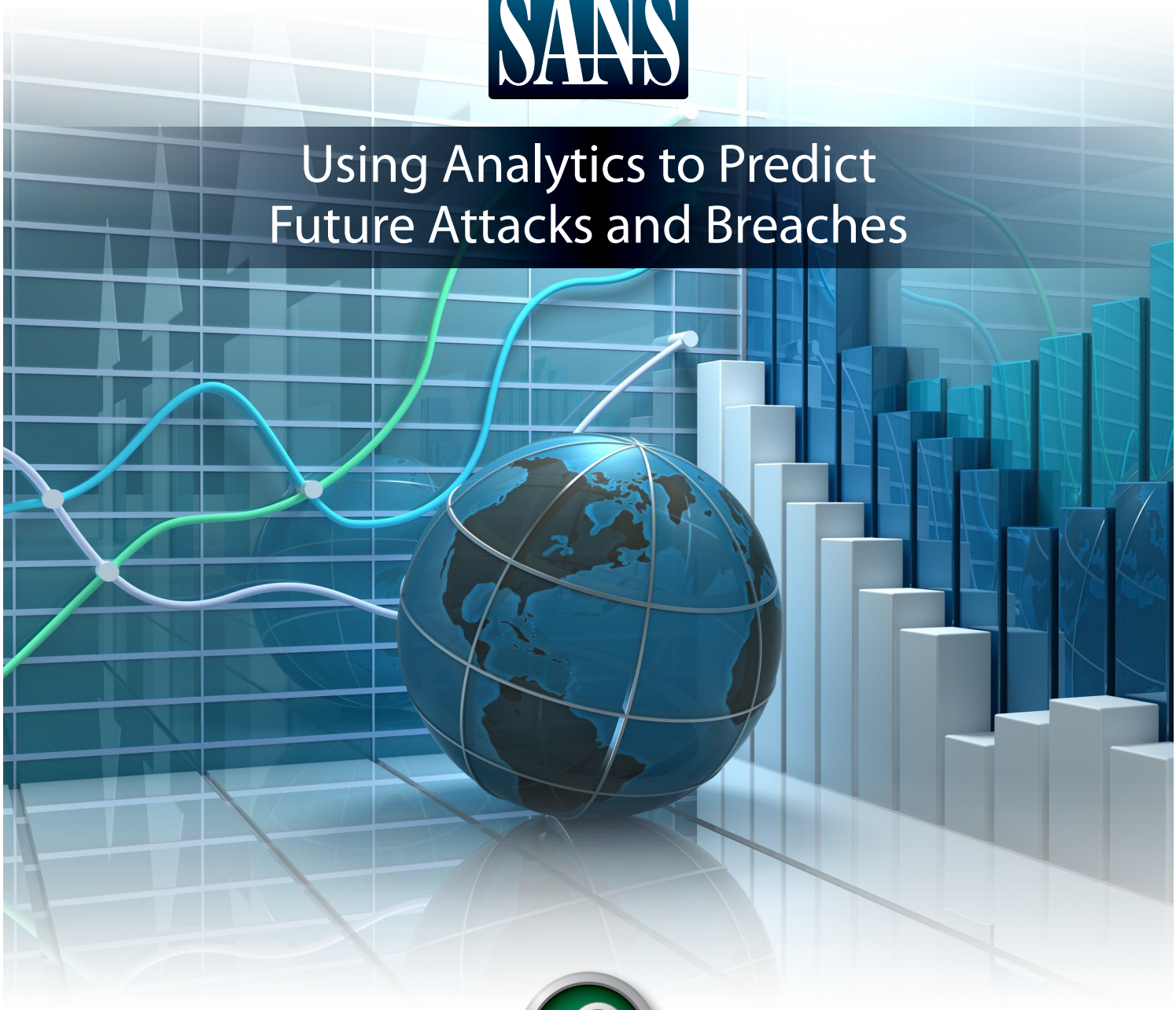

Using Analytics to Predict Future Attacks and Breaches

A SANS Whitepaper

Written by Dave Shackleford

January 2016

Sponsored by
SAS

©2016 SANS™ Institute

Introduction to Today's Monitoring Landscape

The pace and sophistication of data breaches is growing all the time. Anyone with valuable secrets can be a target, and likely already is. According to the Privacy Rights Clearinghouse, at the time of this writing, 884,903,517 records were breached in 4,621 incidents documented since 2005.[1] This number is just an estimate based on publicly disclosed and well-documented incidents; the real number is likely much higher. According to data available from datalossdb.org, the size of the major breaches over the past several years has grown significantly:[2]

- 77 million customer records and possibly payment card information were stolen from Sony in April 2011.
- Adobe Systems infrastructure was hacked, leading to the loss of 152 million names, customer IDs, passwords, encrypted payment card information, and source code in October 2013.
- Target Brands lost 110 million customer records, and credit and debit card numbers due to hacking of point-of-sale and other systems by attackers using sophisticated malware in December 2013.
- Anthem was hacked, leading to loss of 78.8 million records that included personal data and Social Security numbers (announced in February 2015).
- Experian announced a breach of 15 million T-Mobile customer records in October 2015.

