SENSAGE
a KEYW company

Defending Against APTs: Do You Have The Complete Picture?

Executive Summary

Advanced persistent threats (APTs) present a large and growing danger to the information security of government and corporate enterprises.

While numerous approaches are being employed to identify, halt and remediate APTs, most lack a critical element: the ability to analyze large amounts of security-related data over a long period of time. Such long-term analysis can identify patterns that will lead to faster and more accurate assessment of the source of APTs, and even to the prevention of some exfiltration attempts.

KEYW is pioneering a new approach to combatting APTs, combining Sensage technology and policy-based, automated analytics that taps the rich potential of event data.

The Sensage Event Data Warehouse collects, stores and organizes all security-related data and applies powerful analytic tools to identify patterns related to real or potential security threats. The result can be a sharp drop in the time needed to identify the source of a threat and a reduced chance of data loss.

What Are APTs?

Advanced persistent threats are sophisticated, covert attacks intended to surreptitiously steal valuable data from targeted corporations and government agencies.

Their relentless, persistent intrusions typically target key users within organizations to gain access to trade secrets, intellectual property, customer or patient data, financial records, state and military secrets, computer source code, and any other valuable information. No one—from government agencies to start-ups—is immune today. For example, a classic attack vector may rely on distributing phishing emails to employees of a banking company in order to infect workstations with malware. An unsuspecting employee is tricked into clicking on a URL which invokes what is called a “Man-In-the-Browser” attack, in which the user’s browser is redirected to a site that performs a battery of attacks to load different exploits and payloads. Once the hacker has successfully infected a machine, the hacker will attempt to gain access to banking credentials.

How Are APTs Identified?

To guard against APTs, companies and agencies deploy a range of solutions: anti-virus products, intrusion-protection systems, next-generation firewalls, sandboxing solutions, real-time event correlation; the list goes on. While such a layered, multi-point approach seems sensible, breaches still occur. Verizon’s 2013 Data Breach Investigations Report notes that more than 70% of security breaches are detected by third parties who operate outside of the enterprise, such as customers or law enforcement. So why, despite all the tools deployed, are so few threats identified in time to do something about them?

One key reason is that APTs are engineered to lurk surreptitiously beneath the thresholds set in security devices to send an alert. In retrospect, there are often many telltale signs of an APT such as suspicious e-mails, anomalous traffic or unauthorized requests for data transfer. But these signs are often overlooked, either because they are needles in a haystack (one-time events amid thousands of concurrent events each day) or because IT staffs lack tools to put them in the context of a larger picture. By the time an APT is discovered, significant damage most likely has already occurred. In fact, most APTs are discovered through their significant impacts (after the damage is done) rather than through their subtle behaviors (before the damage is done).

APTs evolve slowly and incrementally, so an effective detection strategy must accommodate this reality.

Most anti-intrusion tools offer a snapshot in time. So if the threat doesn’t leave a sign at the exact moment of monitoring, then detection is avoided. It is similar to a lighthouse, whose searchlight rotates constantly but only illuminates a given point a fraction of the time.

In order to detect these slow and methodical attacks, security teams must look deeper into their data, aggregating and correlating the artifacts across many data sources over extended periods of time to detect compromises that otherwise would simply go unnoticed.

APT Warning Signs

Let's examine four stages in the lifecycle of an APT and selected types of warning signs that might typically be found at each stage. All of these warning signs are activity that can be discovered when enough event data is collected, stored and analyzed properly.

Compromise: This is the first stage of an APT, where the attacker seeks an entry point into the target network. Telltale signs at this stage include:

- **Abnormal process, service or path**: These might include attempts to run unauthorized applications, or a process or service running from an abnormal path.
- **Attempts to hide activity**: For example, attempts to stop an audit function, or to clear security, application and event logs. Excessive logging or more frequent rollover also might be detected.
- Unusual activity: Examples might include downloads of abnormally long file names, or .zip files from abnormal IP addresses.
- **Abnormal connectivity**: For example, connections to unfamiliar IP addresses or URLs at unusual times; repeated connections to an IP address or URL at regular intervals; connections to IP addresses in unusual geographies; or a number of accessed URLs by a user that exceeds what's possible for a human to perform.

Escalation: Once an attacker has gained a toehold, there will be a new set of signs that can be used to identify and stem the attack. These include:

- **Privilege escalation**: An attack may try a number of tactics to gain access to higher levels of authorization. Telltale signs include administrator log-ons from an unusual host; administrator log-ons at unusual times, the addition of a user to a privileged group, alteration of a folder owner, or rapid creation and destruction of user accounts.
- **Failed log-ons**: Obviously suspicious activity might include repeated failed administrator log-on attempts, or an account lockout for privileged accounts.
- **Attempts to run applications**: Warning signs here could include abnormal command line activity; new service installations; new scheduled tasks; or abnormal patterns of users performing local installs.
- **Abnormal file access**: One example would be attempts to access restricted information using security overrides coming from unusual sources.

