Splunk, Big Data and the Future of Security

WHITE PAPER

The Changing Nature of Security Threats

Current IT security tools and mindsets are no longer adequate to meet the scope and complexity of today's threats. Internet security has evolved over the last ten years but advanced persistent threats and the sophistication of the malware have fundamentally changed the way security teams must think about these new threats and the tools used for detective controls.

The Evolution of the IT Security Infrastructure

Over the years, security teams have moved from simply protecting the IT infrastructure perimeter with traditional defenses (firewalls), anti-virus and intrusion detection systems (IDS). They have employed sophisticated IDS to prevent attacks and we watched as these evolved into intrusion prevention systems (IPS). Even with this level of protection, it was not possible to track all the events that security teams saw from the security architecture and many of the IDS events were false-positives. Security event management systems (SEM), which correlate data from these systems, evolved to reduce the workload using correlation rules to reduce false-positives, alert on possible threats and provide some visualizations and canned reports that reflect some security metrics.

"…There is now a fairly accepted consensus that the technology (AV, IPS, and Firewalls) aiming to keep malicious actors at bay isn't completely successful."
Chris Silva,
IANS Trend Report 2011[1]

As use of the web and email for business exploded, over time (and in no particular order) web proxy, email security, vulnerability assessment, database security were added to monitor possible attacks, understand what systems were unpatched and vulnerable, and prevent the spread of viruses and malware on systems. This deluge of data proved too much for the SEM and the security Information management system (SIM) was introduced to reduce the workload on the SEM by collecting, normalizing and storing data from the security architecture sending a subset of this data to the SEM.

This combination SIM/SEM technology and any required collection agents became referred to as the security information and event management system (SIEM) and continued the data-reduction or data-distilling strategy first introduced by the SEM. This architecture is usually visually represented as the funnel: Information is gathered from a wide variety of signature-based systems and operating systems and a sub-set of the collected data is sent to the SEM and correlated using a rule-based system focused on known threats such as classic perimeter-based attacks over open ports or brute force attacks.

As many businesses realized that a majority of data exfiltration happened when employees were not happy or left the company, data loss prevention (DLP) was added. Also during this time, due to the business scandals of the late '90s and early 2000s, access controls and system monitoring was mandated and regulatory compliance became a cost of doing business. Funding shifted from security to compliance and vendors rushed to create more canned reports "chasing the money."

While the SIEM "funnel" or data reduction approach was helpful in reducing the amount of data the team had to analyze to recognize attacks, attackers exploited the fact that the security team had only a subset of all data to work with due to data scalability issues. In addition, silos remained between IT operations, IT security and development teams in large IT organizations. Applications closest to the core business product or service mission were monitored by IT operations or applications teams or in some instances not at all. Developers created applications that either didn't log or logged in ways that were only understood by the developers and that were not standardized. This data was not able to be included in a SIEM view due to complexity or expense.

Not having access to application data meant that security teams had a blind spot for risks to the business as malware and viruses that took advantage of undisclosed or zero-day application layer, the lack of context for security incidents, and silos between teams that increased incident response times.

Exploiting Trust

Today, who we are, what we do, and who we know is available on LinkedIn, Facebook and other social networking sites. A simple search on Google can reveal our PowerPoint presentations and the conferences we've attended. Today attackers target the most vulnerable point in the network—exploiting human trust using spear phishing. The exposure to today's sophisticated advanced persistent threat (APT)-style attacks and the pervasive presence of malware is a harsh wake up call for many IT organizations. According to Kevin Mandia of Mandiant Inc., "…there are thousands of companies compromised—actively, right now,"[2] with highly sophisticated malware that has yet to be discovered deposited by persistent attackers.

An appropriate analogy of this new style of targeted attack would be the difference between a car thief that wanders around parking lots at a variety of locations looking for cars that may be unlocked with keys left behind the visor, versus a thief that wants your specific car. The latter type of thief decides to follow you around for weeks or months, watching your habits, learning where you live and getting to know who you talk to, where you go and who you know. He then introduces himself to you at a party as a co-worker of one of your best friends in another department and then asks to borrow your car.

