CyberActive℠

Use Cases for Security Analytics

Author:

Vinod Vasudevan,
CTO, Paladion Networks

PALADION
BETTER SECURITY OUTCOMES

Use Cases for Security Analytics

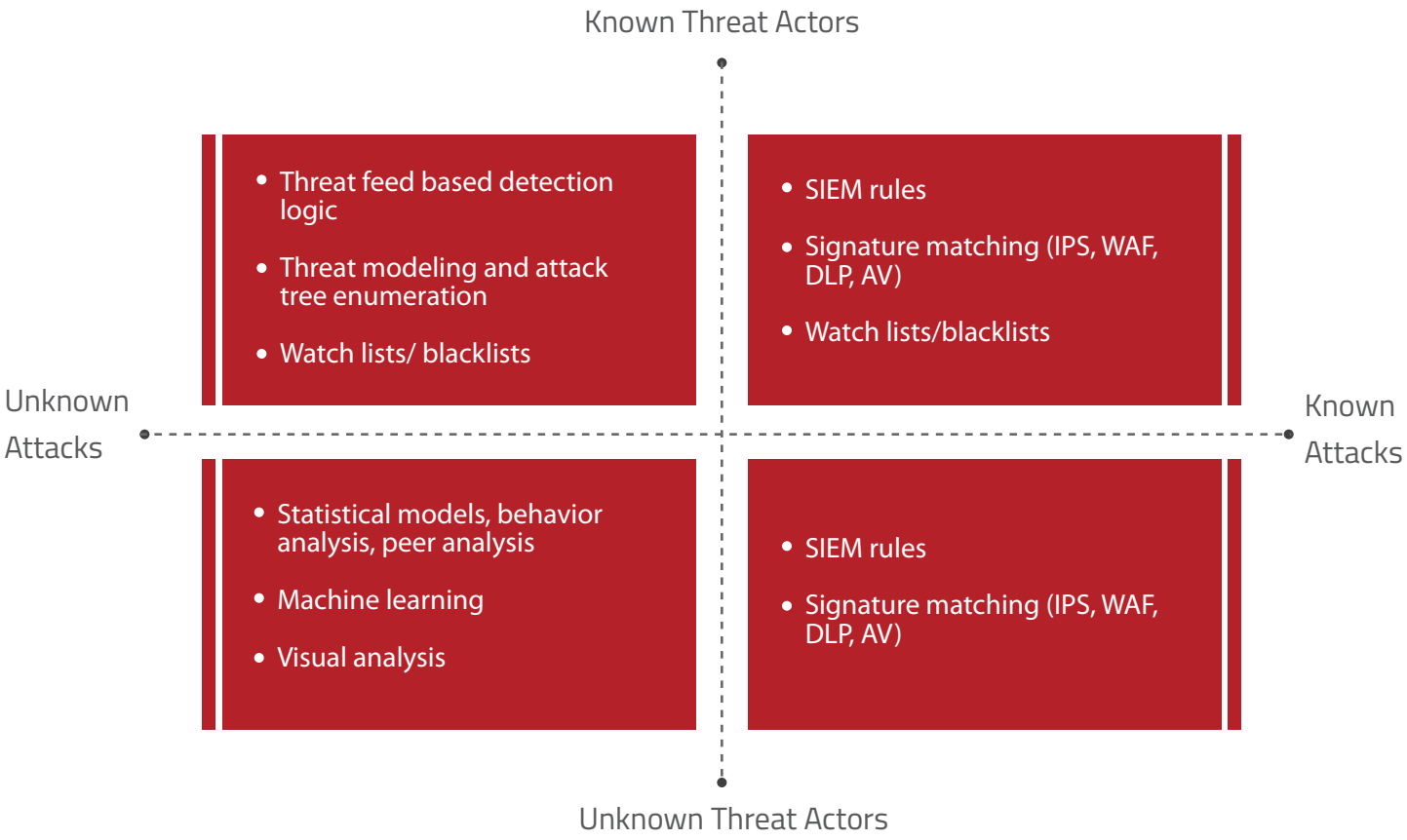
EXECUTIVE SUMMARY

Security analytics can be a valuable tool for detecting advanced attacks. However, it must be applied correctly. Too often, the goal of security analytics is reduced to the construction of a big data platform; running data science algorithms, machine learning, or statistical packages. Instead, the starting point should be to identify the risks that cannot be monitored through conventional security products and then to define use cases in security analytics to monitor those risks. In this paper, we discuss the need for security analytics and how to apply it in a meaningful way to achieve results. We then discuss the technology components required to put security analytics in action.

