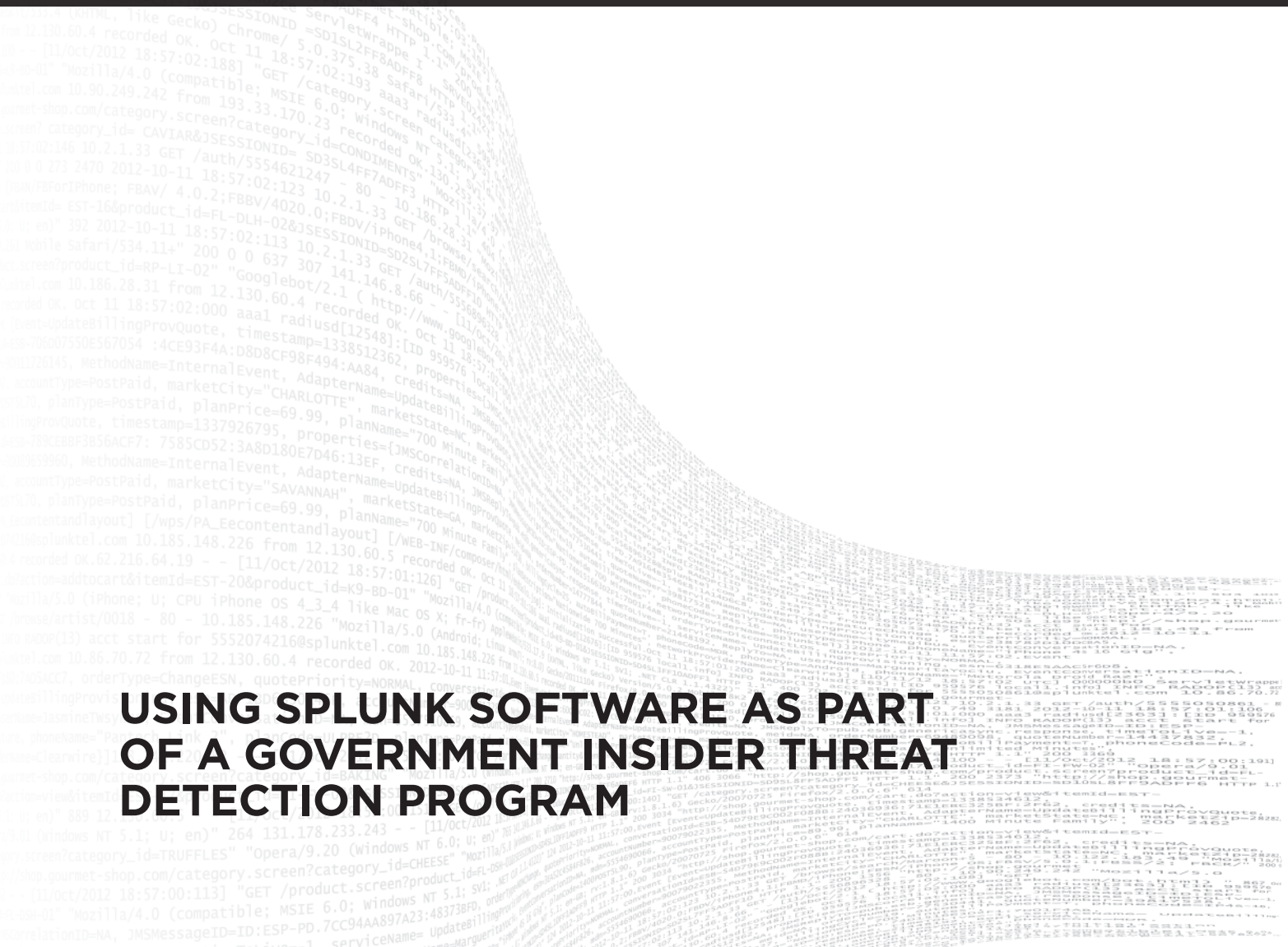
splunk>

WHITE PAPER

USING SPLUNK SOFTWARE AS PART OF A GOVERNMENT INSIDER THREAT DETECTION PROGRAM



Many public sector agencies, both civilian and those in the intelligence community, manage sensitive data that requires special handling, classification and heightened access monitoring for insider threats. Whether it's the Department of Energy keeping the latest U.S. oil and gas reserve data private until 10:30 a.m. EST on a Thursday, the release of drug efficacy test results by the FDA, or the intelligence community needing to safeguard intelligence and tradecraft, data access and dissemination needs to be monitored and controlled.

