RobbLAW

IT Security &
Compliance Consulting

AI/Machine Learning for Cloud Security

Next-Generation Detection Analytics

By: Jonathan Presto, M.S.
Date: January 20, 2018

A lot has changed and developed over the last decade in cloud computing platforms. Cloud service providers are providing their customers with flexibility to customize the security of their environment. Services like serverless computing, docker images and containers, high-level API's for real-time streaming, or toolkits for secure coding of IoT applications are modern cloud capabilities without having the need to necessarily know everything "under the hood". The levels of abstraction have gone higher and computational speeds have scaled up and gone much faster. The challenge today lies with pace of adoption by large enterprises, largely due to the role of DevOps in ensuring secured and compliant infrastructures, maintaining existing operational workflows, training, and other various economical reasons unique to the complexity of the enterprise. Smaller companies or start-ups, on the other hand, do not usually have many complex layers of IT infrastructure, making them the forefront of early adopters for cloud computing platforms. This growth has been the driver for the maturity of cloud security solutions in the market.

Best practices in next-generation security are now coming from the cloud. Security professionals should not think of security as an after thought. Instead, they should design security solutions with tools and controls right from the beginning; as the old saying goes "security should be baked in, not sprinkled on top." Cloud services have integrated security controls and policies such as the ability to configure Identity Access Management, deploy advanced encryption features, use multi-factor authentication, stream encrypted data, and provide proactive tooling and automation. *Automation* is a key component in next-generation security because humans cannot keep up with the increasing number of vulnerabilities. Older systems also lack the ability to proactively detect new threats. Artificial Intelligence and Machine Learning algorithms that integrate with cloud computing platforms enable new threat detection.

Currently available security tools are pretty good at detecting *known* attack patterns. If an attack matches a signature, talks to a known bad place, uses unencrypted protocols, or happens within the infrastructure that you closely monitor, you can reliably detect it as it occurs. What businesses struggle with is having the capability of detecting *unknown* attack types, new malicious behaviors and insider threats. Security professionals also struggle with attackers hiding within a bell curve. In the past, many attacks occurred at 3pm on a Friday just before a three-day weekend. This allowed plenty of time to break in, ransack the place, clean up, and install a back door for persistent access. Today, things have changed and we now often see attacks on Wednesdays at 10am, because our adversaries understand that we are sensitive to volume, and that is the hour of peak network traffic (e.g. caused by high online customer activity) based on well-understood behavior. Our adversaries know how to hide in our normal bell curve of network activity.

When envisioning the future of detection analytics, imagine the current enterprise security landscape as a river delta. The delta consists of streams, rivers, and an ocean. Endpoints produce small streams of operational data. This data then flows downstream, where it gets aggregated into rivers of enterprise log and security data. The rivers include business operations context, IT operations logs, and Information Security events. When you aggregate and monitor these using real-time correlation in a Security Information and Event Management system (SIEM), they anchor the Security Operations Center (SOC), which monitors real-time correlated security events to detect Indicators of Compromise (IOC).

Given the modern threat landscape, you can now picture how real-time capability requires a correlation and longer-term analytical capability as a supplement. You need to expand operational post-hoc analytics to the data ocean by assigning this work to a team of security professionals, especially when an unknown attack takes place. Once an attack type is detected, it is then converted into automated real-time detection, so in the future you can catch it and respond to it in real time. The tactical technologies for breach detection and prevention are in the streams of data (e.g., intrusion prevention), operational monitoring capabilities sit across the rivers of data (e.g., SIEM), and any strategic data analysis for breaches resides in the oceans of data (e.g., security professionals). People and processes are a critical link between these levels.

Artificial Intelligence and Machine Learning have enabled advanced detection capabilities in next-generation detection analytics, which are available in today's cloud computing platforms

via Security as a Service (SECaaS). Advanced *detection* includes advanced statistical analysis, the ability to train models and detect anomalies across any of your security data feeds. A bad actor has to deviate on at least one measurable parameter in order to conduct a malicious activity. Data mining provides the means to *explain* patterns. The main disciplines of data mining are: clustering, classification, correlation, aggregation and affinity grouping:

- **Clustering** - when you cluster security data, you almost inevitably identify large numbers of false positives. When you clean up these false positives on your security, you make the deep, muddy river of data that you are monitoring clearer and shallower.
- **Classification** - which is used to classify data into types for comparative analysis and cross-correlation.
- **Correlation** - which has been around a long time and provides the analytical engine for modern enterprise SIEM deployments.
- **Aggregation** - imagine an aggregate profile of a server, a user, or an IP address based on long-term historical behavior information; this type of aggregation can easily profile an attacker, allowing you to identify similar profiles with the addition of a scoring algorithm.
- **Affinity grouping** - which can be described anecdotally using the Netflix Recommender System. According to a Netflix research paper [Gomez-Uribe, Carlos and Hunt, Neil. The Netflix Recommender System: Algorithms, Business Value, and Innovation. 2015], Netflix categorizes viewer behavior and uncovers interesting insights about them to build their "Because You Watched" (BYW) rows of recommended movies. A BYW row anchors its recommendations to a single video watched by the subscriber. The video-video similarity algorithm (sims) drives the recommendations in these rows. For example, as a personal subscriber and viewer of the "The Crown", I am presented with the recommendation for "Grace and Frankie" because there is a high likelihood of this pattern exhibited by other subscribers; the collection of past actions by subscribers tuned the parameters of the sims algorithm to strengthen this particular affinity grouping between "The Crown" and "Grace and Frankie."

In a security context, malicious Command-and-Control (CNC) infrastructure within your environment has a higher affinity for itself and its peer nodes than for the normal infrastructure surrounding it. This allows for a significant advance in detection capability to further *explore* and *understand* the data. Analytical query takes your big data and turns it into small data,

which you can perform further data exploration. Big Data management takes large amounts of structured and unstructured data and organizes it into small amounts of data that can be retrieved in a reasonable time frame. These are often called queries or data marts, and these are where a security professional applies their advanced analytical capabilities. Technical intelligence is the concept of producing your own IOC to understand the data. Currently, the majority of the industry purchases this intelligence from trusted vendors. It is generic intelligence aggregated from open sources and occasionally augmented by honeypot collection projects. The ability to detect new malware in your environment, detonate it, and produce indicators to be shared across your organization begins to simulate an immune response, and that is clearly a desired direction for any business in today's world.