Towards Benchmarking Intrusion Detection Systems for Virtualized Cloud Environments

Aleksandar Milenkoski
Karlsruhe Institute for Technology
Karlsruhe, Germany
Email: milenkoski@kit.edu

Samuel Kounev
Karlsruhe Institute for Technology
Karlsruhe, Germany
Email: kounev@kit.edu

***Abstract*—Many recent research works propose novel architectures of intrusion detection systems specifically designed to operate in virtualized environments. However, little attention has been given to the evaluation and benchmarking of such architectures with respect to their performance and dependability. In this paper, we present a research roadmap towards developing a framework for benchmarking intrusion detection systems for cloud environments in a scientifically rigorous and a representative manner.**

***Index Terms*—intrusion detection; benchmark testing;**

I. Introduction

The cloud computing paradigm, with virtualization as key enabling technology, is constantly gaining in popularity. However, the wide migration to cloud systems is challenged by security concerns. A common defensive instrument against security threats are intrusion detection systems (IDSes). The IDSes for cloud platforms are usually deployed in the virtualization layer, i.e., in a Virtual Machine Monitor (VMM). We refer to such IDSes as *VMM-based IDSes*. To minimize the risk of security breaches, reliable methods and techniques for evaluating the performance of IDSes are needed. Lack of in-depth IDS evaluations can lead to deployment of an IDS which does not operate optimally in a given environment. To the best of our knowledge, no benchmarking framework specifically targeted at VMM-based IDSes currently exists. We argue that the existing IDS benchmarking solutions do not completely satisfy the requirements for benchmarking such IDSes. Some of these requirements are representative cloud benign and malicious workloads, appropriate metrics and benchmarking methodology. In this paper, we briefly outline our vision towards addressing these issues by designing and developing a benchmark framework for VMM-based IDSes. We build upon existing knowledge on benchmarking IDSes for traditional non-virtualized environments, focusing on the unique issues and challenges that arise in the context of VMM-based IDSes.

II. Benchmarking Approach

We aim to support benchmarking a VMM-based IDS in its target production environment during operation. Our goal is to create a customizable IDS benchmarking framework that delivers representative cloud benign and malicious workloads, novel metrics and scientifically rigorous methodology.

Workloads. Since VMM-based IDSes normally reside in a VMM, they have the opportunity to monitor the network and the host activities of all guest virtual machines (VMs) at the same time. Thus, many VMM-based IDSes possess both network- and host-based intrusion detection sensors [1]. Hence, to benchmark the attack detection accuracy of a typical VMM-based IDS, one needs malicious workloads for both *host and network* intrusion detection sensors. Further, many VMM-based IDSes are designed to correlate sequential, i.e., multi-step, attacks since they are typical for virtualized environments (e.g., attack on operating system (OS) and on VMM in that order) [1], [2]. That raises the need for malicious workloads which contain attacks in a specific *temporal order*. Our approach is to generate attack traces and potentially to acquire real-world traces of attacks targeted at virtualized infrastructures. Attack traces offer controlled and repeatable replay of recorded attack sessions. We intend to address some of the issues related to acquiring real-world traces (e.g., privacy issues) by leveraging close contacts with industrial collaborators. The authors are members of the Cloud Research Working Group of SPEC (Standard Performance Evaluation Corporation)[3]. The members of this group include many established academic and industrial institutions. As a group activity, we are performing a feasibility study for acquiring and sharing traces from cloud providers which may contain attacks. This includes evaluation of approaches to solve legal issues, determination of the type and the scope of shared data and so on. We also intend to generate host and network attack traces, with or without background benign activity, in a testbed environment at Karlsruhe Institute of Technology (KIT). We are currently in the process of setting up a virtualized environment consisting of multiple VMMs in which we intend to perform sequential attacks against representative targets towards a predefined final goal. As representative targets we consider vulnerable VMMs and applications/services commonly deployed in cloud systems for which publicly available attack scripts exist (e.g., Xen, CVE-2012-0217; Remote Desktop Protocol, CVE-2012-0002).

We intend to provide benign workloads as mixed with the attacks or as isolated (i.e., pure). For instance, pure benign workloads can be used for measuring the monitoring performance overhead of an IDS. We argue that a representative cloud benign workload should possess two characteristics: *heterogeneity*, i.e., to be a mix of cloud representative workload types (e.g., streaming, data processing) [4] and *scalability*. The number of guest VMs hosted on a VMM can vary as new VMs