"Too often, security analytics is reduced to the construction of a big data platform. The starting point should be to identify risks and define use cases."

THE CONTEXT FOR SECURITY ANALYTICS

Threat actors have evolved. They now often employ types of attacks that cannot be detected using signature matching or predefined rules. The threat landscape can be mapped using a grid with two dimensions: attacks (whether known or unknown) and attackers (known or unknown). Attacks that are known can be detected using the rule-matching technology of antivirus (AV) software, intrusion prevention systems (IPS), web application firewalls (WAF), data loss prevention (DLP), and security information and event management (SIEM). When attackers are also known, their attributes (IP addresses, URLs, files) can be used to further prioritize the attacks.



When attackers are known through threat intelligence feeds, but the attack methods are not known, a big data platform can be used to collect Netflow, proxy and user access data, and to match this against known malicious IP address or other known indicators of compromise (IOC) related to the attackers. Threat intelligence is also used for carrying out hunting activities within the network. Both of these approaches need specialized tools and may be grouped under the broad umbrella of security analytics.

Signature, rules, and threat intelligence fail when it comes to detecting unknown attacks from unknown threat actors. Undetected attackers can stay within the network of an organization for longer, navigating towards critical or valuable assets. Their attacks usually fall into the category of advanced persistent threats (APT) or advanced targeted attacks (ATA).

FireEye Mandiant report (M-Trends 2015: A View from the Front Lines[1]) states that it takes a median time of 205 days to detect an advanced persistent threat. The damage is often high by the time the attack is detected. Reducing the time to detect such attacks can considerably reduce the impact from breaches.

"Reducing the time to detect such attacks can considerably reduce the impact from breaches."

The critical need for security analytics is for detecting such unknown attacks. By comparison, although incremental value might be gained in applying security analytics in the other three quadrants of the grid, the benefit to effort ratio is small. As a result, whenever security analytics is being considered for attack detection, the acid test is to see if such attacks can be detected through rules, because the attack, the attacker or both are known. If a rule-based approach is insufficient, then the next step is to define the appropriate use case(s) to apply security analytics. We will illustrate this approach with a few examples of use cases in security analytics.

"The acid test is to see if attacks can be detected through rules. If not, the next step is to define use cases to apply security analytics."

1 https://www2.fireeye.com/rs/fireye/images/rpt-m-trends-2015.pdf

Use Case 1: Detection of Malware Beaconing Activity

Security analytics can be used to detect beaconing activity from unknown malware and from that detect the unknown malware itself. This applies to malware used in APTs and ATAs. A rule-based system such as SIEM or a signature-based system including IPS, anti-malware, and WAF cannot detect such malware attacks. Even a sandbox technology approach for detecting malware will fail, because today's malware stops executing when a virtual environment is detected.

Before deciding to use security analytics, we should recognize that there are two other ways to identify unknown malware. One of them is to use external threat feeds. Most malware sends out regular heart beat information to its command-and-control (C&C) server. Using external threat intelligence, these communications can be identified and flagged to security investigators. The other technique is to look for a fixed pattern of beaconing: for instance, data packets of a certain size sent at a certain frequency. If the size and frequency is known, such rules can be modelled in SIEM.

Advanced attackers use command-and-control (C&C) servers not yet on any threat feed. They also use beaconing tactics that do not follow a known pattern of size and frequency. This is the use case we now need the security analytics to solve – advanced malware with no known C&C server and beaconing pattern.

"Advanced malware with no known C&C server and beaconing pattern must be solved by security analytics."

There are nonetheless traces of malware activity in different sources in the IT environment. One such source is a proxy. It will contain the heart beat traffic, even though that traffic cannot be detected by rules. Using analytics, this heart beat can be detected by applying entropy techniques.

