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.

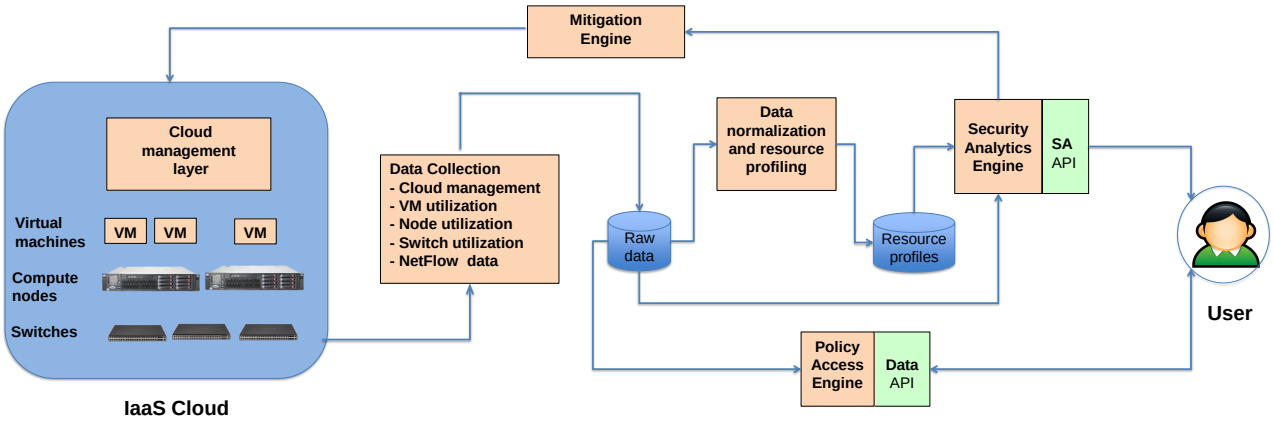

Fig. 1. High-level overview of MOSAIC architecture

REFERENCES

[1] Kevin M.. Carter, Nwokedi Idika, and William. W. Streilein. Probabilistic threat propagation for network security. *IEEE Transactions on Information Forensics and Security*, 9, 2014.

[2] Cloud Security Alliance. The notorious nine: Cloud computing top threats in 2013. Report available from www.cloudsecurityalliance.org, 2013.

[3] Pratyusa K. Manadhata, Sandeep Yadav, Prasad Rao, and William Horne. Detecting malicious domains via graph inference. In *Proc. 19th European Symposium on Research in Computer Security (ESORICS)*, 2014.

[4] Massachusetts Open Cloud. info.massopencloud.org.

[5] Terry Nelms, Roberto Perdisci, and Mustaque Ahamad. Execscent: Mining for new c&c domains in live networks with adaptive control protocol templates. In *Proceedings of the 22Nd USENIX Conference on Security*, SEC'13, pages 589–604. USENIX Association, 2013.

[6] Netskope. Cloud report. Report available from www.netskope.com/netskope-cloud-report, 2015.

[7] Alina Oprea, Zhou Li, Ting-Fang Yen, Sang H. Chin, and Sumayah Alrwais. Detection of early-stage enterprise infection by mining large-scale log data. In *Proc. 25th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2015.

[8] Douglas Lee Schales, Xin Hu, Jiyong Jang, Reiner Sailer, Marc Ph. Stoecklin, and Ting Wang. FCCE: highly scalable distributed feature collection and correlation engine for low latency big data analytics. In *Proceedings of the 31st IEEE International Conference on Data Engineering*, ICDE '15, pages 1316–1327, Washington, DC, USA, 2015. IEEE Computer Society.

[9] Ted E. Senator, Henry G. Goldberg, Alex Memory, William T. Young, Brad Rees, Robert Pierce, Daniel Huang, Matthew Reardon, David A. Bader, Edmond Chow, Irfan A. Essa, Joshua Jones, Vinay Bettadapura, Duen Horng Chau, Oded Green, Oguz Kaya, Anita Zakrzewska, Erica Briscoe, Rudolph L. Mappus IV, Robert McColl, Lora Weiss, Thomas G. Dietterich, Alan Fern, Weng-Keen Wong, Shubhomoy Das, Andrew Emmott, Jed Irvine, Jay Yoon Lee, Danai Koutra, Christos Faloutsos, Daniel D. Corkill, Lisa Friedland, Amanda Gentzel, and David Jensen. Detecting insider threats in a real corporate database of computer usage activity. In *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013*, pages 1393–1401, 2013.

[10] R. Sommer and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *Proc. IEEE Symposium on Security and Privacy*, 2010.

[11] Joseph Steinberg. Massive Security Breach At Sony – Here's What You Need To Know. http://www.forbes.com/sites/josephsteinberg/2014/12/11/massive-security-breach-at-sony-heres-what-you-need-to-know, 2014.

[12] Ata Turk, Hao Chen, Ozan Tuncer, Hua Li, Qingqing Li, Orran Krieger, and Ayse K. Coskun. Seeing into a public cloud: Monitoring the Massachusetts Open Cloud. In *Proc. USENIX Workshop on Cool Topics in Sustainable Data Centers (CoolDC)*, 2016.

[13] Ting Wang, Fei Wang, Reiner Sailer, and Douglas Schales. Kaleido: Network Traffic Attribution using Multifaceted Footprinting. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, SDM '14, 2014.

[14] E. Weise. Massive breach at health care company Anthem Inc. Available from www.usatoday.com, 2015.

[15] Ting-Fang Yen, Victor Heorhiadi, Alina Oprea, Michael K. Reiter, and Ari Juels. An epidemiological study of malware encounters in a large enterprise. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, pages 1117–1130, 2014.

[16] Ting-Fang Yen, Alina Oprea, Kaan Onarlioglu, Todd Leetham, William Robertson, Ari Juels, and Engin Kirda. Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks. In *Proc. 29th Annual Computer Security Applications Conference (ACSAC)*, 2013.