A Love Affair: Cyber Security, Big-data and Risk

Mark Seward, Senior Director Security and Compliance, Splunk Inc.

Professional Techniques - Session 31

Security – what's at stake

"On average, organizations are experiencing a staggering 643 Web-based malicious events each week – incidents that effectively penetrate the traditional security infrastructure."

FireEye Advanced Threat Report – 1H 2012 Released
August 29, 2012

Security – what’s at stake

We need to stop what has been the “greatest transfer of wealth in history” that U.S. companies lose to foreign hackers.

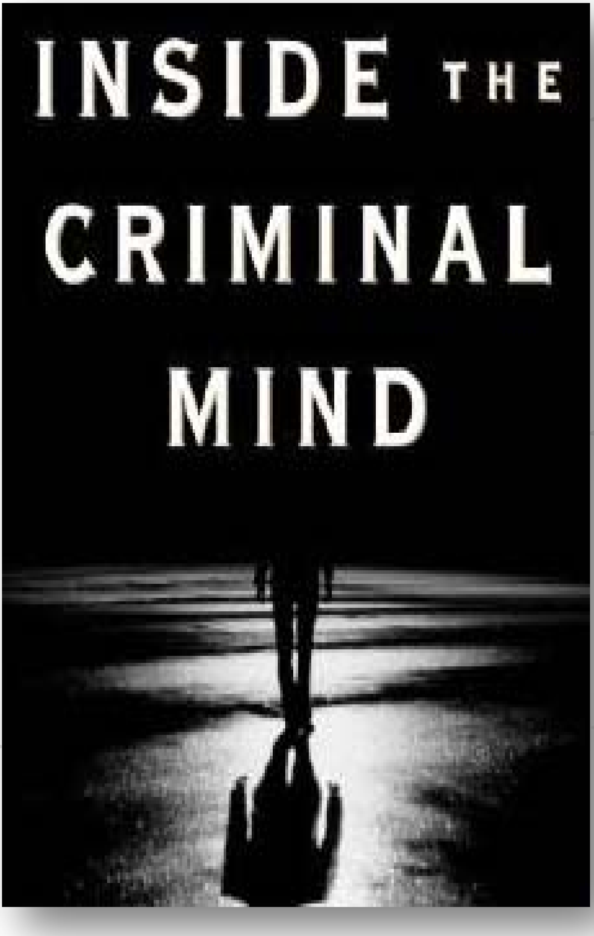
Army Gen. Keith B. Alexander, NSA Director

Understanding Unknown Threats, or 'Thinking like a Criminal'

Where is the most important and valuable data?

What are the typical security defenses?

What structural information silos that exist for the security team?

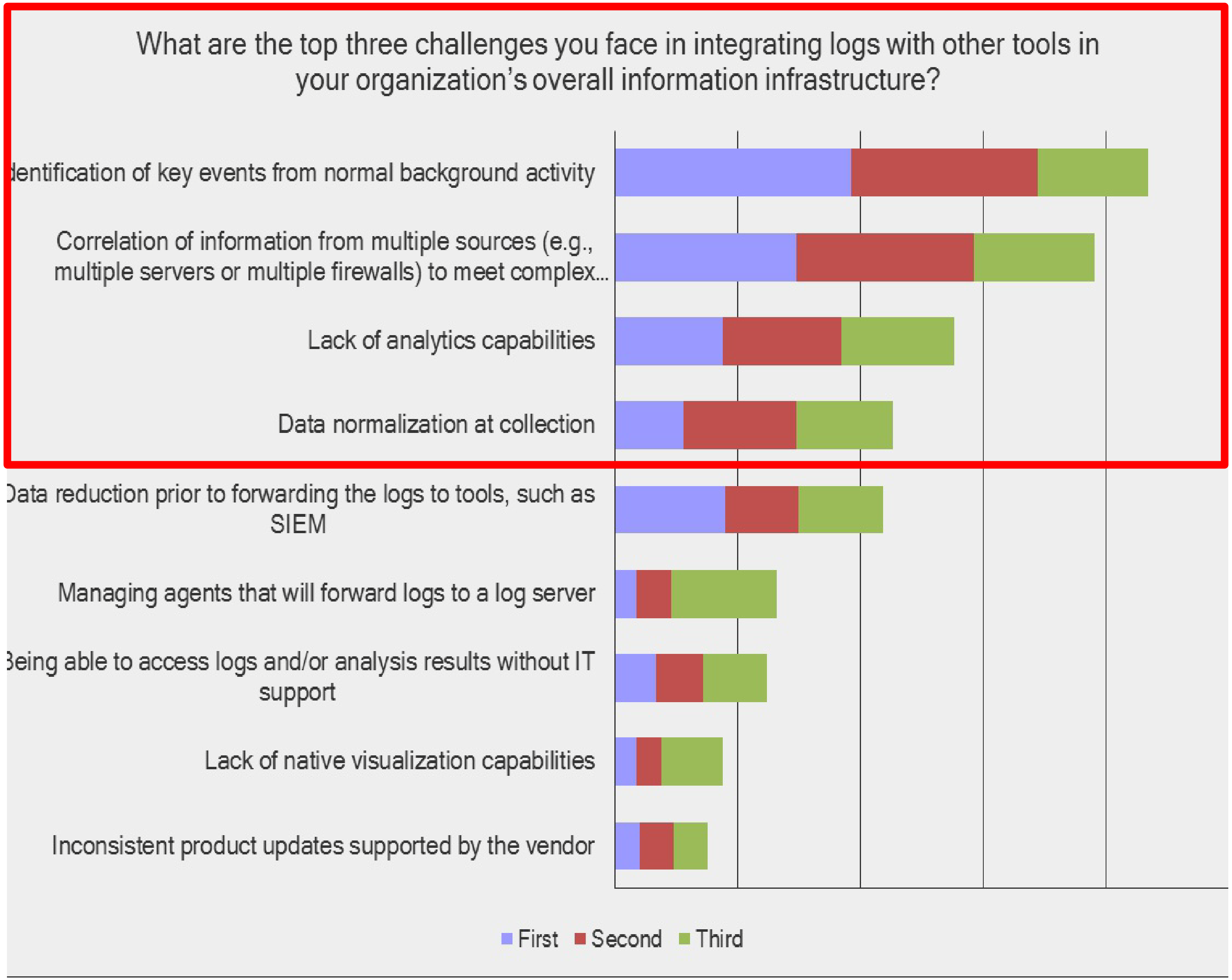


What's the typical patch cycle for applications and operating systems?

How does the IT team prioritize vulnerabilities?