In the last four years, the U.S. federal government has faced a number of challenges from malicious insiders that have disclosed sensitive information. The Wikileaks incident and the Snowden releases have both compromised national security. In October 2011, after the Wikileaks incident, the president issued an executive order, creating a Senior Information Sharing and Safeguards Steering Committee made up of representatives across departments and agencies. The steering committee was charged with setting government-wide policy for the "deterrence, detection and mitigation" of insider threats.

New Priorities to Detect Insider Threats

One of the steering committee's initial priorities was to create minimum standards governing agencies' individual insider threat programs. In 2012, the steering committee established the initial priorities as:

1. Mitigating risks inherent to removable media
2. Reducing anonymity
3. Establishing insider threat programs
4. Improving access control
5. Enhancing enterprise audit

Following the tragic incident at the Washington Navy Yard in 2013, incident review led by the Office of Management and Budget (OMB) led to a review of the security clearance process, resulting in a recommendation of continuous evaluation of cleared personnel. These insider threat incidents and reviews led to a consolidated set of recommendations produced in a February 2014 White House memorandum from Lisa Monaco, Assistant to the President for Homeland Security and Counterterrorism. The memorandum provides detailed guidance and a timeline for department and agency implementation of the steering committee's recommendations, and singles out thirteen departments and agencies considered to be at risk for "high-impact unauthorized disclosures." The recommendations also include and apply to "private sector entities that are approved to handle classified information."

The Role of Machine Data in Insider Threat Investigations

As the definitive time-series record of machine-to-machine and human-to-machine behavior, data generated by applications and servers, inside or outside a department or agency, can play a role in determining anomalous employee behaviors. In some cases, the motivational context for those behaviors can signify an insider threat. The importance of this data is highlighted in specific data collection requirements within the February 2014 memorandum.

Data Collection Requirements of ICS 500-27

In a subsection on data management, the memorandum suggests starting with access to information contained on information sharing portals. More specifically, SharePoint and Wikis are identified as places where controls to accessing data may be weak. As a first step, the memorandum suggests conducting a continuous enterprise audit using audit data to evaluate and monitor access. A footnote in the memorandum indicates that the terms "enterprise audit" and "audit data" are defined in Intelligence Community Standard (ICS) 500-27.

A close examination of the ICS 500-27 (last updated June 2011) document reveals a highly prescriptive list of access activity data that should be monitored (see Appendix A). This list includes removable media events.

Data Collection Requirements of ICS 700-2

The Navy Yard shooter represents a different kind of insider threat. In this case, a contractor had access to firearms, held a high level clearance and also had a recent history of mental health issues with several documented police interactions. With over six million persons in the U.S. with high level clearances, and a cleared personnel review process that only occurs once every five years, it's easy to conceive how a single person might change over time and become an insider threat and a danger to their coworkers. Continuous evaluation of cleared personnel helps agencies address changes to individuals as evidenced in their data footprint as they interact with online systems.

ICS 700-2 contains information about the types of data that can and should be used directly as an indicator of a potential insider threat from a cleared individual, and also provides context for data collected as part of the ICS 500-27. The crux of ICS 700-2 is:

"As appropriate and in accordance with applicable law and policy, audit data collected pursuant to ICS 700-2 shall be analyzed in conjunction with other available data to support detection, mitigation or assessment of insider threats. Other data may include, but is not limited to:

a. Facility access information;
b. Foreign contact information;
c. Foreign travel information;
d. Personnel security information; and
e. Financial disclosure information."

This expands the scope of data that should be collected to include:

- **Physical access information** – Physical access to government facilities across the globe
- **Foreign contact information** – Data indicating contact with foreign nationals as may be evidenced in Outlook and other date book programs
- **Foreign travel records** – Data from public reservations systems
- **Personnel security information** – Gun background check information, police records that are publically accessible, hospital records and other data that may be collected about the individual in other non- or quasi-government systems (may include health data) for individuals with high level clearances
- **Financial disclosure data** – Motives for insider threat can often be seen in credit scores, business startups and other data that may indicate unusual financial activities

Depending on each employee's classification and employment agreement, the activities of the employee, as recorded in data, should be part of the big picture when assessing insider threats. These activities are used either as context for auditable IT events (ICS 500-27) or as standalone data (see Figure 1).

Figure 1 - The data required for insider threat context and analysis.

Additional Data to Consider

To fully understand insider threats, there's additional data that should be made available as needed, and is not mentioned in either ICS 500-27 or ICS 700-2 (a specific list of examples is shown in Appendix B).