In most of these cases, sophisticated attackers targeted the companies and organizations and deliberately went after some of their most sensitive data. It's clear that the security strategies we've used in the past are increasingly less effective against these new types of attacks.

Many tools and security processes have been more focused on prevention than on detection and response, and attackers are taking advantage of the fact that organizations are not finding the indicators of compromise within their environments soon enough, nor are they responding to these incidents and removing them quickly enough.

In addition to rapid event detection, correlation and response, we need the capability to predict future trends based on past and current behavior, which is where security analytics may prove useful. This paper explores the growing necessity of security analytics and looks at some sample use cases to support its adoption.

[1] www.privacyrights.org/data-breach

[2] http://datalossdb.org/index/largest

The Current Landscape of Detection

Organizations are often compromised in minutes, and attackers can persist within organizations for extended periods of time. Why is this? The answer may have something to do with the detection and response tactics we've been using. Are they failing us? Let's explore some of those tactics here and see what has worked well and what may need to change.

In the realm of detection, we've leveraged a number of technologies and tools over the years, as represented in Figure 1.

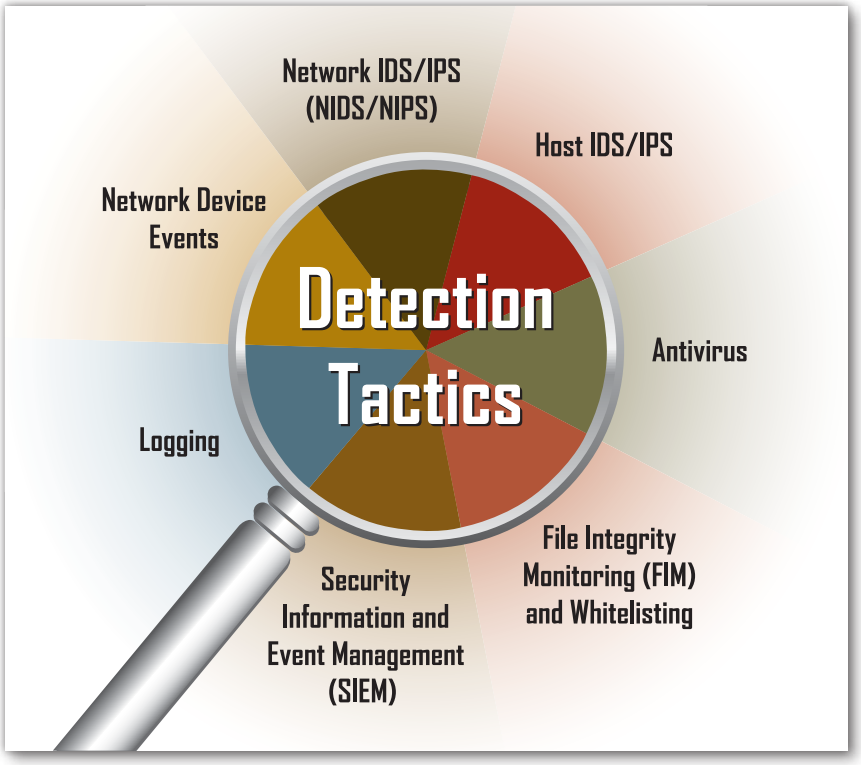

Figure 1. Detection Tactics

Some of the more prevalent ones, with benefits and limitations, include:

- **Logging:** Centralized logging is a primary control for detecting security incidents today. Organizations increasingly work to gather and aggregate logs for analysis, with use cases ranging from IT operations troubleshooting to security event analysis. With this wealth of important data has come a significant challenge, however—many organizations are not able to wade through the data and successfully detect and prioritize the most meaningful events in their environments due to complexity of data involved, volume of data, lack of skill sets, and tools that do not facilitate rapid and intuitive searches. While logging and log analysis is a critical, and often required, aspect of a security program, a fair amount of operational time is needed to properly develop analysis rules and alerts that benefit security teams.