Entropy in data science terms refers to "uncertainty of data". When we look at proxy data in general, the data related to user interaction with URLs is expected to be randomly distributed in terms of size of interaction. After all, people visit a variety of websites and upload and download a variety of data. Therefore, data sizes are expected to be highly variable. In this case, when we examine the entropy of the byte size of the communication between a user and a URL, the entropy and "uncertainty of data" should be high.

The heart beat information being beaconed out by a lot of malware is characterized by similarity and regularity, even though neither are known a priori for us to build rules. If we apply an entropy function to this data, the entropy will be very low, since the byte size and frequency are relatively uniform when interacting with the C&C URLs. This enables us to detect the "unknown" attack, even though the attack signature or attacker is unknown.

"Security analytics enables us to detect cyber attacks, for which the attack signature or attacker is unknown."

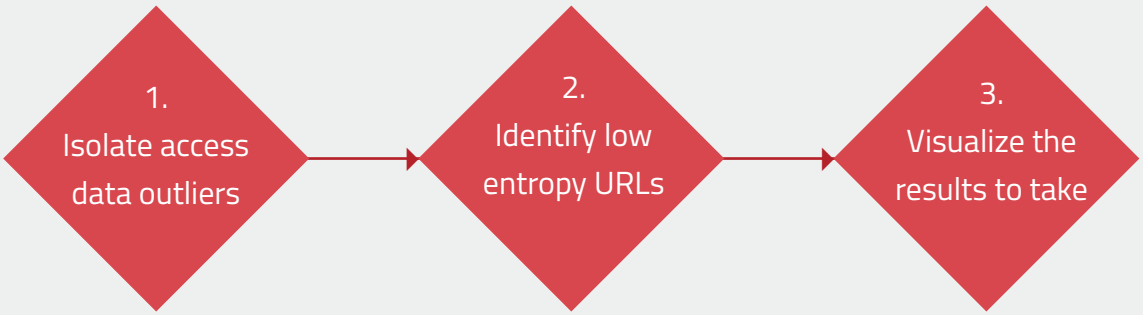
Use Case 2: Detection of a Watering Hole Attack

A 'watering hole' attack is used to infect hosts by luring users to a location (URL/IP) where the malicious code is hosted. In a similar way to the previous discussion, if the URL hosting malicious code is represented in blacklists of threat feeds, this is not a use case for security analytics. It can be detected using SIEM with external threat feed integration. Also, if the file size or file pattern is known, it can be spotted by writing specific rules based on items such as IPS signatures, URL filters, and sandboxing for any one of a number of security products.

On the other hand, we need security analytics to detect infection through water holing where the file size, URL, and file pattern are all unknowns.

Detection of a waterhole attack requires three steps. First, outliers in URL access data must be identified, based on the number and frequency of accesses over the past few days. Second, an entropy function is applied to the file size/download content to identify low entropy URLs. This generates a repository of all potentially compromised hosts and likely water holing URLs they have visited. As the third and final step, this data is converted into a tree map to help security analysts quickly visualize the hosts and URLs, and start investigating the larger nodes of the tree map.

"Detection of a waterhole attack requires three steps – isolate access data outliers, identify low entropy URLs, and visualize the results to take action."



Use Case 3: Data Exfiltration

The application of security analytics to data exfiltration makes it possible to detect new scenarios in data leakage beyond those identified by standard data loss prevention (DLP) solutions. The standard solutions can detect leakage if we know the data or the pattern of the data we want to protect: for instance, card information, specific data fields, or file signatures. However, when we want to detect instances of data leakage beyond known data or when the communication is encrypted, we need security analytics.

As an example, HTTP POST is a common method used for the upload of files. Such uploads could be valid business requirements. However, they could also be malicious data exfiltration by malware or a rogue insider. Detecting such illicit exfiltration is possible by base lining HTTP POST traffic from each source system in an organization and then detecting outliers with respect to this traffic. This enables us to detect any abnormal traffic movement from any system in the enterprise, and to take steps to mitigate it. In a self-learning system that builds on itself, baselines are created and updated over a period of time. There are many other channels (e.g. email) to which similar analytical techniques can be applied to detect data exfiltration.

