Toward characterization of cloud execution environments

Mots clés :
- Directeur de thèse : refik MOLVA
- Co-encadrant(s) :
- Unité de recherche : Laboratoire de recherche d'EURECOM
- Ecole doctorale : Ecole Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Divers

Résumé du projet de recherche (Langue 1)

An increasing number of Internet services rely on Cloud infrastructure. Storage services such as Dropbox rely on the cloud storage of Amazon S3. Major Internet sites such as Netflix rely on the Infrastructure as a Service (IaaS) of Amazon AWS. Technicolor considers using the cloud for media related applications. Despite its commercial success, it is commonly accepted that delegating the execution of a task or storage of data to the Cloud comes with security issues. The abstraction layers provided by the cloud model hide the actual execution environment such as the hardware, the hypervisor used... It is not always clear which technology and architecture is actually hidden behind cloud products, even for products with specific security features (e.g. Amazon VPC, Amazon dedicated instance, domelS SaaS firewalls). As a consequence, an administrator has little information about the execution environment of its virtual machines, processes and tasks and may have difficulties to evaluate the security issues that might arise. The question that an administrator might want to ask can be as follows: On which hypervisors are my virtual machines running? Is it the latest version of the hypervisor? Where is the host machine located? In the U.S., Europe, India? What type of hardware is it running on? Which drivers are used? Is my virtual network properly isolated? Which are the other machines on my virtual network? Are there other virtual machines running on the same hypervisor? etc. A precise characterization of cloud execution environments is necessary. This characterization should be considered from a security perspective, thus take into account possible attacks, subsequent corrective actions and countermeasures. During this PhD we propose to investigate this characterization, following these items: - Evaluate the actual variability in security of existing cloud execution environments. - Propose and evaluate fingerprinting methods that characterize cloud technologies and architectures from inside and outside the cloud. - Propose and evaluate adaptive actions or countermeasures once the execution environment has been fingerprinted. These items and a methodology on how to address them are described in greater detail in the following sections. (Evaluate the variability) On paper, the cloud computing services proposed by the different cloud vendors are similar. However, in practice they rely on different technologies and the service provided may vary from a performance and security perspective from one cloud provider to another. Recent work [5, 6] show that important performance and cost variations exist between cloud providers and also within a single cloud provider. These works only considered performance and cost aspects but not security. We propose to characterize the variability in security of existing cloud execution environments. This step encompasses a measurement period where we observe architectural aspects of cloud deployments, such as n-tier isolation or DNS and firewall services provided by the cloud service providers. The observations will be conducted from different point of views: from a network point of view, virtualization point of view, physical machines etc. A partial map of the cloud infrastructure may be constructed from these observations, relying on observations such as traceroutes, port scans, fingerprinting of network elements etc. This approach differs from related work [1, 2, 4, 3] by the fact that we consider more architectural aspects and that we are interested in highlighting the differences in security between different cloud providers. (Propose and evaluate fingerprint methods) Fingerprinting is the action of gathering information about objects such as machines or drivers in order to identify them. Our objective is to design fingerprinting methods that can identify the technologies and architectures used within a given cloud. A process or machine running in the cloud should be able to determine in which environment it is executed just by gathering some observable characteristics of its execution environment. Fingerprinting often relies on machine learning algorithms. In this work we also adopt a machine learning approach by (i) clustering (unsupervised learning) a set of observed characteristics and (ii) labeling observations in a learning dataset and classify an object of the cloud thanks to its observed characteristics (supervised learning). The datasets used will be based on the previously conducted measurements. We will carefully evaluate the accuracy of the fingerprinting methods proposed, and also consider to which extend the methods can be attacked. One of the characteristics that may be used to fingerprint the physical machine and the hypervisor is the behavior of the clock skew based on work of Kohno et al. [9]. Chen et al. [13] observed that virtualized hosts have a more perturbed clock skew behavior and could determine the type of the hypervisor based on this behavior. While this work is promising, it is not sure that it can be applied in a more complex setting such as the cloud. Similarly, clock skews [14] have been used to reveal nodes within the Tor network. Such methods may reveal other architectural elements of the cloud and typically help identifying the physical machine a virtual machine is running on. Related questions that will arise during this study are as follows: What can be fingerprinted in the cloud? The cloud provider? The cloud virtualization technology; up to the version? the subversion? The physical device? The location? country? datacenter? container? rack? Many contributions have been made in the field of fingerprinting. Tools like NMAP, SinFP [7], CronOS/RING [8] perform OS/TCP stack characterization. Kohno et al. [9] characterize clock skews of remote hosts through a TCP/IP connection and similarly Jana et al. [10] characterize clock skew of wireless access points. Zalewski et al. [11] characterize the random generator used in different flavors of IP stacks. Neumann et al. characterize wireless devices [12] by observing frame inter-arrival times. While these techniques may be inspiring to design new fingerprinting method, they generally do not apply to the specific technologies, such as hypervisors and firewalls used by the cloud. (Propose and evaluate adaptive actions) Once an execution...
environment has been accurately fingerprinted, the concerned process or virtual machine should adapt accordingly. This may consist in aborting the current execution, migrating to another cloud environment, or adding additional layers of security such as asking for a trusted computing based (TCB) execution [15, 16], firewalls, obfuscating etc. The different adaptive actions that may be applied in the context of a cloud environment will be investigated. It is not clear today, how and to which extend the security solutions that exist in a non-cloud environment can be applied to a cloud environment. The adaptive actions will be evaluated from a performance point of view (overhead introduced) and from a security point of view. 

References