When thinking about insider threats, it's rare that an individual has a career goal of becoming a threat. More often, a person undergoes life changes that

prompt him or her down that path. The digital fingerprints of these changes can often be seen in data that resides in human resource and time management systems. For example, an unusual number of home address changes in a specific amount of time, no vacation time usage in a two year period, habitually coming in early and leaving late, changes in marital status and poor performance reviews may indicate that additional data may need to be viewed in the context of unusual network and access activities, since they can provide a broader notion of potential insider threats.

All HR data should be considered sensitive employee information; however, with the right processes in place, it's possible that investigations beginning with anomalous IT systems or application activity are eventually referred to a group with the authority to review HR data for cleared personnel. This group can then consider the anomalous activity with motivational and behavioral clues.

Using Splunk Software to Assess Insider Threats

Splunk software provides a big data analytics platform that's capable of ingesting massive quantities of structured and unstructured time-series machine and log data without regard to format. Ad hoc or automated queries can be run across multiple petabytes and multiple data sources using the Search Processing Language (SPL™). Splunk software can also pull data from Hadoop data stores (HDFS), traditional databases and any other relevant data source via API to provide additional context for the data.

Splunk software can create correlations between disparate data sources and normalize different data types at search time. For example, a single user may be represented in log data as an employee number, but in an HR system as the employee's full name. Splunk software will help you normalize the ways the data is represented. This allows you to take full advantage of Splunk software's statistical analysis capabilities, which can help you monitor for activities that are statistical outliers at a variety of levels of standard deviation.

Dashboards, reports and alerts also support and address continuous evaluation of cleared personnel, allowing the user to increase or decrease the depth of an investigation. Splunk software makes the monitoring and investigation of insider threats a seamless process across time and data types.

Splunk has an ecosystem of over 700 apps that can be used to jump-start and support implementations. Splunk Enterprise Security and the Splunk App for Microsoft SharePoint (see Figure 2) offer out-of-the-box content for analysis of access events to help you audit user information, portal activities and access behaviors.

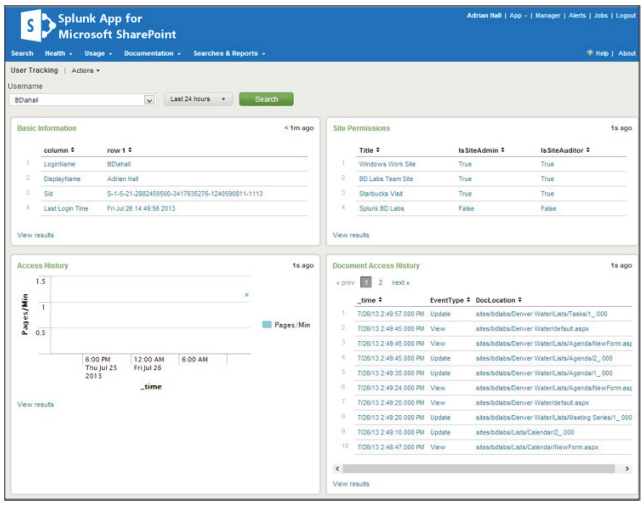


Figure 2 - User Document Access History in Microsoft SharePoint.

Look-ups to other local and non-local systems allow users to incorporate organizational and external context into security or insider threat investigations on-demand, scheduled or as a real-time activity. This provides the insider threat analyst access to context at scale. Splunk software is the bridge between anomalous IT behaviors and the context needed for analysis.

Summary

Many federal agencies create or collect data that can be a tempting target for a malicious insider for financial gain, as a contentious objector or willful destruction. For the benefit of all citizens, careful and thoughtful processes and big data solutions should be applied for continuous evaluation and monitoring for insider threats. There are three key data types that are required to discover malicious insiders:

IT system and security logs (as listed in ICS 500-27): Any data generated by the user as a result of credentialed activity generated from human-to-machine activities

Organizational context: Information about the employee, usually contained in a business system such as an HR database or time management system

External context (as listed in ICS 700-2): Data that an employer can access as a result of an employment agreement, especially for those handling sensitive information or intellectual property

Splunk software removes barriers to asking the right questions, helping you gain greater insight into insider threat motives and actions. Getting answers to behavioral questions requires deeper levels of investigation and data analysis, such as:

Statistical analysis: How many times or how often has something happened? How long does an activity take? Are there normal activity baselines that this employee exceeds on a regular basis? Has the ratio of website types visited changed beyond a statistical norm?

Personal activity comparative analysis: What can be seen in time management or HR system data to indicate out of the norm behavior? For example, lack of vacation time, changing home addresses multiple times in a short period, or variation in work schedule.

