Challenges and Opportunities in Historical Data Analytics for Cyber Security

Dr. Edward G. Amoroso
Chief Executive Officer, TAG Cyber LLC
Mob: (201) 454 – 1854
eamoroso@tag-cyber.com

Abstract
Historical analysis of data for cyber threat is the most traditional means for security protection. Advances in virtualization, hunting techniques, and virtualized managed security services will effectively address many of the challenges found in historical security analysis of data in the past few years. (Analyst Note: This paper was inspired by discussions about threat analysis of historical data with Faizel Lakhani and Tony Thompson of SS8 in May 2016.)

Introduction
One of the most basic concepts of cyber security, tracing its roots back to Dorothy Denning's early work on the IDES model in the 1980's, is the idea that historical records contain security-related information. Every modern Security Information Event Management (SIEM) system, for example, was designed specifically to reach back in time to analyze behaviors and to enable security analysts to make forensic and predictive determinations about security.

The historical approach can be illustrated by first looking at a typical event trace of session activity in a target environment:

Figure 1. Typical Event Trace for Several Sessions
On first glance, the event trace shown here provides the analyst with the step-by-step progression of activity for each session. Correlation of events between sessions can be identified in an ad hoc visual manner, but the development of security analytic tools automates the task, presumably highlighting patterns not immediately obvious.

**Figure 2. Correlative Patterns Identified Across Several Sessions**

This theoretical goal of using historical analysis for cyber security is generally pretty easy to explain and illustrate. Furthermore, the marketplace is rich with offers from SIEM and log management vendors that have been indispensable in preventing and analyzing attacks.

That said, the sad reality is that many enterprise security analysts have discovered that serious technical, operational, and management challenges exist which have prevented historical analysis from reaching its full potential. These challenges include the following:

- **Collection** – The massively distributed complexity of the modern enterprise has made it difficult to deploy sufficient sensor coverage to allow for uninterrupted views of relevant activity. The result is an historical pattern with significant time gaps resulting from spotty sensor coverage.

- **Signatures** – The use of predefined patterns to detect a security event (as depicted with the A-B-C-D-E-F-G pattern in Figure 2 above) have been susceptible to variant techniques by bad guys to thwart the analysis (e.g., skipping step C in the pattern shown above). This complicates automated processing as in intrusion detection.

- **Analysts** – The practical reality is that identifying trained security analysts may not be feasible for many companies. Furthermore, expensive managed services may also be out of budgetary reach.

These are real challenges that have at least contributed to the massive number of enterprise and government breaches, such as the Office of Personnel Management (OPM) attack, that have occurred in the past few years. The good news, however, is that many advances in modern technology and enterprise architecture show excellent promise in easing these challenges and helping to establish historical analysis, sometimes referred to as *hunting*, as a more dependable and sustainable technique in the modern enterprise.
Collection via Centralized SDN

The essence of software defined networking (SDN) is a return to centralized control in modern networking to optimize the performance, and minimize the overhead, of moving traffic from one virtual workload to another. This provides immediate relief to the cyber security team trying to deploy data collection sensors across an enterprise. Rather than having to find natural chokepoints, increasingly rare in complex Intranet, SDN affords the opportunity to perform collection in one place.

Since SDN controllers are designed specifically to support extensible processing through the northbound application interface, security analytic support can be deployed specifically to reside on this interface. The result will be increased visibility, more holistic coverage, and greatly simplified deployment without the need for hardware (SDN is all-virtual).

Figure 3. Security Analytics in the SDN Controller

Product support for SDN in the security analytics marketplace is currently not the highest priority, but as CISO teams come to recognize the benefit, this is likely to change quickly. With such deployment the potential increases that a team can actually reach back in time via a locally stored archive of historical records to identify threat indicators. This task will be done in the virtual data center and the SDN-powered WAN, LAN, and mobility infrastructure.