The "data-thief" you linked to on Linkedin wasn't really a long forgotten coworker but someone you've never met who knows you very well. The 'data-thief' sends you an email with a PDF attachment called "organizational changes" that contains zero-day exploit code. With one click, your system is "owned" and your email, private information, your website content or other company intellectual property data is potentially compromised.

What's worse is if you happen to have admin rights to your system and stored credentials on your computer to other critical applications on the network. The deposited malware can change settings on your Windows group policy object and change the Windows DLL startup order, guaranteeing that the malware will start at boot time and stay persistent. It can also replace or switch on a little used networking service to spread the malware to other systems. This type of malware may also stay resident but dormant on several other systems as a backup in case the first compromised system is found. Each of these steps may take place several days or weeks apart.

While the threat landscape has changed, the mindset of the security team has remained focused on a conventional approach to security dealing with known threats, which includes:

- Protecting all information assets
- Maintaining controls that are primarily signature-based and preventative
- Paying the most attention to the egress points in the network
- Collecting logs for compliance and post incident forensics
- Getting smart on the current malware threats
- Finding and removing malware and infections
- Preserving the network: a successful outcome defined as no attackers get in.[3]

Our Overdependence on SIEM

With hundreds of variants to social engineering-based scenarios, many ways to compromise a system, zero-day vulnerabilities in applications and operating systems and thousands of malware variants, what are the odds that one of your 200+ SIEM rules will fire and alert you to a problem? "Traditional security information event management (SIEM) systems typically don't detect a relentless targeted attack designed to avoid raising any red flags: they're tuned to catch unusual activity, not stealthy attacks that hide behind legitimate user credentials or normal traffic."[4]

"'Can I be compromised?' is no longer the right question to be asking."
Jason Rebholz,
Mandiant, 2011

If we are lucky, and a piece of our security infrastructure finds a compromised host with persistent malware, we declare victory, clean or re-image the affected host and are satisfied with our success. Most SIEMs act in a serial fashion—the security event detected at the SIEM level is presented to us as the "end-of-the-story" supported by evidence from signature based systems. With this approach, the SIEM will not tell us that the security event that was alerted on was really compromised host number 117 and that the attack really started with a different host over a year ago. The information to support this possibility has long since passed into the history books. The SIEM is not architected in a way that will let you easily re-examine an old attack from a year ago with new information. In July of 2011, McAfee announced that it had found active malware deposited on systems in 2006. Re-analyzing five years of log data is outside the reach of traditional SIEM and more suited for a big-data solution.

"Most security software prevents or detects a high number of known threats. While you need to have these capabilities in order to detect the botnets and viruses that cause interruptions to your organization's daily operations, they miss the advanced threats being used to target your most sensitive information. Additionally, much of this software — although not all of it — is designed to limit your control over what threats are detected, how the detection occurs and when you remediate."[5]

When the attacker sees that one of their "owned" hosts is no longer available, the attacker does their own post mortem on the discovered host and shifts tactics to get more firmly embedded on hosts in the network. New command and control instructions are sent to other dormant malware on hosts on the network and the persistent attack continues.

The Implications of New Modus Operandi

The attacker wants their malware to behave like a normal application so that log data doesn't cause any security systems alarms. If an attacker knows you well and establishes trust with you over a real or imagined long period of time, then there's a good chance you will follow a request of an attacker telling you to check out a website, click on an attachment or give to a charity that needs help now (and pay with a credit card).

Because of the current mindset, security professionals always find these words hard to admit: completely preventing these kinds of attacks is futile. However, in the current environment, the security team can work to minimize the damage by quickly spotting new user or machine patterns of behavior in very large data sets that may be worth investigating.