"A self-learning system, which builds on itself, detects data exfiltration."

The above three use cases are samples. There are many more use cases that can be addressed using security analytics. Paladion webpage http://www.paladion.net/cyber-active/ offers more examples of security analytics use cases.

Use Case Best Fit

Overall, we can classify use cases in three broad categories: real time rule-based use cases, real time security analytics use cases, and batch security analytics use cases.

Real time rule based use cases – Use cases for attacks or attackers that are known and that do not need to be compared with past attack history. They can be defined and detected using rule-based approaches such as SIEM, IPS, WAF, and DLP. As an example, an attack originating from a blacklisted IP address corresponds to a simple rule for matching the source IP address of the event with the available blacklist or global threat database. Similarly, rules for known attacks are signature rules in SIEM, IPS, WAF, and other rule-based systems. Known indicators, such as a high number of login failure attempts within a short time, can also be defined in rule-based systems. Compliance-related measures including logins after office hours, logins from suspicious geographies, and access using non-standard remote clients can all be configured as use cases in an SIEM system.

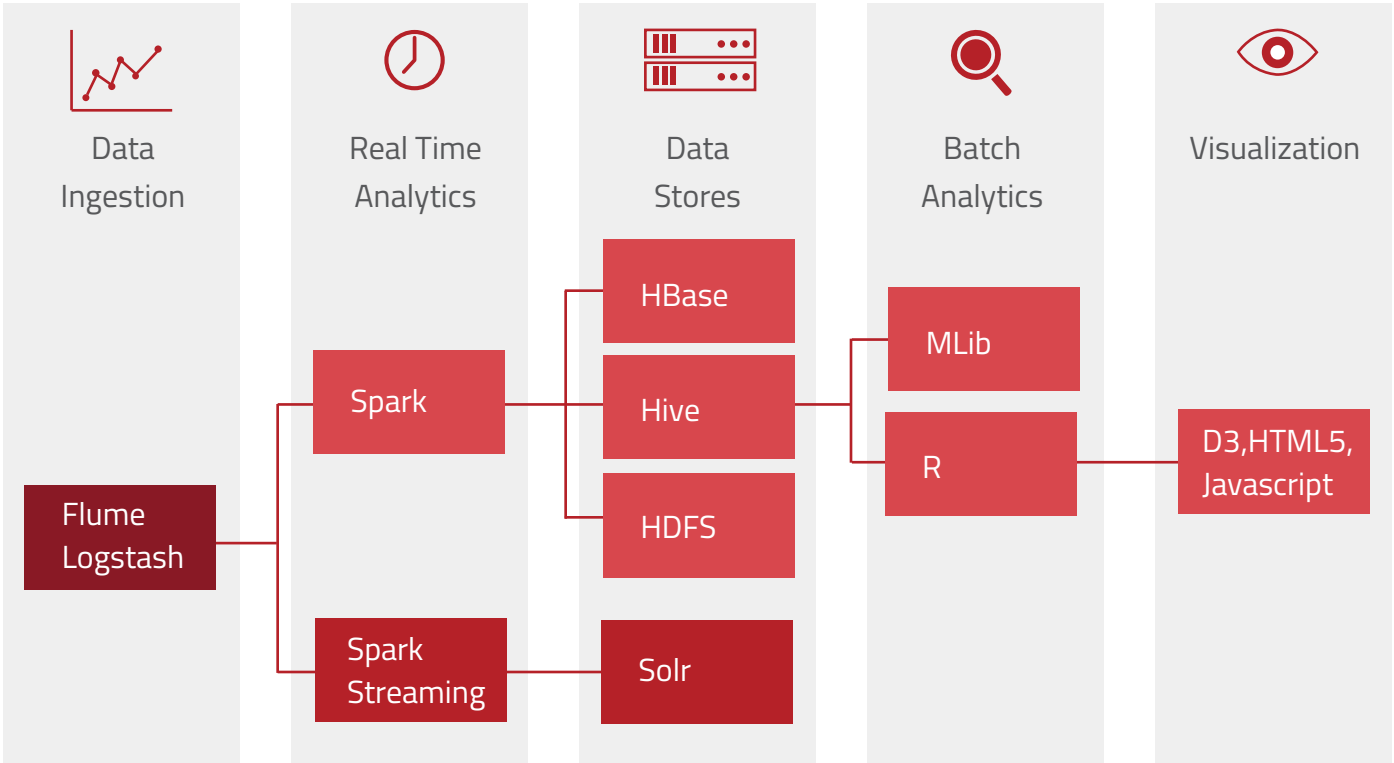
Real time security analytics use cases – Use cases for triaging incoming alerts from other real time systems (SIEM, IPS, WAF, etc.) or matching a pattern that needs longer period data for detection. As an example, an SIEM alert can be better prioritized by quickly assessing the attacker's IP address in real time for the past volume of attacks from that IP address, the severity of those attacks, other destinations targeted by the same address, and user parameters including vacation information. There are other similar parameters that can also be used to ascertain if the event is worth the effort of remediation. All of this is achievable using real time security analytics to leverage statistical models and rapid searches of large datasets.