Are 'normal' IT service user activities routinely monitored and correlated?

Top challenges in log management



SANS Log Management Survey 2012

What’s Big-data?

- The Three Vs
  - Data volume
  - Data variety
  - Data velocity
- All too much for a traditional data store

Unknown threats – defining Security Intelligence

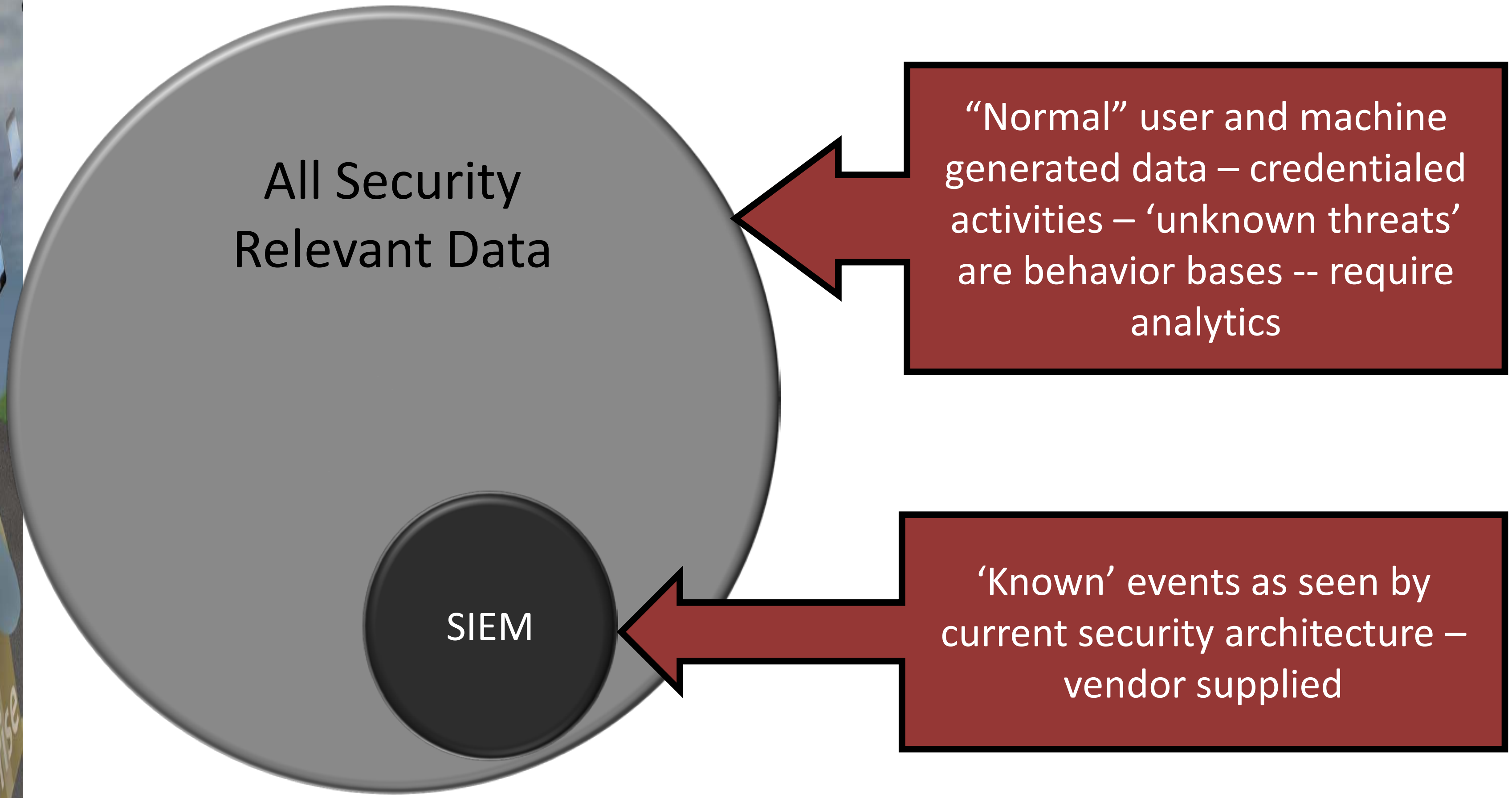
Enterprise Security Intelligence is:

- The collection of data from **all** IT systems in the enterprise that could be security relevant and
- The application of the security team's knowledge and skill
- Resulting in risk reduction

Gartner

Prepare for the Emergence of Enterprise Security Intelligence,
Joseph Feiman, Gartner, June 29, 2011