Lateral Movement: This is the stage at which an attacker's actions may leave an increasing number of artifacts, as they try to navigate through the network to reach the sensitive information. Critical categories of activity include:

- **Abnormal log-ons**: These can include a range of suspicious activity, such as abnormal user ID and target host combinations; abnormal user ID and target host log-on times; and abnormal source-host and target-host combinations.
- **Brute force log-ons**: Examples include unusual numbers of log-on attempts over a period of time, or from multiple hosts, or log-on attempts to multiple hosts.
- **Failed log-ons**: Abnormally high number of failed log-on attempts due to incorrect user IDs or passwords.
- **Unusual activity**: Suspicious patterns can be found in dimensions such as host access to target, host access to application, host-to-host connections, the time of connections for a given user or host, and unapproved software installation.

Exfiltration: This final stage of an APT is one you hope you don't reach. But even if data is being extracted, such activity can leave numerous signals that can help stem the attack and mitigate the damage. These signals include the following:

- **Unusual activity**: There is a wide range of factors where unusual trends can be spotted, among them:
  - Abnormal time and/or volume of file access by a given user
  - Abnormal time and/or volume for file upload or download
  - Unusual volume of screen captures
  - Large numbers of files sent to unknown IP addresses or geographies
  - Abnormal printer volume within a defined period
  - Abnormal printing times

- **Abnormal connectivity:**
  - Connectivity to unfamiliar IP addresses or URLs
  - Repeated connectivity to an IP address or URL at regular intervals
  - A workstation performing server functions
  - Abnormal volume of network activity

Not all the factors listed above will necessarily produce evidence of an APT. They are cited because they are real-life examples of the types of information available in event logs that – when properly gathered, managed and analyzed – can be of great value in spotting, stopping or mitigating an APT.

Why Do Some APTs Escape Detection?

While APTs progress through a lifecycle described above, the entry point for an APT is rarely the ultimate target. Most commonly, an intruder needs to find a vulnerable entry point; a user who may engage in risky behavior, such as accessing large numbers of websites outside the enterprise, or carelessly clicking on links in e-mails that should raise suspicions.

Once inside the enterprise, an intruder must navigate laterally through the network in order to reach his ultimate target. At each stage of the process, an APT will leave telltale signs. Why aren't these signs picked up? There are several possible reasons.

One is siloed processes. Security teams tend to focus on their division or area of specialization, while successful attackers cross all boundaries without regard to domains, IT layers or organizational structures. Another issue may be a lack of personnel trained or experienced in detecting intrusions.

And then there is a set of related issues that can be summarized in two words: insufficient data. Many organizations are simply not collecting and analyzing easily available data that could give them greater, earlier insight into the source and progress of APTs.

Where is this data, and why isn't it being collected?

The sources of data lie in the event logs of all the equipment connected to a network, such as servers, routers, firewalls, and intrusion-detection systems. The logs of these systems contain the telltale signs of an APT or other intrusion. Any event with a time stamp – a system log-on or file transfer, for instance – can potentially turn out to be evidence of an attack. But unless these events are stored, saved and analyzed, the clues they contain will never be seen.

The barriers to collecting and analyzing this data are familiar: "We generate way too much data to store, and even if we did store it, we don't have the people or tools to make any sense of it."

Before answering those objections, here is more context to the limitations of solutions being used to address APTs barriers to using log-data analysis to reduce damage from APTs.

Limitations of Current Approaches to APTs

What's holding back a comprehensive solution to identifying APTs? Many issues are, including:

- "Stovepiped" security products that don't correlate information or share policies
- The fear of false alerts, or uncertainty about how to deal with them
- No enterprise-wide reporting or analysis
- A shortage of experts who have time to bridge the gaps in existing systems
- No "metrics-minded" culture committed to measurement, analysis and continuous improvement

Another essential missing element is a long-term perspective on the nature of APTs. By their very nature, APTs occur over varying periods of time, sometimes months or even years. According to the Verizon study, an attacker can successfully compromise a system in as little as minutes or hours, but it can take days, weeks or months from the time of compromise to the point of data exfiltration. There are numerous points in the process from compromise to exfiltration that the attack could be identified and stopped.