Batch security analytics use cases – Unknown attacks and attackers are best handled in batch or near real time analytics use cases. In these cases, detection involves using deeper statistical models and profiling large data sets. These are periodic analytical jobs that run on data to produce output that is then visualized. The time to process this type of output is usually a function of the type and quantity of hardware that can be deployed. Accordingly, the periodicity of the analytical models used may range from a day to less than an hour. The use cases discussed above of malware beaconing, watering hole attacks, and data exfiltration fall into this category.

"Unknown attacks and attackers are best detected in batch or near real time analytics use cases, with deeper statistical models and large data sets."

Security Analytics Architecture

The architecture must support the functionality that we have discussed above. After data input using technology such as Flume or Logstash, real time analytics on large data sets can be performed with Spark and Spark streaming. Since datasets are large, they will need to be stored in suitable data stores such as Hive and HBase. The fast search feature on datasets to match patterns can be supported by Solr. Batch analytics that require deeper statistical models can be performed using statistical packages such as R and MLib. Interactive visual querying can be achieved using D3 (Data-Driven Documents) functions. The components described here are simply to illustrate possibilities. Other similar products can also be used to build the big data analytics platform.



Summary

Security analytics needs clear definition of use cases before designing the platform and data science packages. A suitable method for defining use cases is to identify the risks that an organization wants to monitor and then find the gaps in existing rule-based systems. These gaps can be addressed using data science techniques. Some of the techniques must be executed on a real time platform. Others are better served by a batch processing platform.

PALADION

BETTER SECURITY OUTCOMES

ABOUT PALADION

Paladion Networks is a specialized partner for information risk management providing end-to-end services and solutions in the US, Europe, Asia and the Middle East. Paladion is rated and has been recognized and awarded by Gartner, Asian Banker and Red Herring, amongst others.

For over 15 years, Paladion has been actively managing information risks for over 700 customers. Paladion provides a complete spectrum of information risk management comprising of security assurance, compliance, governance, monitoring, security analytics and security management services to large and medium-sized organizations. Paladion is also actively involved in several information risk management research forums and has authored many books on the same. With a staff of over 800 dedicated security experts, Paladion has 6 Security Operations Centers (SOCs) across the world.

Head Office: Bangalore: Shilpa Vidya, 49 1st Main, 3rd Phase, JP Nagar, Bangalore- 560078
Phone : +91-80-42543444, Fax : +91-80-41208929
Abu Dhabi: +971-55-9891227, Bangkok: +66 23093650-51, Doha: +974 33559018, Dubai: +971-4-2595526, Jakarta: +62-8111664399, Kuala Lumpur: +60-3-7660-4988, London: +44(0)20 7148 7475, Mumbai: +91 022 33655151, Riyadh: +966 (0) 11 4725163, Stuttgart: +49-711-7224-9626, Toronto: +1-416-273-5004, Virginia: +1-703-8713934, Muscat: +968 99383575 (Business Associates)
sales@paladion.net | www.paladion.net

PALADION