The spheres of security data

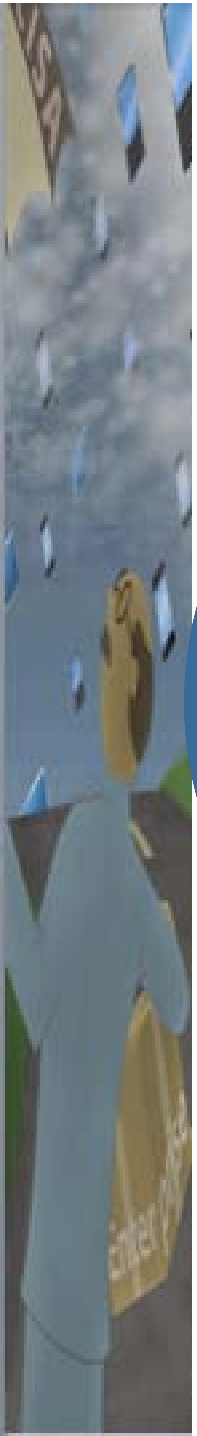


All Security Relevant Data

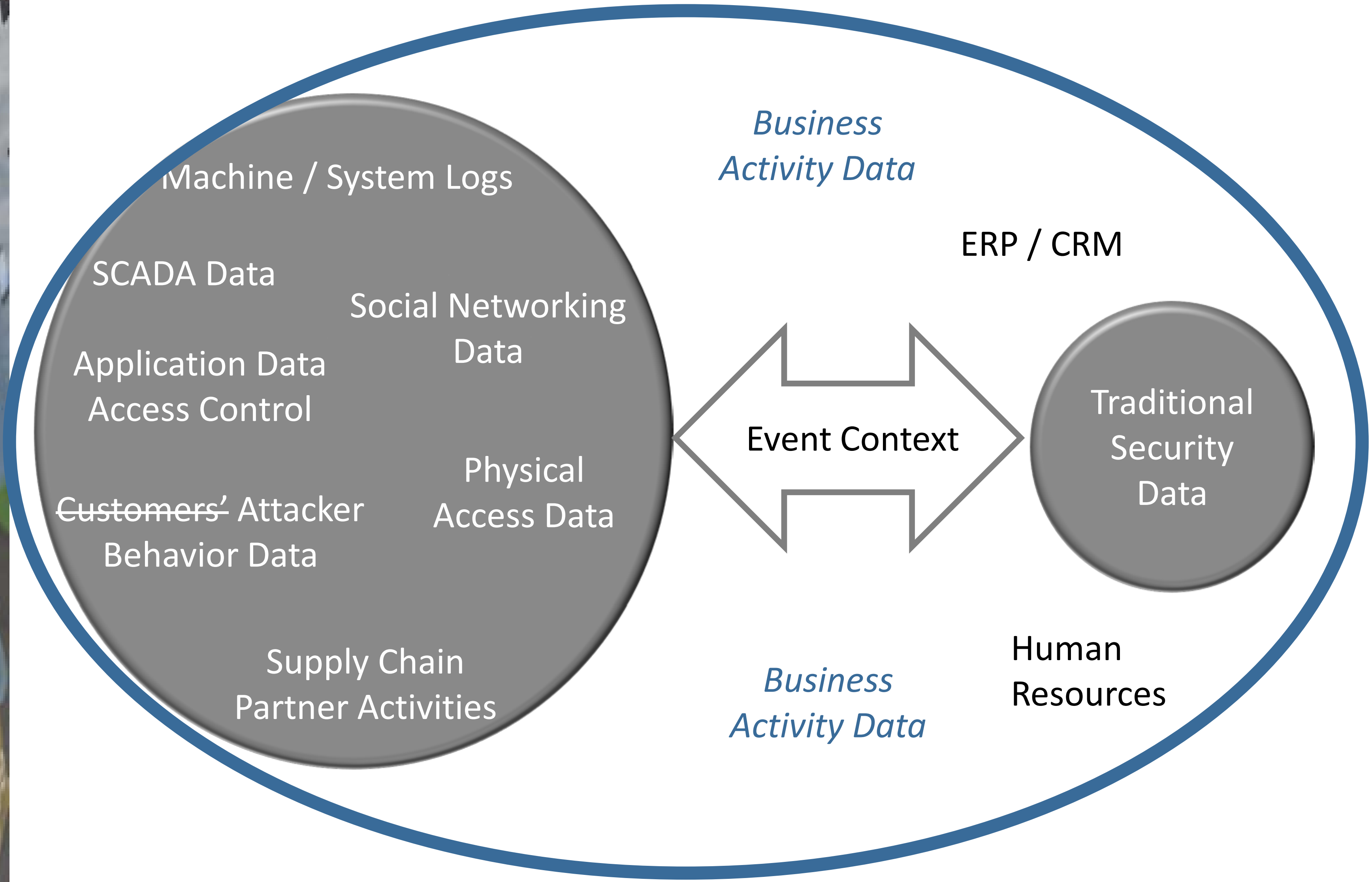
SIEM

“Normal” user and machine generated data – credentialed activities – ‘unknown threats’ are behavior bases -- require analytics

‘Known’ events as seen by current security architecture – vendor supplied



Cybersecurity Risks are Business Risk



Machine / System Logs

SCADA Data

Social Networking Data

Application Data Access Control

Physical Access Data

~~Customers'~~ Attacker Behavior Data

Supply Chain Partner Activities

Business Activity Data

ERP / CRM

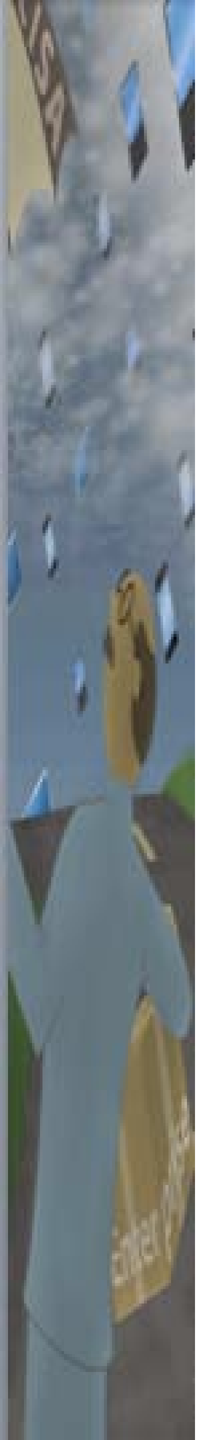
Event Context

Traditional Security Data

Business Activity Data

Human Resources

Knowing what's normal and what's not requires big-data, context, and analytics

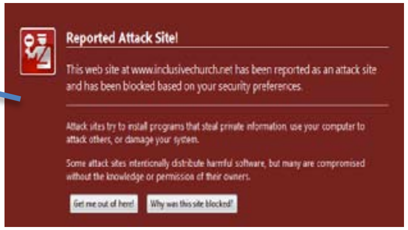


Detecting the malicious insider – data required

Email

Time

Locatio
n



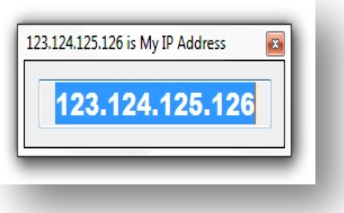
Reported Attack Site!

This web site at www.inclusivechurch.net has been reported as an attack site and has been blocked based on your security preferences.

Attack sites try to install programs that steal private information, use your computer to attack others, or damage your system.

Some attack sites intentionally distribute harmful software, but many are compromised without the knowledge or permission of their owners.