978-1-908320-08/7/$25.00©2012 IEEE

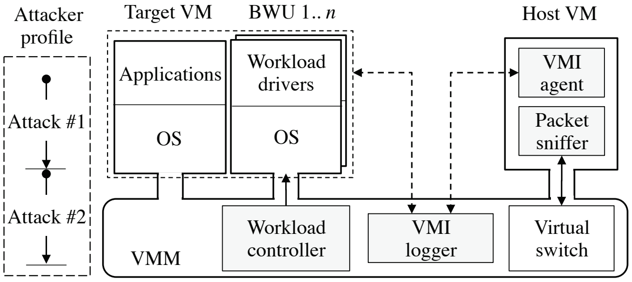

Figure 1. Architecture of an environment for attack trace generation (single VMM)

are added or existing ones are removed. Each VM is likely to execute workloads with various characteristics which can affect the performance of a VMM-based IDS due to shared hardware resources. To achieve scalability and heterogeneity, we are designing *workload controller*, i.e., a VMM module which uses the VMM functionalities to manage the activity of Benign Workload Units (*BWUs*), i.e., virtual machine images with deployed realistic mix of automated workload drivers. The task of the workload controller is to activate or deactivate BWUs at a specific time as defined in a benign workload scenario specification. We are developing workload controllers and BWUs for the VMMs which are native environment for most existing VMM-based IDSes, e.g, Xen. Besides for generation of pure benign workloads, we intend to use BWUs in our test-bed to record attacks mixed with background benign activity. In Fig. 1, we depict the architecture of our virtualized environment (single VMM) for attack trace generation. We use an *attacker profile* (e.g., novice, expert attacker) as a specification of different attack scenarios against the test-bed environment (e.g., an expert attacker might use sophisticated IDS evasion techniques specialized for virtualized environments [5]). In Fig.1, we depict two sequential attacks on a single target VM, one remote attack to gain access to the OS, followed by a local attack to gain access to the VMM. To record remote attacks, we deploy network packet sniffer in the host VM to capture network activities of the target VM(s) and the BWUs by monitoring the virtual switch. VMM-based IDSes use virtual machine introspection (VMI) routines/tools to monitor user- or kernel-level system activities of guest VMs relevant to attack detection (e.g., system calls, kernel routines). Thus, to record local attacks, we are creating *VMI logger*, i.e., set of VMI routines/tools used by VMM-based IDSes (e.g., XenAccess routines) which we extend to be able to capture system state and activities. The invocation of these routines/tools during the recording session is managed by a configurable *VMI agent*. To replay the *VMI logger* records, we are designing *VMI replayer*, i.e., set of VMI routines/tools which replace original VMI routines/tools and deliver recorded system activities instead of performing introspection. Although the proposed approach is VMM-dependent, most of the VMM-based IDSes are designed for a very small number of open-source VMMs which indicates its practical feasibility.

Metrics. We distinguish between two metric categories: security-related metrics (e.g., precision, recall) and performance-related metrics (e.g., capacity, resource consumption metrics). The current metrics for IDS benchmarking are defined with respect to a *fixed* set of hardware resources available to the IDS. However, cloud systems have *elastic* properties, i.e., hardware resources might be provisioned and used by an IDS on-demand. That might significantly affect the behavior of an IDS under test. Therefore, the metrics for benchmarking VMM-based IDSes should be calculated as a function of an elasticity metric. Such metric would quantify the ability of the cloud platform to dynamically scale up or down the hardware environment in which a VMM-based IDS resides. As part of our involvement in SPEC, we are participating in an effort to define elasticity metrics and to place them in context of IDS benchmarking metrics. An approach under consideration is to trigger the resource management mechanisms of a cloud platform to adapt to a given specific demand, and then to measure the adaptation time with respect to the amount of allocated/released resources.

Methodology. Currently, we are working on benchmark tests for the categories *capacity* (processing capacity, alarm reporting capacity, state tracking capacity), *resource consumption* (CPU, network, memory utilization), *performance overhead* and *attack detection* (attack detection accuracy, attack coverage, attack detection speed, resistance to evasion techniques). Given these categories, one may investigate important trade-offs, such as the one between the resource consumption and the attack detection accuracy of the IDS under test. An IDS benchmarking methodology would depend on the characteristics of the IDS under test. Thus, we are developing benchmark tests focusing on the most common characteristics of the existing VMM-based IDSes (e.g., use of anomaly- or misuse-based attack detection techniques). Regarding implementation-related concepts, the framework will integrate: (i) trace replay tools to replay the workload traces (e.g., tcpreplay for network trace replay, VMI replayer); (ii) workload controllers and BWUs, and (iii) tools for measurement, gathering and visualization of the benchmark metrics.

III. Conclusion

In this paper, we presented our current work and research roadmap towards developing a benchmarking framework for VMM-based IDSes which addresses challenges related to workload, metrics and methodology.

References

[1] F. Lombardi and R. Di Pietro, “Secure virtualization for cloud computing,” *Journal of Network and Computer Applications*, vol. 34, no. 4, pp. 1113–1122, Jul. 2011.

[2] S. Bharadwaja, W. Sun, M. Niamat, and F. Shen, “Collabra: A xen hypervisor based collaborative intrusion detection system,” in *Proc. ITNG’11*, 2011, pp. 695–700.

[3] Rg cloud working group: Spec research group. [Online]. Available: http://research.spec.org/working-groups/rg-cloud-working-group.html

[4] C. Binnig, D. Kossmann, T. Kraska, and S. Loesing, “How is the weather tomorrow?: towards a benchmark for the cloud,” in *Proc. DBTest’09*, 2009, pp. 9:1–9:6.

[5] A. Srivastava, K. Singh, and J. Giffin. (2008) Secure observation of kernel behavior. [Online]. Available: http://hdl.handle.net/1853/25464