It's no surprise that organizations feel they cannot get their hands around the problem of collecting and storing security data. For example, one investment bank with 5,000 employees says it captures 25GB of security-related data every day. And in the case of an even larger enterprise, the amount of data can be truly staggering. For instance, a single server can generate 1,000 events per hour, while each employee

can generate some 800 events a day when in the office and perhaps 200 or more events when using a mobile device. All told, an enterprise IT infrastructure can generate in the neighborhood of 6 billion events per day. Multiplied by a year or two, the storage requirements can be in the petabytes.

The reason to collect and save so much data is not so you can find the one piece of information that will solve the case. It's so you can conduct an analysis across time and systems that will reveal patterns that just cannot be identified by point-in-time approaches.

So, it is very important to capture data from all relevant sources, and store that data for more than the normal 3-6 months. But it isn't sufficient to just have the data. You have to be able to conduct rapid, complex analysis of the data such that an investigation can be completed in time to interdict the threat. This means using tools that are specifically built to the task. The key here is workflow automation. The process of collecting, structuring and analyzing huge amounts of data is beyond the ability of humans alone. Workflow automation – policy-based methods carried out automatically across all relevant sources of data – can overcome the objections of cost and complexity, and produce results that will lessen the damage of an APT.

Most security systems and tools in place today examine a point in time, looking for instances where, for example, large amounts of data are transferred in an unusual pattern. But as experience shows, most APTs are successful over a long period time and involve small, incremental intrusions or extractions of data. There seldom is a single piece of evidence that will break open a case. Much more often, the solution comes from spotting a pattern of events. And patterns are only detected when the investigator has enough data points. Possible patterns could entail devices emitting beacon signals or a radical change in a user or machine profile (for example, a production server acting as a workstation desktop in terms of network patterns).

It's like looking at a great painting in an art museum. If you put your eye an inch away from the canvas, you will see a dab of color or a stroke of a brush. But step back a few feet – taking in all the dabs and strokes – and the subject becomes evident; the picture becomes clear.

Confronting APTs requires a similar approach: looking for the big picture.

Implementing a Comprehensive Solution

What are the key elements of a comprehensive solution to detecting APTs? They include the following:

- **Collecting data from all sources** – as noted earlier, the more sources of information tapped, the greater the chance of finding relevant information.
- **Retaining native logs** – the temptation is strong to reduce storage demands through normalization, data filtering or other techniques, but these actions can limit analytic capabilities and in the case of regulatory compliance can be counter to requirements.
- **Centralizing event data management** – when data is pulled to a central location, trained security personnel can more easily look across multiple sources in search of trends, anomalies or other valuable insights.
- **Providing flexible analytic capabilities** – to perform complex analysis on event data, trained security personnel need tools that are tailor-made to handle the unique challenges posed by high volumes of event data and APTs.
- **Adopting a "metric-minded' attitude throughout the enterprise** – setting goals and measurement criteria are essential steps in a plan to improve organizational efficiency and performance.

Understanding Event Data

We have seen that detecting and combating APTs requires a concerted effort to collect and analyze all relevant sources of information that may contain clues to the source and progress of a threat. And because the volumes of information are so large, the task is as much a data-management issue as a security issue. In this context it is critical to understand the unique characteristics of event data and how it is best handled.

Among the key characteristics of event data are the following:

- **Non-transactional**: Once event data is stored, it will never be updated. In fact, for

compliance purposes, altering and deleting event data should be strictly prohibited.

- **Time-based**: Event data is a collection of data about a particular event, at a specific point in time. This means that every event will have a time stamp associated with it.
- **Repetitiveness**: Event data is generally highly repetitive (e.g., company employees logging on and off), so finding anomalies involves the analysis of high volumes of repetitive events.
- **Near real-time**: Event data is created in real time and must be loaded at least as fast as it is created. Although all event data does not need to be available in real time, load rate must keep pace with creation rate in the long run.

The foundation of most data management for the past quarter-century has been the relational database management system (RDBMS). As Security Information Management (SIM) vendors discovered the need to manage event data beyond real-time requirements, they followed the well-established practice of incorporating event data into the relational model and using RDBMS technology to store and analyze it.

However, as the demand to manage greater volumes of event data emerged, the limitations of RDBMSs became apparent. There are many reasons for this,[1] but the quick summary is: *event data and RDBMS technology are, at best, poorly matched*. The mismatches can be roughly categorized as either a) RDBMS overhead not required to manage event data, or b) lack of technology to support the unique requirements of event data management.

Pioneering a New Approach

With the acquisition of Sensage technologies, KEYW has pioneered a new approach to security management by delivering a high-performance, scalable means for organizations to centrally aggregate, cost-effectively store, dynamically monitor and efficiently analyze massive volumes of event log data over long periods of time while retaining the original source data.