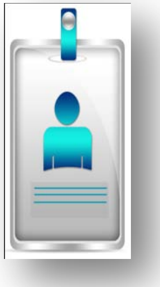
Proxy data
Browsing History



123.124.125.126 is My IP Address

123.124.125.126

IP Address
DHCP / DNS



Badge

Supervisor

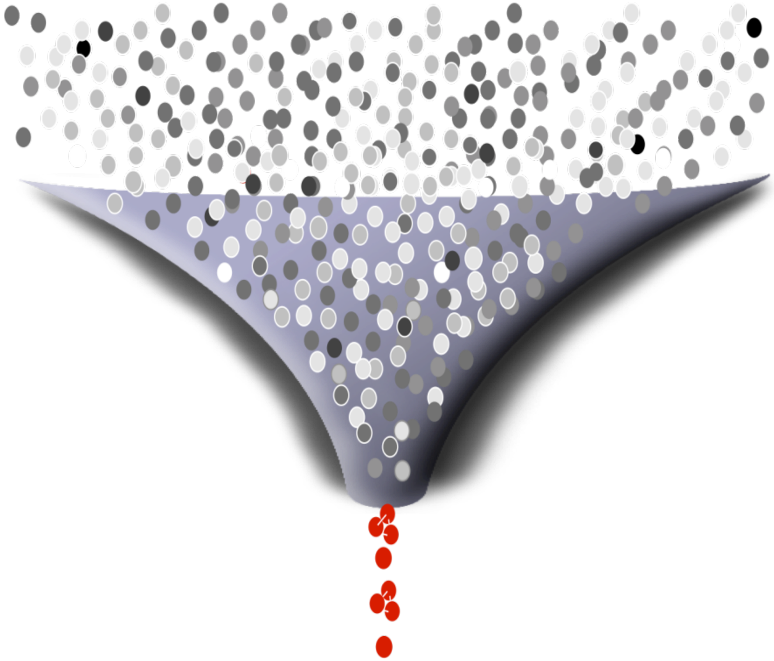
September

Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

Date

Problem with traditional log management / SIEM architectures

Typical SIEM Architecture



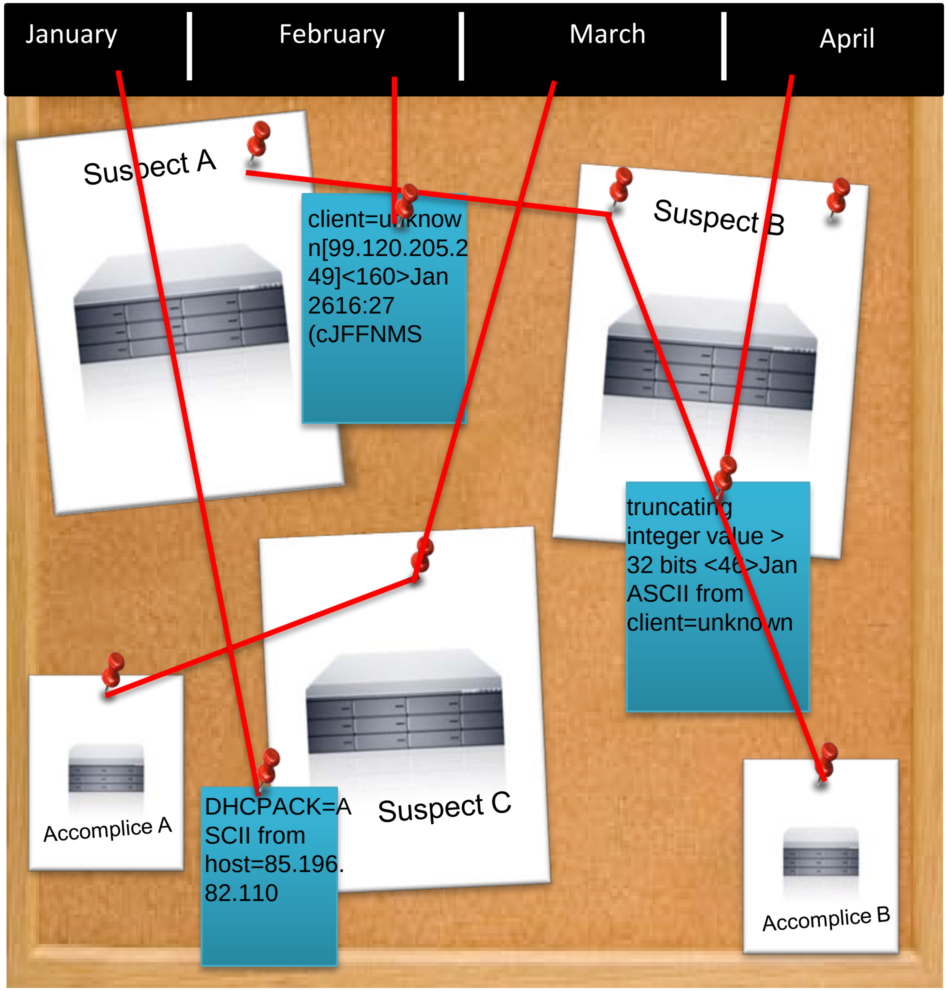
Data Reduction Model

- ✓ Have to know what you need for investigation before you need it
- ✓ Useful data can come from anywhere – not just what's supported by the vendor
- ✓ Lack of scalability restricts visibility
- ✓ Creates vendor dependancy
- ✓ The 'cold case' problem

The ‘Cold Case’ Problem with SIEM

Reinvestigating the ‘crime’

Not possible to add new information to old security events



client=unknown[99.120.205.249]<160>Jan 2616:27 (cJFFNMS

truncating integer value > 32 bits <46>Jan ASCII from client=unknown

DHCPACK=ASCII from host=85.196.82.110

How do you match wits with the creative attacker?

Creativity -- Convergent and Divergent Thinking

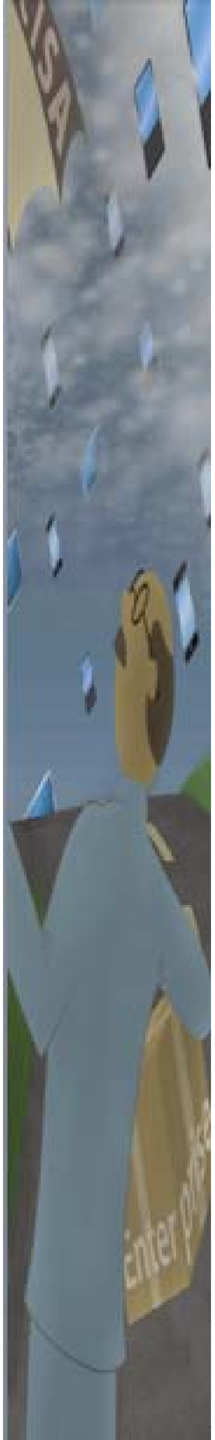
Big-data and Creative Security Thinking

Divergent Thinking:

- The Aha moment / Spontaneous epiphany
- Remote associative processes
- Pattern-based thinking

Convergent Thinking:

- About analysis and attention
- The act of ‘un-concealing’ – chiseling away at a problem
- Write a symphony / poem / solve an algebraic equation
- Stick with a problem till it ‘cries uncle’