Thinking Like an Attacker to Find Unknown Threats

For years security professionals have been in reactive mode when a security threat was recognized. As a threat was recognized, the team was dispatched to deal with what was detected. If it was determined that the issue was with a user's behavior, the team would try to educate the user about the importance of data protection and move on to the next problem. But for each incident the team was able to address, there were others that went undetected. The current security tool sets by design (including SIEM) have us thinking like a victim and not like an attacker. To understand the full scope of an attack we need to start thinking like an attacker.

Thinking like an attacker means you must understand:

- Asset and data criticality
- Location of the most important company data assets
- Ways your systems and data can be accessed
- Means by which malware can be spread in the organization
- Means by which malware can be made persistent
- Who might be the most "attractive" victims, what level they are in the enterprise and what data they might have access to

- Knowing what would be considered unusual accesses to important data based on time, frequency or location
- Discovery of a single compromised host should not end the security investigation

Mandiant's "Anatomy of a Hack"

Understanding the attacker means monitoring large data sets of normal user activity data looking for patterns of activity that are not normal in context of time, place, or appropriateness. This has given rise to a new role on the security team called the security intelligence analyst.[6] These individuals:

- Take the "actor view" to understand the identities, goals and methods of potential adversaries
- Work with management, lines of business and operations personnel—this knowledge makes them aware of the threats posed by persistent adversaries
- Assess actions and determine if a pattern of threatening behavior is emerging
- Map and visualize threat behavior patterns against big-data sets of normal IT activities with analytics

"The core of the most effective [APT] response appears to be a new breed of security analytics that help quickly detect anomalous patterns—basically power tools in the hands of a new and important sub-category of data scientists: the security analytics expert." [7]

APT style attacks have de-positioned SIEM moving it from a solution to simply a tool for monitoring mainly known threats.

"Because adversaries are intelligent, well funded and patient, they can afford to take weeks to probe their targets and months to plant malware inside the organizations in order to exfiltrate data."
Chris Silva,
IANS Trend Report 2011 [8]

Dealing with Unknown Threats Using Big Data and Analytics

Just as businesses use business intelligence solutions to monitor large amounts of customer data and watch for patterns that allow them to better understand customer behavior, security professionals need similar solutions for the infrastructure.

They need to monitor network, host, and application behaviors in a contextual way across IT data to understand the depth and breadth of persistent malware in the IT environment.

"There are some emerging use cases for information security which can only be handled with big data capabilities."
Neil MacDonald,
Gartner, April 12, 2011

Enter Splunk—The First Big Data System for Security

Splunk Enterprise is the engine for machine data. Splunk software enables enterprises to gain operational intelligence by monitoring, reporting and analyzing real-time machine data as well as terabytes of historical data located on-premise or in the cloud. With Splunk you can leverage an analytics command language to map and visualize any potential attack scenario against the business' most important data assets. These scenarios can be easily aligned with the business risk-based modus operandi of potential attackers. Automated searches can continuously monitor for abnormal patterns of behavior in host, network and application data. Combined with an understanding of where critical data is stored, who should have access to it, time based analysis of typical user behavior (i.e., how much mail is sent per day, normal data access times, physical access and normal host network behaviors), abnormal patterns can be detected. Adaptive monitoring of the active phase of the attack over time presents opportunities to detect abnormal behaviors on hosts and networks.

"Splunk has approximately 3,000 customers, the vast majority of which are using Splunk to solve big data problems—providing operational intelligence to make machine data accessible, usable, and valuable.
Frank Sparacino,
First Analysis, September 15, 2011

Specific sets of automated Splunk searches of normal user activities may comprise multiple scenarios. A single search can trigger several other searches in a decision tree fashion and could confirm the existence and spread of malware. An anomalous behavior detected can be analyzed along with other time-sequenced IT data and changes to host configuration files. Old security events can and should be reviewed and reanalyzed on a regular basis as a means of preventing re-infestation and as a way of determining whether the first compromised victim found was the only one, the first one or victim 112.