- **Network device events:** Network devices such as firewalls, proxies, routers and switches can generate events in the form of standard syslog logs, vendor-specific event formats, and SNMP events. Access control rules, firewall rules and authentication logging can all provide important insight into what is happening on the network, as well as the configuration and management of the devices themselves. These devices play an important role in preventing attacks, as well as in detection of potential or existing adversaries looking to gain access into a network environment. As with general-purpose logging, however, the volume of network events is often overwhelming for today's security teams. It needs to be carefully analyzed and sorted to properly weed out false positives and prioritize the most important security information.
- **Network IDS/IPS (NIDS/NIPS):** Network intrusion detection and prevention systems are staples of network event generation that can often detect well-known signatures of attacks or unusual patterns in network traffic. These platforms are exceedingly common in a defense-in-depth security architecture today, but they also generate a staggering array of alerts until properly tuned and are also prone to false positives. Signature-based NIDS/NIPS platforms are also limited to detection of known attacks.
- **Host IDS/IPS:** These applications can provide enormous amounts of useful data, especially when combating malware and attacks directed toward client-side software on endpoints. Much like network IDS/IPS, however, host-based IDS/IPS must be properly tuned to have true value for security analysts.
- **Antivirus:** Antivirus tools are becoming less and less valuable over time, largely due to the system overhead used and the fact that many antivirus agents are focused primarily on signatures of well-known malware. Management and control of distributed agents on all endpoints can be an operational challenge as well. Aside from compliance requirements to have antivirus installed, security analysts may still find value in antivirus alerts, as they can indicate a broad array of malware infection attempts. In some cases, malware can be blocked or quarantined as well, giving incident responders time to properly contain the issue.

- **File Integrity Monitoring (FIM) and Whitelisting:** FIM and whitelisting are two additional host-based security controls that focus on changes to critical files and rules that allow only "known good" actions or content access by approved users. Both of these technologies can act as excellent prevention and detection controls, with valuable events generated and correlated with other system and environment information. Once tuned and installed, both can add enormous value to detection and response programs. The biggest challenge in using FIM and whitelisting agents is the up-front requirements to properly configure and tune the policies necessary for proper functioning.
- **Security Information and Event Management (SIEM):** SIEM platforms can gather and correlate information from numerous devices and applications, allowing security analysts to monitor the environment more thoroughly and develop much more sophisticated detection rules. Many security operations center (SOC) teams use SIEM tools as the primary monitoring and detection platform in the environment. Although SIEM tools are immensely powerful and provide advanced security monitoring and response features, they can also be complex to configure and manage.

All of these detection and prevention tools can play a role in incident response (IR), and many other tools are also available today. One that is gaining significant traction in many organizations is malware sandboxing (sometimes called "detonation" platforms). These systems can analyze the behavior of malicious code as it tries to execute in an isolated and contained environment.

Despite these tools, some security teams are less effective than they could be because these disparate tools and platforms generate an overwhelming amount of data. Security teams are trying to incorporate numerous controls with detection events into their response processes, and it can be easy to miss events and indicators of compromise. In the 2015 SANS Analytics and Intelligence Survey, 53% of respondents said they were dissatisfied with visibility into external adversary infrastructures based on intelligence and analytics processing.[3]

[3] www.sans.org/reading-room/whitepapers/analyst/2015-analytics-intelligence-survey-36432

In addition, many teams tend to store a great deal of event data with the intention of analyzing it later, only to never find the time.

Many security teams are also using manual processes to initiate incident investigations and follow through with containment and elimination steps. This can be a slow process, leading to attackers moving laterally throughout many networks before investigators can detect, properly respond to and remediate intrusions.

Overall, most organizations are in "reactive" mode today, trying to discover evidence of attacks and potential breaches from finite events using correlation tools such as SIEM and log management. However, the attacks today are much more subtle than before and may be detected only by looking into larger and longer-term patterns of behavior in the IT environment. Are analysts asking the right questions? Do we have the tools to do this today?

Why Many Monitoring Tools and Strategies Fail

Most IR tools generate event data. Log management and SIEM platforms have traditionally been able to aggregate these events and allow security teams to search and correlate all this data. While that's still the case, today we are drowning in data. Plus, we're dealing with many unique types of data that require analysis, compounding the problem. At the same time, security teams are facing pressure to detect attacks and respond to them more rapidly.

To facilitate more effective detection, we need to better accommodate larger datasets and analyze the data using more advanced techniques.