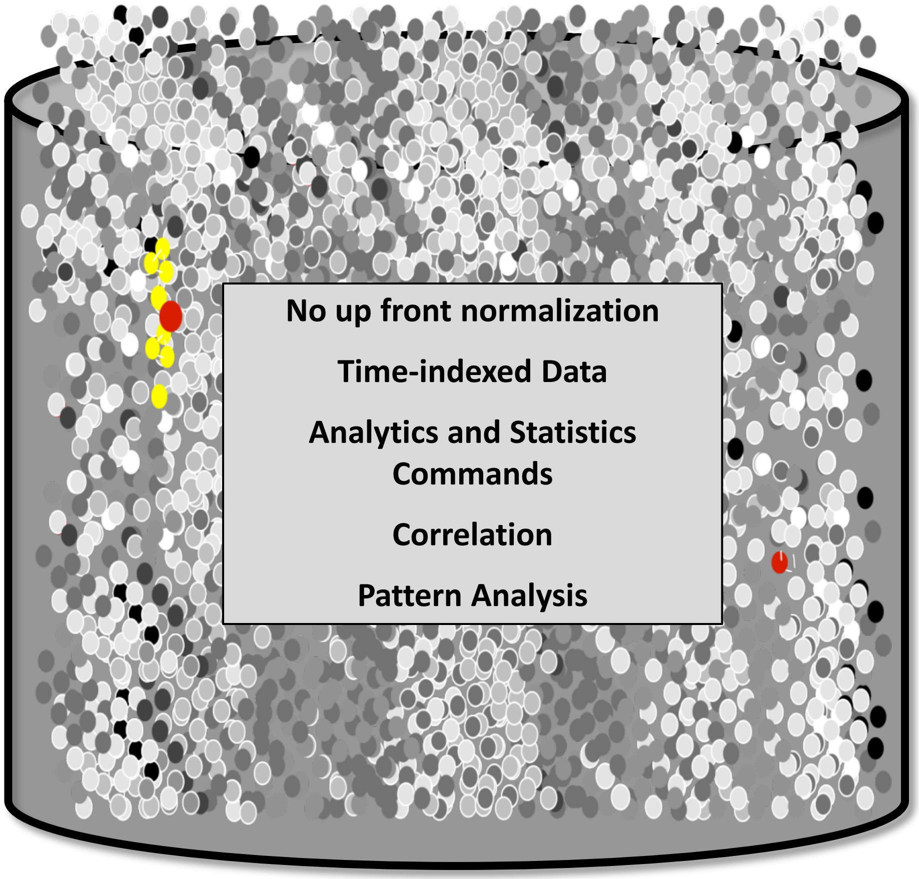


Security Intelligence Needs a New Architecture

Specific behavior based pattern modeling for humans and machines

Based on combinations of:

- Location
- Role
- Data/Asset type
- Data/Asset criticality
- Action type
- How long did the action take
- Time of day



No up front normalization

Time-indexed Data

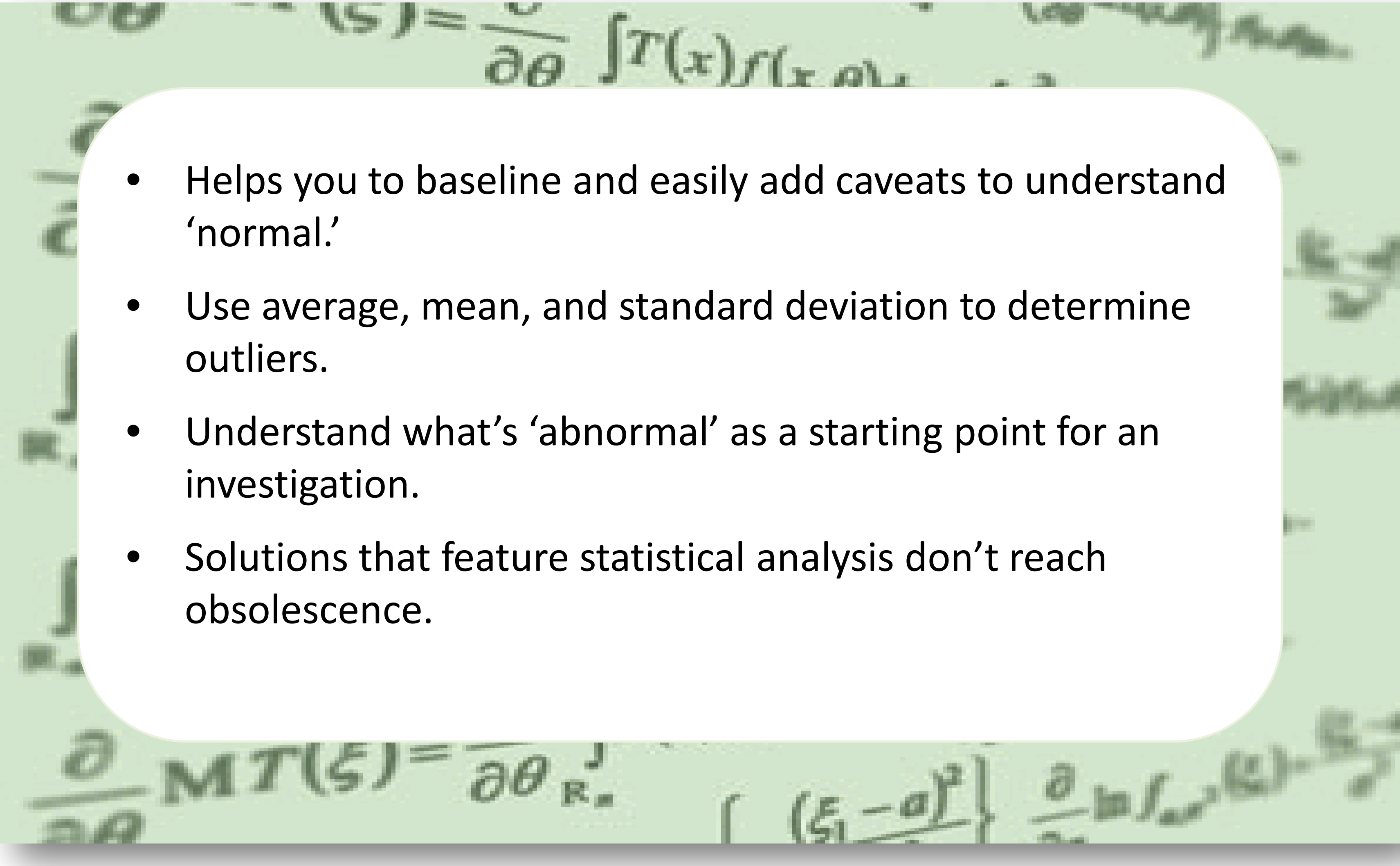
Analytics and Statistics Commands

Correlation

Pattern Analysis

Data Inclusion Model

Mathematics and statistical analysis



- Helps you to baseline and easily add caveats to understand ‘normal.’
- Use average, mean, and standard deviation to determine outliers.
- Understand what’s ‘abnormal’ as a starting point for an investigation.
- Solutions that feature statistical analysis don’t reach obsolescence.

