality involve the detection of Kelihos and Zeus under static analysis and live-migration using a 12 dimensional system-level dataset. NAE performance is tested under static analysis against DoS using a nine dimensional network-level dataset and against Zeus using the nine dimensional network dataset and a 21 dimensional joint-level dataset (i.e., system and network).

V. CONCLUSION

Overall, this work performs online anomaly detection under two pragmatic cloud scenarios, based on suggestions by cloud operators, which emulate “static” detection as well as detection under the scenario of VM “live” migration. The results obtained by strictly utilizing system-level data in our SAE detection, which was supported by an automatic SVM specific parameter selection process, have shown excellent detection for all samples of malware under a variety of conditions (i.e., static and migration analysis) with an overall detection accuracy rate of well above 90 percent. Hence, we have demonstrated that the extracted features for classifier training were appropriate for our purposes and aided towards the detection of the investigated anomalies under minimal time cost throughout the training and testing phase. Nonetheless, in order to further the investigation, this feature set can easily be expanded to include statistics derived from CPU usage and a deeper introspection of process handles, which could be beneficial for the detection of highly stealthy malware. However, the consideration of new features would naturally invoke a computational trade-off, since deeper introspection will require more system resources.

VI. ACKNOWLEDGEMENTS

This work is sponsored by UK-EPSRC funded EPSRC IUATC project, grant agreement no. EP/J016675/1 in collaboration with British Telecommunications (BT); and the EU FP7 Project SECCRIT (Secure Cloud Computing for Critical Infrastructure IT), grant agreement no. 312758.

REFERENCES

[1] A. Marnerides, C. James, A. Schaeffer, S. Sait, A. Mauthe, and H. Murthy, “Multi-level network resilience: Traffic analysis, anomaly detection and simulation,” ICTACT J. Commun. Technol.,Special Issue Next Generation Wireless Netw. Appl., vol. 2, pp. 345– 356, Jun. 2011.
[2] J. P. G. Sterbenz, D. Hutchison, E. K. C¸ etinkaya, A. Jabbar, J. P. Rohrer,M. Sch€oller, and P. Smith, “Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines,” Comput.Netw., vol. 54, no. 8, pp. 1245–1265, Jun. 2010.
[3] A. K. Marnerides, M. R. Watson, N. Shirazi, A. Mauthe, and D. Hutchison, “Malware analysis in cloud computing: Network and system characteristics,” in Proc. IEEE Globecom Workshop, 2013, pp. 482–487.

[4] M. R. Watson, N. Shirazi, A. K. Marnerides, A. Mauthe, and D. Hutchison, “Towards a distributed, self-organizing approach to malware detection in cloud computing,” in Proc. 7th IFIP/IFISCIWSOS, 2013, pp. 182–185.
[5] M. Garnaeva. Kelihos/Hlux Botnet Returns with New Techniques. Securelist [Online]. Available: http://www.securelist.com/en/blog/655/Kelihos_Hlux_botnet_returns_with_new_techniques, Feb. 2012.
[6] H. Binsalleeh, T. Ormerod, A. Boukhtouta, P. Sinha, A. Youssef, M. Debbabi, and L. Wang, “On the analysis of the zeus botnet crimeware toolkit,” in Proc. 8th Annu. Int. Conf. Privacy Security Trust, Aug. 2010, pp. 31–38.
[7] T. Brewster. (2014, Jul. 11). GameOver Zeus returns: Thieving malware rises a month after police actions, Guardian Newspaper [Online]. Available: http://www.theguardian.com/technology/2014/jul/11/gameover-zeus-crimina l-malware-police-hacking
[8] A. K. Marnerides, P. Spachos, P. Chatzimisios, and A. Mauthe,“Malware detection in the cloud under ensemble empirical modeldecomposition,” in Proc. 6th IEEE Int. Conf. Netw. Comput., 2015,pp. 82–88.
[9] L. Kaufman, “Data security in the world of cloud computing,”IEEE Security Privacy, vol. 7, no. 4, pp. 61–64, Jul. 2009.
[10] M. Christodorescu, R. Sailer, D. L. Schales, D. Sgandurra, andD. Zamboni, “Cloud security is not (just) virtualization security:A short paper,” in Proc. ACM Workshop Cloud Comput. Security,New York, NY, USA, 2009, pp. 97–102.
[11] N. Gruschka and M. Jensen, “Attack surfaces: A taxonomy forattacks on cloud services,” in Proc. IEEE 3rd Int. Conf. Cloud Comput., Jul. 2010, pp. 276–279.

BIOGRAPHY

Aditya K.V, is a Assistant Professor in the Department of MCA, Narayana Engineering College, Nellore. His research interest in Increasing Anomaly Detection Accuracy In Cloud Networks.

KiranKumar K, MCA student in Narayana Engineering College, Nellore. His research interest in Increasing Anomaly Detection Accuracy In Cloud Networks.

NarayanaReddy K, MCA student in Narayana Engineering College, Nellore. His research interest in Increasing Anomaly Detection Accuracy In Cloud Networks.