- Security event monitoring uses a broad variety of data and data types, including logs and events from systems, applications, network and security devices, and other sources covered earlier. This data typically resides in traditional relational databases, central log stores, or some type of SAN or NAS platform. Security responders are still missing a vast array of other IT data, especially data related to sensitive information, so organizations need even more data types for advanced correlation, and data analytics platforms must be able to process this data.
- Security teams also have to look for event types they don't know about yet, so security analytics platforms should allow for the creation of new correlation rules and the discovery of new trends and behavioral patterns in the environment.
- Having more data will allow security analysts to more readily develop trends and patterns that can be used for accurate predictive security intelligence, learning from the environment over time and developing new types of alerting. External threat intelligence feeds can be a huge element of this new model of security data analysis, too, providing additional context from new outside data sources.

We need more data from more different sources, over longer periods of time, to really develop an understanding of what's happening in the network environment.

By themselves, signature-based detection methods don't work well anymore. Signatures for intrusion detection and anti-malware tools catch some known threats, but cannot keep pace with the subtle and varied attacks perpetrated by today's more advanced adversaries. Sophisticated attackers are behaving more stealthily, seeking to avoid detection from traditional signature-based tools such as antivirus software and NIDS, and we ultimately need more data from more different sources, over longer periods of time, to really develop an understanding of what is happening in the network environment.

For example, consider an attack workflow that begins with a user receiving a phishing email. If the user clicks a link that directs him or her to a malicious website, a previously unknown piece of malware may execute on the system, giving the attacker an initial foothold in the environment. The attacker may then compromise the user's credentials, connecting to other systems and moving laterally through the network masquerading as the original user as well as other users along the way. This is a simple and very common scenario seen in attacks today, yet most signature-based tools would not detect this type of activity.

Unfortunately, most organizations are struggling to find the right skill sets to properly operate and maintain a security analytics platform for detection and response. In fact, this problem was overwhelmingly cited as the top impediment to discovering and following up on attacks in the SANS 2014[4] and 2015[5] Security Analytics Surveys. In 2015, 24% of respondents reported that this was their top impediment to discovering and following up on attacks.

Most teams should be looking for some form of abnormal behavior, but they don't really know what they are looking for or how to spot it.

Most security event management tools and activities have been focused on detecting events happening right now, and possibly developing some behavioral trends and analytics. Detection and response teams hunt for evidence of malicious activity in the datasets they gather, and this is becoming more challenging than ever before. More attacks and attackers are out there, much larger datasets must be combed through, and one fundamental problem looms above all: Most teams should be looking for some form of abnormal behavior, but they don't really know what they are looking for or how to spot it.

In addition to rapid event detection, correlation and response, we've been sorely lacking the capability of predicting future trends based on past and current user and attacker behavior, and this is one of the goals of security analytics. At the same time, having more data on which to perform root cause analysis and forensics after the fact can only help security teams improve their ability to look for specific threats, thus leading to a more proactive monitoring approach.

[4] www.sans.org/reading-room/whitepapers/analyst/analytics-intelligence-survey-2014-35507

[5] www.sans.org/reading-room/whitepapers/analyst/2015-analytics-intelligence-survey-36432

Looking Toward "Real" Security Analytics

This is really what we are asking of security analytics platforms today: Collect large and disparate internal datasets, comb through that data looking for patterns and creating correlations, and provide guidance on anomalies and the potential threats we face.

What exactly *is* security analytics? Security analytics analyzes large datasets with technologies that enable rapid and accurate analysis, correlation and reporting to identify events and patterns of interest that may indicate malicious behavior in the environment. With more data to work with, security teams can analyze the environment based on:

- Timing of events
- Sequences of occurrence (and time)
- Differences in data from various sources
- Real statistical analyses (time series plots of risk and behavior, machine learning, etc.)

For example, security analytics could allow more advanced visibility into indicators of compromise, such as:

- Phishing in mail logs, using trends and correlation capabilities to assess the affinity of senders to the organization
- Slow data exfiltration in proxy/firewall logs, looking at the number of bytes and sessions over time
- HTTP-based malware command and control channels (C&C) in Web proxy logs, looking for long URLs without referer fields and Base64 POST variables in the headers

Analytics is not just another term for security event management. Analytics systems need to perform rigorous analysis of many disparate types of data, and also provide strong correlation and statistical tools for security analysts to use in developing baselines of normal behavior. In large networks, few tools are able to digest petabytes of network traffic, billions of network flows, and numerous other alerts and data types that can align with network data to determine whether unusual activities are occurring.

In addition to the tactical detection and response capabilities, analytics tools should enable investigators to perform deep root cause analysis of incidents and develop predictive models of future behavior based on knowledge of patterns in the data center.