The Way Forward – Risk-base Scenario Thinking

- Operationalize ‘capture-the-flag’ or red-team blue team exercises
- Or, how would you do to steal data from your organization without them knowing?
- What activities would you perform?
- How would these manifest in log data (or the absence of log data)?
- What would the activity patterns look like over time?
- How can we implement ‘call and response’ to automate the investigator mindset?

Find Email and Web Drive-by attacks

Action	Phase	Source	Splunk Search	Why
SQL Injection	Infiltration	WebLogs	len(_raw) +2.5stddev	Hacker puts SQL commands in the URL; URL length is standard deviations higher than normal
Password Brutes	Infiltration	Auth Logs	short delta _time	Automated password guessing tools enter credentials much faster than humanly possible
DNS Exfil	Exfiltration	DNS logs/FW Logs	count +2.5stddev	Hackers exfiltrate the data in DNS packet; standard deviations more DNS requests from a single IP
Web Crawling	Reconnaissance	Web/FTP Logs	count(src_ip) +2.5stddev	Web crawlers (copying the web site for comments, passwords, email addresses, etc) will be the source IP behind page requests standard deviations higher than normal
Port Knocking	Exfil/CnC	Firewall	count(deny) by ip	Threat does inside-out port scan to identify exfiltration paths

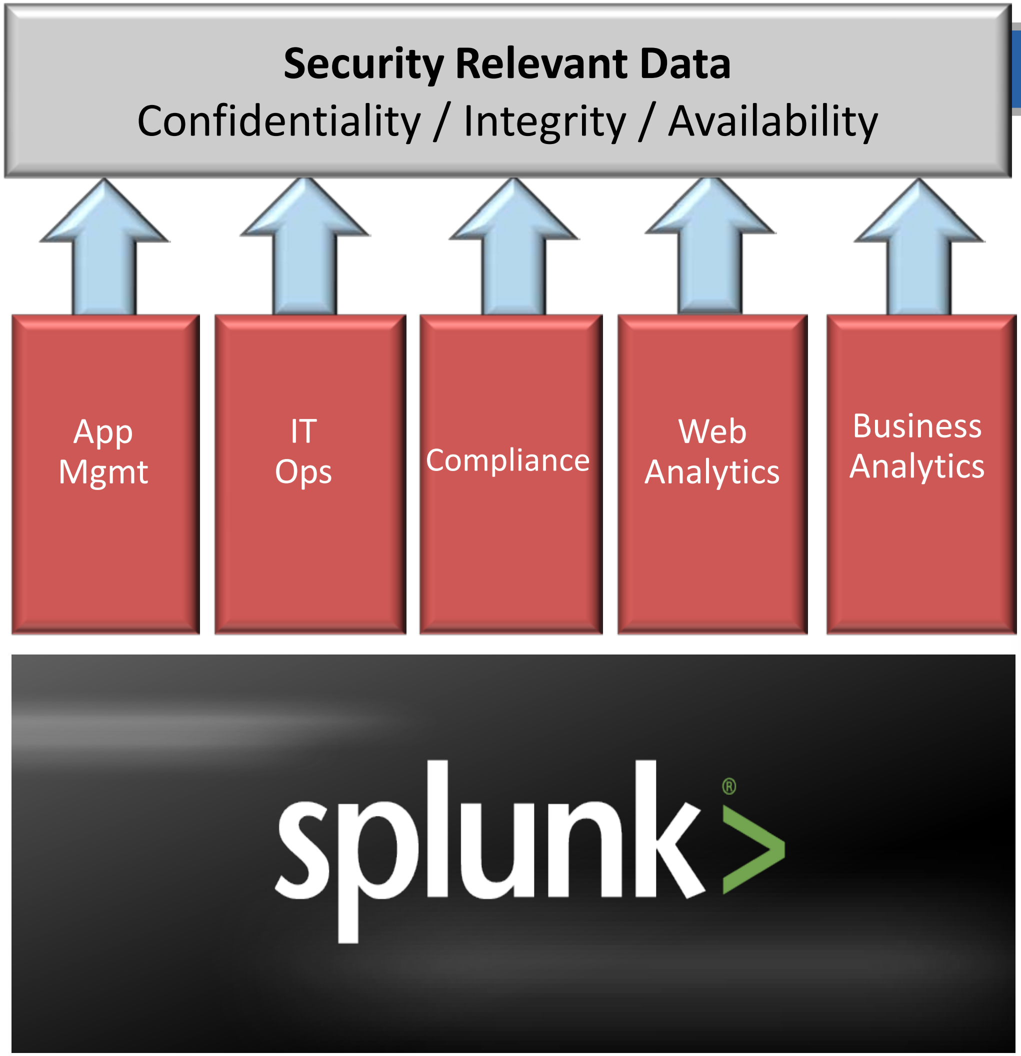
Statistical Analysis

Action	Phase	Source	Splunk Search	Why
Spear Phishing	Infiltration	Mail Logs	Affinity of Sender	Spear phishing sender address has likely never been seen by the company's mail servers
Bad Mail Links	Infiltration	Mail Logs	Domain Affinity	URL likely has never been seen by the company's web servers --fingerprint attackers
Low/slow exfil	Exfiltration	Proxy/FW logs	Avg(bytes)/GET	Small amounts of data leaving in many sessions over time
Form based exfil	Exfiltration	Proxy Logs	Transaction: Post w/o GET	Large amounts of data leaving in few sessions. POST without GET implies automated process
HTTP CnC	Exfil/CnC	Proxy Logs	Long URL w/o Referrer	Botnets often embed long CnC message in the URL

Account take-overs and statistical analysis

- Account takeover – statistical analytics and thresholds
  - Behavior of logins and password changes and resets
  - Analysis of same IP – multiple password resets
  - Multiple IPs -- resetting the same account
- How many times people change their bank information
- How many times they change their credit card information