From Signatures to Hunting

The simplest means for historical analysis of cyber security involves the oft-maligned signature. Strictly speaking, a signature is a snapshot of some known pattern that an analyst should be on the lookout for in poring through historical (or current) event logs. The reason signatures have been such a problem is that once a signature has been established, the mechanized parsing of data for this pattern can be sidestepped by bad guy who understand what is being used.
For example, if a signature is established citing three failed login attempts within thirty seconds as a clear indication of a hacker, then the signature would look like this:

If login fails, then start a clock and watch for more
If two more fail within 30 sec, then send an alert
If two more don’t’ fail within 30 sec, then reset

The trivial means for getting around this signature is to develop a variant automated attack that introduces sufficient delay as follows:

Repeat
  Time = 0
  Hack login and quit if it succeeds;
  If it fails, try again, quitting if it succeeds;
  If it fails, pause until Time > 30 sec
Continue Again

This simple variant illustrates the simplicity with which historical patterns can be avoided by hackers who understand the automated signatures being used. Luckily, cyber security experts have improved their game significantly, and historical analysis can now provide much better evidence of real attacks. For example, in the case sited above, the mechanized attack would clearly avoid a programmed intrusion detection system, but the pattern would be obvious to a cyber analyst looking carefully through the data.

The process by which cyber security analysts examine historical data in conjunction with a competently designed analytic toolset is now known as hunting. The advantages of hunting over signature analysis are many. The most obvious is that employing human analyst hunters allows for insights, hunches, and human intelligence to play a role in detecting subtle issues. Cyber security history includes many cases where the human saw something unusual, which led to important breach detections.

In fact, this notion of “finding something unusual” has led to the use of behavioral profiling, machine-learning, and complex pattern analysis as the best ways to assist the hunter. In essence, both the human hunter and the automated analytic tools are working in concert to detect anything that appears out of the ordinary. Such management and analysis by exception has been a powerful tool for many years in industries such as telecommunications. Vendors and service providers who have previously implemented CALEA and other lawful telecommunications data collection, processing, and analysis are thus well suited to support the emerging hunting task.

It is therefore only natural that profiling and exception management would combine with hunting to give organizations the best possible means for detecting cyber attacks. This is good news for CISO teams concerned that historical data analysis will inherit the issues found in traditional intrusion detection.

Reducing Analyst Need via Virtualized MSS
If hunting is the new best practice for cyber security analytics, then the management question emerges: Where does one find all these analysts? Furthermore, is it reasonable to envision a future where every business has a team of real-time hunters watching for attacks to the corporation? This type of activity is clearly beyond the budgetary reach for most companies, and it is a distraction from the core mission of just about every type of organization. What is needed is a comfortable blend of sensor coverage, automated analytic support, and hunters available to dive into the historical data if something is truly amiss.

The best solution to this puzzle can be found in the evolution to virtualization in the data center and enterprise network. With virtualization comes a fundamental architectural change in how businesses operate – and it is this change that enables a solution to the analytics and hunting puzzle. Specifically, with virtualization comes simplicity. And with simplicity comes the potential to gain the best of all worlds because the historical analysis of a cloud workload might actually be automated.

Consider that virtualization begs the development of cloud workloads as the new unit of modern enterprise computing. This is a big change from the applications, servers, and endpoints that CISO teams try to protect using a vast perimeter. Simple cloud workloads allow for the possibility that a portion of the analytics can direct a machine-learned mitigation response. This is possible, again, because cloud workloads are simpler and more predictable than a large enterprise network.

Take, for example the development of a cloud workload deployed to a hypervisor environment in an organizational infrastructure, which might be public, private, or hybrid. If the workload is protected with a micro-segment, perhaps using simple utilities provided by cloud providers such as Amazon or with custom-designed protections using the Security Groups function in OpenStack, then the possibility exists that simple fixes can be automated into the micro-segment. Dynamically pushing a new rule into a micro-segment firewall is an example – and this automation is only possible because the workload is simple.

![Figure 4. Micro-Segment Automated Mitigation](image)
The key here is the following: Virtualization leads to workload simplicity; workload simplicity leads to simpler historical analysis of a more compact and focused data set; simpler historical analysis leads to more tractable automation of mitigation; and automated mitigation reduces the burden of finding competent hunters so that the ones who are available, can focus on the truly complex tasks that span workloads, networks, and even different organizations. Such human support can easily be integrated into future virtualized managed security service offerings.

Readers might believe there is some hand waving going on here – and the truth is that addressing shortages in analysts through automation is certainly not a new idea. But the fact remains that simplifying infrastructure is always a good idea, and if such simplicity can allow for more automated historical analysis of data with more automated mitigation, then this should be embraced.

References

About TAG Cyber LLC
Founded in 2016 by former AT&T Chief Security Officer, Dr. Edward G. Amoroso, TAG Cyber LLC is focused on bringing world-class, military-grade cyber security analysis, training, consulting, and media services to enterprise CISO teams around the world. The TAG Cyber Security Annual is its flagship annual publication, offered to enterprise security teams as an eBook or free PDF download from select cyber security vendors.