Machine Learning

Much has been discussed in the information security community about "machine learning" in information security for improving our detection and response capabilities. In a nutshell, machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

This is really what we are asking of security analytics platforms today: Collect large and disparate internal datasets, comb through that data looking for patterns and creating correlations, and provide guidance on anomalies and the potential threats we face. At the same time, many security teams have started incorporating threat intelligence into their datasets to gain perspective on what is being seen by others. Regardless of the individual datasets in use, analytics should provide us with predictive capabilities beyond what we've traditionally had with today's event management technology. Ideally, we may be able to develop baselines that allow us to detect new and unusual patterns for attacks both known and unknown.

Use Cases and Examples of Analytics in Action

In a blog post, Gartner Research Vice President Anton Chuvakin suggests that security analytics tools could easily provide value in expanding our capabilities to perform network forensic analysis.[6] Here, we will look at four main use cases for security analytics today.

The first is helping security teams reduce the mean time to detect (MTTD) an attacker in the organization to speed IR. For example, analytics could potentially help with the detection of web-based malware delivery, installation and control associated with a phishing attack. In this case, analytics can help answer the following questions:

- Is this the first time this person has received email from the recipient?
- Is the website link in the email on a known list of bad websites?
- Are there changes to any host configuration settings or files closely tied to a website visit?
- Are there DNS requests to known bad sites or are the IP addresses of the DNS URL request and responses the same or different?
- Have there been any unusual examples of port and protocol usage?
- Are there any similarities in the size of transmissions, particularly across multiple time periods?

Security monitoring is a second use case for security analytics. Analytics could help security operations teams understand the impact of individual assets on the business. Using this information, analysts could then prioritize security-monitoring efforts accordingly. Let's take the example of a widget manufacturer with network-connected machines on the plant floor. One machine's abnormal behavior could have a negative downstream impact on overall widget production. Here, security analytics could help answer the following questions:

- Are there network assets that should be monitored more frequently because of their importance to the business?
- What is the impact on other network assets if this one is compromised?
- What is the impact on other network assets if this one cannot authenticate to the network?
- Should we monitor some assets more frequently because of the amount of use they get?

[6] http://blogs.gartner.com/anton-chuvakin/2015/01/12/security-analytics-finally-emerging-for-real

In the third use case, analytics can be used to assist in configuration management activities, allowing IT to optimize devices across the network. Analytics could help provide information on the known relationship between systems for better network documentation. Security analytics could help answer the following questions:

- What is really happening on the network versus what we think should happen?
- What system may be operating outside of policy that wouldn't generate a security alert?
- Based on current system workloads, can anything be virtualized?
- What impact would an outage or downtime of a particular system have on the business?
- Are there any assets that can be completely decommissioned?

The fourth and final use case for security analytics benefits the organization's management. Security analytics can provide a data-driven assessment to guide executive decision making. Armed with this data, CxOs and board members can better understand the organization and its risk so they can address the following issues:

- What is our overall risk posture?
- What are our high-value targets?
- What are the risks if our high-value targets are compromised?
- What are the most cost-effective ways of reducing risks?

Clearly, analytics could help in more complicated and advanced use cases for threats against our organizations today.

Conclusion

According to Verizon's 2014 Data Breach Investigations Report, 66% of organizations have been breached for months or longer before they realize it.[7] This problem continues to worsen, and even though there are indicators of compromise in many of these cases, most IR teams are too overwhelmed to see these or react quickly when they do. With the sophistication level of attacks on the rise, it's vital that new models of IR (along with the tools to get the job done) advance as well. Leveraging more advanced data analytics platforms to take in more and different types of data, focusing on increased network threat visibility, and automating detection and response actions may help security teams now and in the future as they evolve to meet these challenges.

[7] www.verizonenterprise.com/DBIR/2014/reports/rp_Verizon-DBIR-2014_en_xg.pdf

About the Author

Dave Shackleford, a SANS analyst, instructor, course author, GIAC technical director and member of the board of directors for the SANS Technology Institute, is the founder and principal consultant with Voodoo Security. He has consulted with hundreds of organizations in the areas of security, regulatory compliance, and network architecture and engineering. A VMware vExpert, Dave has extensive experience designing and configuring secure virtualized infrastructures. He previously worked as chief security officer for Configuresoft and CTO for the Center for Internet Security. Dave currently helps lead the Atlanta chapter of the Cloud Security Alliance.

Sponsor

SANS would like to thank this paper's sponsor:

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.108130_S150442.0216