Data Convergence for Security and Risk



Security Relevant Data
Confidentiality / Integrity / Availability

App Mgmt

IT Ops

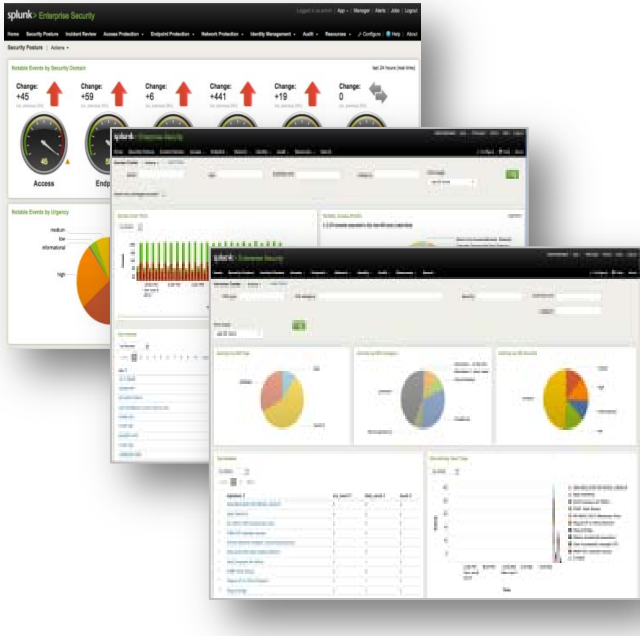
Compliance

Web Analytics

Business Analytics

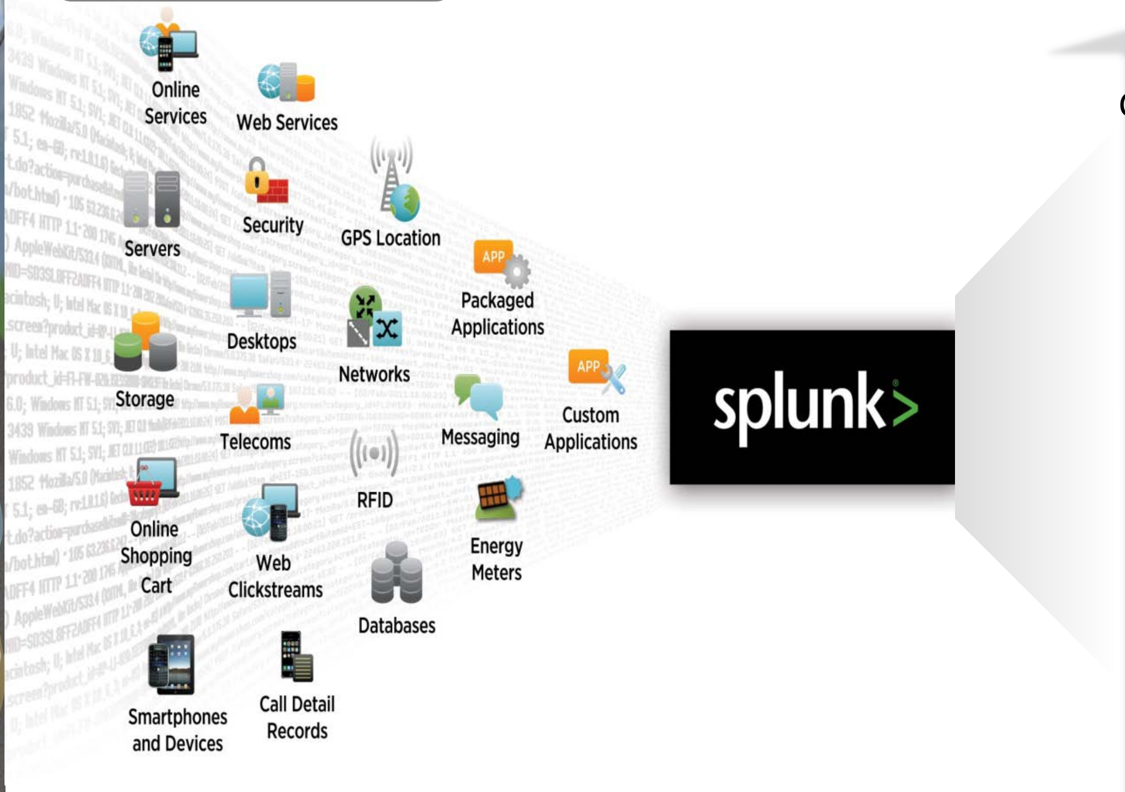
splunk>

CSO / CIO / CEO Views



The Security Intelligence Platform

Traditional and non-traditional data sources



Security Intelligence for Business

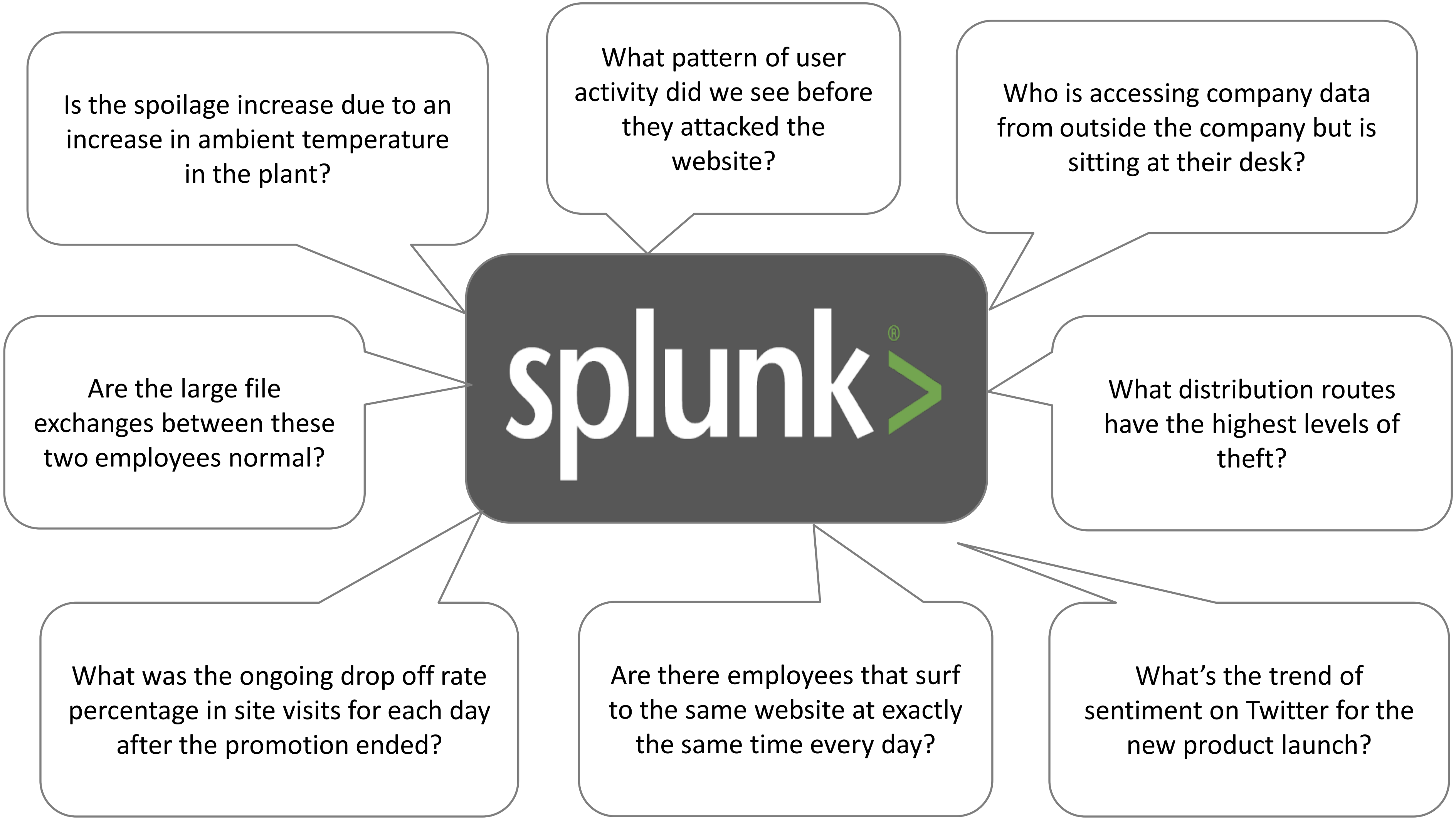
Security Visualizations for Executives
Continuous monitoring of security posture, compliance with internal and external mandates