The Sensage Event Data Warehouse approach eliminates the unnecessary overhead imposed by standard RDBMS technology, and materially increases the performance and capacity to manage massive volumes of event data.

The Sensage solution features:

- Server clustering (massively parallel processing architecture)
- Data compression
- A non-transactional model
- Seamless access to online and archived data in a single query

This solution provides significant benefits to customers needing advanced event data management.

- High-Performance Analysis: Perform analysis in minutes or hours, where RDBMS searches often take hours or days.
- High-Volume Loading: Data loading keeps pace with enterprise-wide event collection for gigabit class networks, with no degradation based on the volume of data stored.
- High-Volume, Low-Cost Storage: Using low-cost Linux servers to store highly compressed data. No expensive RDBMS licenses are required. Servers are more efficiently utilized due to the elimination of RDBMS overhead.
- Low Cost of Ownership: No database administration resources. Data organization is simple and self-tuning.
- Incremental Scalability: Additional servers can be scaled incrementally to provide increased capacity and throughput to match business growth.
- High Availability: Built-in redundancy allows continued operation even with a server failure.
- Data Protection: Event data is protected against any modifications by outside sources. Data redundancy protects against loss of data in the event of component failure.
- Data Archiving: Allows for archiving data from faster direct storage to slower, cheaper storage while still keeping data online and available for queries.

[1] For a detailed comparison of event data and RDBMS technology, see the KEYW white paper entitled "*The Event Analysis and Retention Dilemma,*" available at http://www.sensage.com/sites/default/files/sensage_event_analysis_dilemma.pdf

Advanced APT Detection

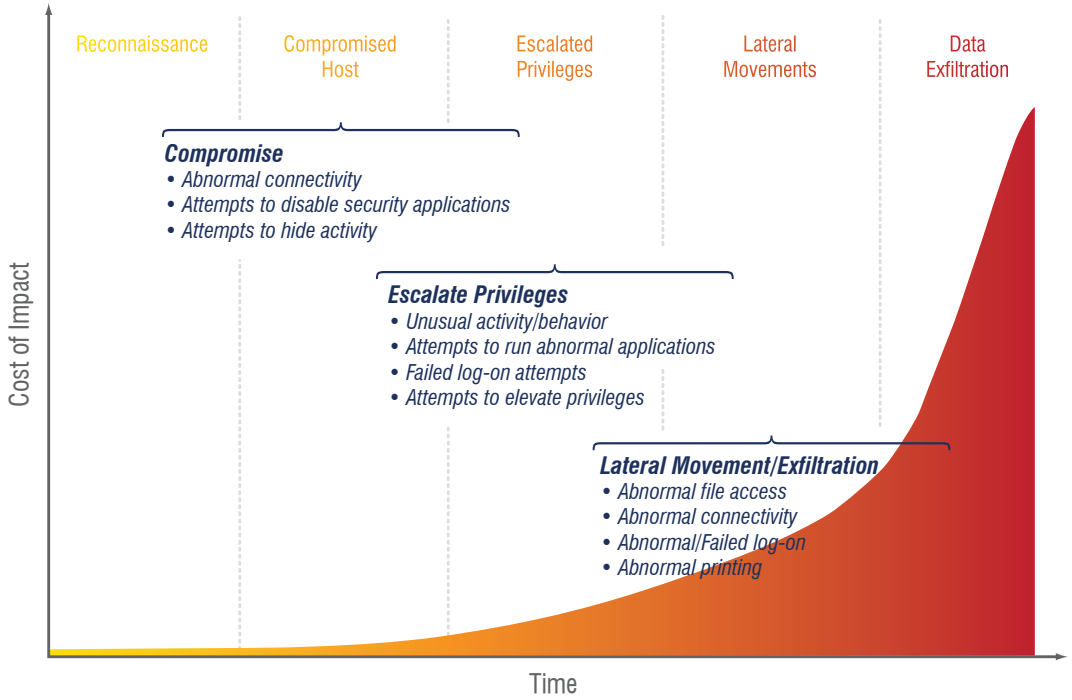
How do customers use this information to block an APT or mitigate its damage? Let's look at two possible scenarios.

The first envisions an APT in its early stages, where the attacker is looking for a vulnerable entry point by trying to get an end-user in the targeted enterprise to click on a link, or visit a URL, or download a file.