An advanced approach—one that uses big-data and analytics for tracking and discovering malware left behind by persistent

attackers--can move the security team from a conventional approach to a more flexible and advanced approach more suitable for the newest security threats. Best practices of this new approach include:

- Focus efforts on most important data business assets
- Use detective controls linked to data analytics watching for behavioral outliers
- Seek, model and dissect attack patterns
- Develop deep understanding of attackers' modus operandi in context of the organization's key assets and IT environments
- Realize that attackers will sometimes get in, but are detected quickly and impact (risk) is minimized

The security team's creativity and imagination are supported through ad-hoc exploration of data and modeling of attacks based on business risk. This allows security teams to speculate on potential attack vectors in advance of any actual attack. Different critical alerts can be created to support a search that has found a particular issue on one host versus another.

Dealing with Known Threats

While the focus of this paper is a discussion of a new paradigm and a new way of thinking about detecting unknown threats, it is not meant to be a condemnation of practices around monitoring known threats. Script-kiddies and canned attack tools are still out there. There is a still a need to monitor security operational metrics for continuous improvement. System patching, IPS attacks, firewall accepts/denies, DNS logs, data loss prevention (DLP) systems, anti-virus and other endpoint security systems should still be monitored.

Splunk offers alternatives to users who would typically purchase a SIEM to monitor their security infrastructure only monitoring known threats. Splunk (as of this writing) offers over 30 security apps free of charge that can monitor specific security point problems. Splunk also offers the Splunk App for Enterprise Security, which monitors over 100 security metrics, provides over 160 reports, identity correlation and a complete set of the most important correlation searches to offer SIEM functionality. It includes incident workflows and supports drill-down into raw data as well as workflow actions to launch cross data-type views of incident data. Just as with the core Splunk product, real-time alerts can be generated.

Summary–Reaching for Security Intelligence

Finding anomalous patterns in massive data sets over time and in context for unknown threats is the key to detecting advanced persistent attackers and the malware they leave behind. SIEMs that are set up to monitor security infrastructure watching for known threats do not solve the APT problem and have security teams in constant cleanup mode thinking like the victim and not like the attacker. Only big-data solutions with strong analytics and visualization capabilities can provide insight into anomalous behavior.

Security teams need to start using their creativity to think about the modus operandi of the attacker and work with the business, assigning risk to data. Thinking like an attacker and modeling attacks that start with spear-phishing against the most important business assets aligns the security team with business objectives through prioritization of data assets and risk. This type of thinking is a valued skill.

Splunk is a security intelligence solution for monitoring large datasets and gives you the ability to tell the difference between humans interacting with IT systems and behaviors that may be caused by malware. Splunk can cover known threats via information from signature and rule-based systems but also can be used to monitor for unknown threats based on risk-based scenarios translated into Splunk's analytics language.

Free Download

Download Splunk for free. You'll get a Splunk Enterprise license for 60 days and you can index up to 500 megabytes of data per day. After 60 days, or anytime before then, you can convert to a perpetual Free license or purchase an Enterprise license by contacting sales@splunk.com.

1 IANS Trends: A Behavioral Approach To Threat Modeling, 2011

2 Report Details Hacks Targeting Google, Others, WIRED, Kim Zetter, February 3, 2010

3 When Advanced Persistent Threats go Mainstream, Security for Business Innovation Council, July 11, 2011

4 APT Shaping SIEM, Kelly Higgins, Security Dark Reading, 10/3/2011

5 MANDIANT M-Trends Report 2011, Mandiant Inc.

6 IANS Trends: A Behavioral Approach To Threat Modeling, 2011

7 When Big Data Met Security: Is The New Era Beginning? Chuck Hollis, VP – CTO, EMC Corporation, April 12, 2011 http://chucksblog.emc.com/chucks_blog/2011/08/when-big-data-met-security-is-the-new-era-beginning.html

8 IANS Trends: A Behavioral Approach To Threat Modeling, 2011

250 Brannan St, San Francisco, CA, 94107 info@splunk.com | sales@splunk.com 866-438-7758 | 415-848-8400 www.splunkbase.com