User activity context analysis: What are the unusual activities that might be considered red flags? Examples can include an employee beginning to use printers on other floors of a facility, accessing code repositories and network information shares that contain network diagrams, or attempts to gain access to restricted physical spaces. Is a 'nontechnical' employee performing technical activities?

Appendix A

ICS 500-27 Set of Auditable Events:

Auditable Events or Activities

Authentication events:

Logons (Success/Failure)

Logons (Success)

File and Object Events

Create (Success/Failure)

Access (Success/Failure)

Delete (Success/Failure)

Modify (Success/Failure)

Permission Modification (Success/Failure)

Ownership Modifications (Success/Failure)

Writes/Downloads to external devices/media (e.g.,

A-Drive, CD/DVD drives, printers) (Success/Failure)

Uploads from external devices/media (e.g. CD/DVD drives) (Success/Failure)

User & group management events

User add, delete, modify, suspend, lock (Success/Failure)

Group/Role add, delete, modify (Success/Failure)

Use of Privileged/Special Rights events (Success/Failure)

Security or audit policy changes (Success/Failure)

Configuration changes (Success/Failure)

Admin or root-level access (Success/Failure)

Privilege/Role escalation (Success/Failure)

Audit and log data access (Success/Failure)

System eboot. Restart & Shutdown (Success/Failure)

Print to a device (Success/Failure)

Print to a file e.g., PDF format (Success/Failure)

Application initialization (e.g., Netscape, Lotus Notes, IE, etc.)

Initialization (Success/Failure)

Export of information (Success/Failure)

Import of information (Success/Failure)

Auditable Event Details Information Elements

Date and time of the event using the common network time (Network Time Protocol (NTP) Protocol)

Type of event (e.g., login, print, etc.)

Identifier indicating the source system of the event activity

Identifier indicating the identity of the subject or actor (e.g., UserID, ProcessID, etc.)

Details identifying any object or resources accessed or involved (aka Resource List, e.g., files (including location), document ID, peripherals, storage devices, etc.) Outcome (Failure/Success)

Attributable Events Indicating Violation of System/Target

Malicious code detection

Unauthorized local device access

Unauthorized local executable (may be evidenced in patterns of host system behaviors)

Unauthorized privileged access

After-hours privileged access

System reset/reboot

Disabling of the audit mechanism

Downloading to local devices

Printing to local devices

Uploading from local devices

Appendix B

Additional Data Sources/Actions and Sample Correlations for Discovering Insider Threats

Mail of document from printer

Scan of document from scanning device (e.g., printer)

Access to Database (Success/Failure)

Access to SharePoint (or other document sharing system) (Success/Failure)

Access to Active Directory (LDAP) (Success/Failure)

Access to code repository (Success/Failure)

Access to network documentation drive/folder (Success/Failure)

Access to HR or time management system (Success/Failure)

Create shared drive (Success/Failure)

Access to HVAC system (Success/Failure)

Physical access (facility and sensitive areas) (Success/Failure)

File transfers – internal network (using FTP, SFTP, Telnet, other) (Success/Failure)

File transfers – external (Success/Failure)

Use of Gmail, Hotmail and other public email services

Use of instant messaging clients for file sharing

Change in ratio of types of websites visited (proxy data)

Password resets

Use of automated password reset system

Local changes to GPO

Universal changes to GPO

Transaction ID (used to track actions across an application stack)

Behavioral events indicating abnormal behaviors (may be seen as mathematical outliers against a "normal" baseline of activity)

Number of password reset attempts over time

VPN initialization on local network

Printing to printers not adjacent to the user

Higher than normal number of print jobs

Print jobs at unusual times of day

Print jobs while on vacation

Post employee termination access attempts

Access from IPs or subnets not normal for the employee

Access to sensitive data/systems while on vacation

Wireless access while on a wired network connection

Changes in facility access

Access to applications and network resources by employee from geo-locations not typical

Context data: Used to reduce false positives and add context to unusual events as evidenced in IT data.

Vacation schedule data: Those persons that have not taken a vacation in the last 24 months are people that do not want to relinquish control of their work to others.

Lack of vacation taken should be used to assess local connections to applications and other network resources.

Time management system data: Employees that arrive before the rest of the workforce and leave after may be people that do not want others to see what they are doing.

HR data: Recent demotions/disciplinary action watch list. This should be a simple list of names maintained by HR that are not visible to the security staff (scrubbed or redacted at the presentation layer). 50 percent of insider threats were from disgruntled employees.

Watch list of contractors whose contracts will end in the next 30 days. Evidence of data exfiltration should be compared with this list (application developers have a sense of entitlement to what they create and source code is often reused form one job to the next).