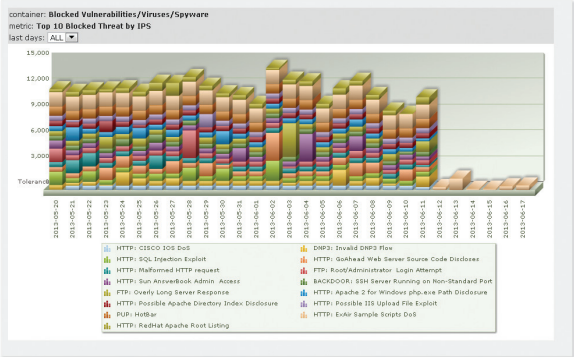
In this situation, it is highly unlikely that any single instance of an attempted entry will set off any alarms inside the enterprise's security operation. The importance of any one event becomes clear only when it is viewed in context – and context requires two things: a view of all conceivably relevant activity over a significant period of time; and the ability to correlate large volumes of data in order to identify meaningful patterns, trends or anomalies.

With the Sensage solution deployed, an enterprise would set thresholds for the various types of factors described above (whichever ones they felt were most useful), then monitor dashboards to spot unusual patterns that might indicate an APT in progress. The result would be the best possible outcome: spotting an APT before it has made much progress.

The second scenario envisions an APT that has been underway for some time and has already succeeded in extracting valuable data. In this case, the specific user and file would be known, so the system would be used in a very different way – examining all activity related to that user in the time before and during exfiltration, and looking for activity that would identify an unusual or suspicious source of access or destination of file transfer. The key here again is context – having enough event data over a long enough period that analysis can be performed to identify abnormalities.



Example of Analytics Across APT Lifecycle



Example of Security Analytics Dashboard

The end result in this scenario would be a two-fold improvement: a clear path to identifying the source of the APT, and a stop to further exfiltration from the targeted user.

The Approach of a Metrics-Minded Organization

None of the steps outlined in this paper can be implemented instantly across a large enterprise. And the application of high-volume, rapid-response analytics to event data requires careful planning and implementation. At the same time, it is important for enterprises concerned about APTs to start somewhere, anywhere.

At first glance, this may seem like a cavalier recommendation, but it's not. An organization can start small by, for example, adding a certain category of devices to the list from which event logs are regularly uploaded. Another place to start would be to look at upload/download traffic to spot patterns of behavior by department or function or geography and then, over time, to learn how to spot anomalies.

The point is to get a start on tapping the value trapped inside mountains of data that may have seemed impossible to tap.

There also are several aspects of high-volume event data analytics that are much closer to organizational management than they are to technology. One is the need to cut across organizational boundaries to ensure that all relevant data is being collected, centrally stored and effectively managed by experts.

Another is the need to find pilot projects that can be used as learning experiences and that can be expanded elsewhere in the organization.

It also is important to, as early as possible, gather a cross-functional team of metrics-minded individuals to build out the plan around collecting, analyzing, reporting, interpreting and responding to security intelligence.

There is much to learn every day about security intelligence. To evolve, organizations must innovate, learn from what works and prune what doesn't, and adopt new disciplines around metric management and continuous improvement.

Conclusion: Employing a Comprehensive Analytics Solution

In the course of this paper, we have noted several important points about effectively combatting APTs:

- Detecting APTs and minimizing their damage is as much a matter of data management as it is an issue of security policy.
- Collecting event data from the broadest range of sources, and over as long a time frame as possible, will increase the chances of uncovering meaningful information and insights.
- Effectively analyzing event data in the context of APTs is a complex task in which human intervention must be supplemented by advanced technologies specifically tailored to the unique characteristics of high-volume event data.
- Organizations must apply techniques such as metrics management, continuous improvement and cross-functional teams to ensure that technology solutions are properly implemented.
- Combatting emerging threats requires a high-performance, scalable solution for organizations to centrally aggregate, cost-effectively store, dynamically monitor and efficiently analyze massive volumes of events over long periods of time, while retaining the complete original source data. This empowers organizations to respond to business threats, conduct thorough investigations, and fortify broad audit compliance processes.

About KEYW

KEYW provides agile cyber superiority, cyber-security, and geospatial intelligence solutions for intelligence, defense, and commercial customers. We create our solutions by combining our services, products and expertise with hardware, software, and proprietary technology to meet our customers' requirements.

KEYW's subsidiary, Sensage, delivers a platform, analytics and services that protect organizations around the world.

The Sensage solution combines powerful data warehousing with business intelligence to address the most advanced use cases in Security Information and Event Management, log management, and Call Detail Record (CDR) retention and retrieval.

KEYW's commercial products division is also the creator of Project G, dedicated to automating and advancing the cyber awareness and cyber defense of the most sophisticated enterprises and government agencies.

SENSAGE
a KEYW company

Sensage, Inc.
2800 Campus Drive
Suite 150
San Mateo
CA 94403
www.sensage.com