MOSAIC: A Platform for Monitoring and Security Analytics in Public Clouds

Alina Oprea  
CCIS  
Northeastern University  
Email: a.oprea@neu.edu

Ata Turk  
ECE Department  
Boston University  
Email: ataturk@bu.edu

Cristina Nita-Rotaru  
CCIS  
Northeastern University  
Email: c.nitarotaru@neu.edu

Orran Krieger  
ECE Department  
Boston University  
Email: okrieg@bu.edu

**Threats in public clouds.** Public clouds have enabled a number of new computing-intensive applications (e.g., personalized medicine, real-time speech recognition and machine translation) that positively impact our daily lives. Compared to traditional computing environments, public clouds offer many economical advantages to both users and service providers. However, the shared, large-scale infrastructure of public clouds amplifies well-known security risks and introduces new security threats compared to traditional organizational networks or private clouds. According to the Cloud Security Alliance (CSA) [2], the top security threats public clouds experience are related to: data breaches (malicious party gaining unauthorized access to data); data loss (permanent loss of data); account and service hijacking (attackers gaining access to critical credentials); denial-of-service attacks (inducing system slowdown and performance degradation); and abuse of cloud services (e.g., hosting of malware infrastructure in the cloud).

**Why existing defenses are not sufficient.** Reputable cloud providers implement a range of security functionality, such as data encryption and integrity, key management, replication, intrusion detection systems, and multi-factor authentication. While these improve the security posture of applications running in the cloud, they are not sufficient to prevent all possible threats experienced by cloud computing infrastructures. For example, in many cloud breaches attackers obtain access to valid user credentials [6], [11], [14] and use them to access sensitive data stores without detection by existing defenses. Similarly, insider attackers that exfiltrate sensitive information over the network are most of the time not detected by firewalls, intrusion detection systems and other security controls. These activities might induce a different pattern of access compared to historical user behavior and could be detected with machine learning techniques. As machine learning has been successfully applied to protect against a number of attacks in corporate networks and private clouds (e.g., [1], [3], [5], [7]–[9], [13], [15], [16]), we believe that it offers an opportunity to improve the security posture of public clouds, as well.

**MOSAIC platform overview.** We are designing a platform called MOSAIC for performing detailed monitoring of public Infrastructure-as-a-Service (IaaS) clouds at multiple layers. A preliminary design was described in [12]. We plan to construct analytics-based security services on top of the monitoring platform that use a variety of machine learning algorithms to profile the legitimate activity of cloud users and applications, and detect anomalous activities related to a wide range of attacks. In designing MOSAIC, we need to overcome a number of challenges related to the platform's scalability, performance overhead, as well as typical challenges encountered when designing machine learning algorithms for security applications (the limited availability of ground truth information, the validation of detected incidents, reducing false positive rates) [10]. An important emphasis in our design is to explore the tradeoffs between users’ privacy (relative to data collected by the cloud provider) and security protection of their resources.

**MOSAIC components.** In more detail, MOSAIC provides the following components (see Figure 1 for an overview):

- **A monitoring platform** for collection of metrics from different layers of the cloud (including the physical, virtual, networking and cloud management layers);
- **A data normalization and profiling architecture** to retain historical information of cloud utilization and application patterns over long periods of time;
- **An analytics-based security service** that employs a variety of machine-learning algorithms to detect anomalies relative to the behavior profiles and identify those related to security incidents;
- **Data and analytics APIs** enabling users to query and run analytics on the historical and real-time data relevant to their own workloads, without exposing sensitive information on other users’ workloads;
- **A set of mitigation strategies** that enables isolation of suspicious workloads, and investigation of detected suspicious behavior.

We are currently implementing the monitoring platform MOSAIC in the Massachusetts Open Cloud (MOC) [4], a public cloud used by five major universities in the state of Massachusetts for various research projects. We envision that our analytics-based security service will generate alerts of suspicious activities consumed by cloud administrators, offering an additional protection layer compared to traditional security defenses. MOSAIC will also enable cloud users to either use the analytics API or run their own algorithms to achieve a better security posture.
Fig. 1. High-level overview of MOSAIC architecture

REFERENCES