Statistical Analysis
IT risk scenario based thinking, 'Thinking like a criminal',

Proactive Monitoring
Monitoring of security infrastructure, automation of forensic analysis searches

Search and Investigation
Forensic investigation for security, root cause analysis, application security awareness, transaction monitoring

What kinds of business risk questions could you ask you data?



Is the spoilage increase due to an increase in ambient temperature in the plant?

What pattern of user activity did we see before they attacked the website?

Who is accessing company data from outside the company but is sitting at their desk?

Are the large file exchanges between these two employees normal?

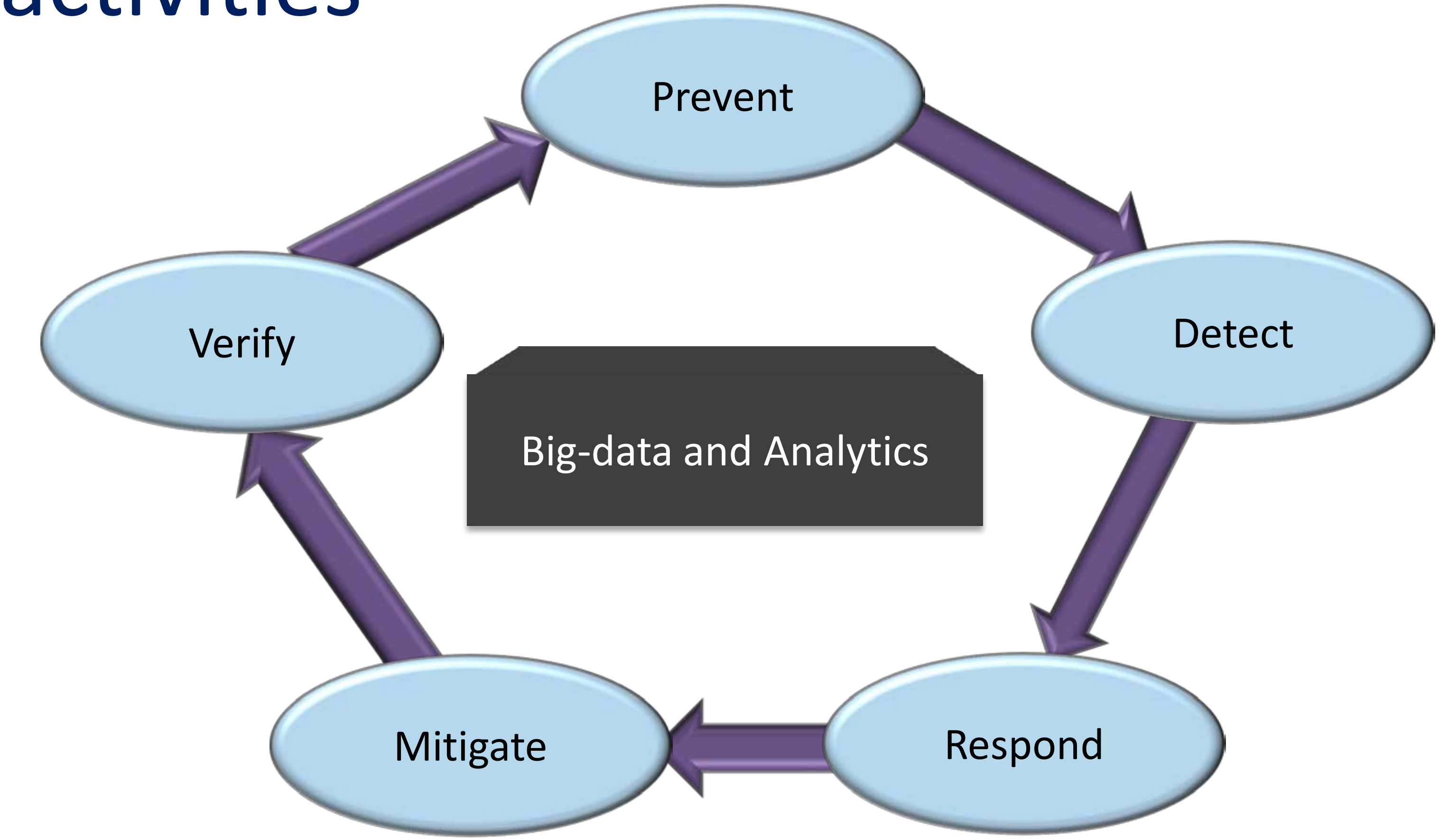
What distribution routes have the highest levels of theft?

What was the ongoing drop off rate percentage in site visits for each day after the promotion ended?

Are there employees that surf to the same website at exactly the same time every day?

What's the trend of sentiment on Twitter for the new product launch?

Security Team Activities – five key activities



Creating and supporting the security mental model

Big-data -- implications for compliance

- Better visibility in to processes
  - Audit supply chain
  - Audit business processes
- Centralized Analytics
  - Audit results are more reliable and consistent
  - Audit process is more consistent, efficient and repeatable
  - Audit costs significantly reduced

Questions

mseward@splunk.com