The Robot in the SOC
What is SOC automation

By Rod Soto
@rodsoto

$Whoami

Rod Soto has over 15 years of experience in information technology and security. Currently working as a Director of Security Research at JASK. He has spoken at ISSA, ISC2, OWASP, DEFCON, DerbyCon, Underground Economy, Hackmiami, Bsides and also been featured in Rolling Stone Magazine, Pentest Magazine, Vice,Univision and CNN.

Rod Soto was the winner of the 2012 BlackHat Las vegas CTF competition and is the founder and lead developer of the Kommand && KonTroll/NoQRTR competitive hacking Tournament series.

What is a SOC

A security operations center

“An **information security operations center** (**"ISOC"** or **"SOC"**) is a facility where enterprise information systems (web sites, applications, databases, data centers and servers, networks, desktops and other endpoints) are monitored, assessed, and defended.” *

What is a SOC

What do you need a SOC for?

A SOC is related to the people, processes and technologies that provide situational awareness through the detection, containment, and remediation of IT threats. A SOC will handle, on behalf of an institution or company, any threatening IT incident, and will ensure that it is properly identified, analyzed, communicated, investigated and reported. The SOC also monitors applications to identify a possible cyber-attack or intrusion (event), and determines if it is a genuine malicious threat (incident), and if it could affect business.

Regulatory requirements

Establishing and operating a SOC is expensive and difficult; organisations should need a good reason to do it. This may include:

- Protecting sensitive data
- Complying with industry rules such as PCI DSS.[1]
- Complying with government rules, such as CESG GPG53

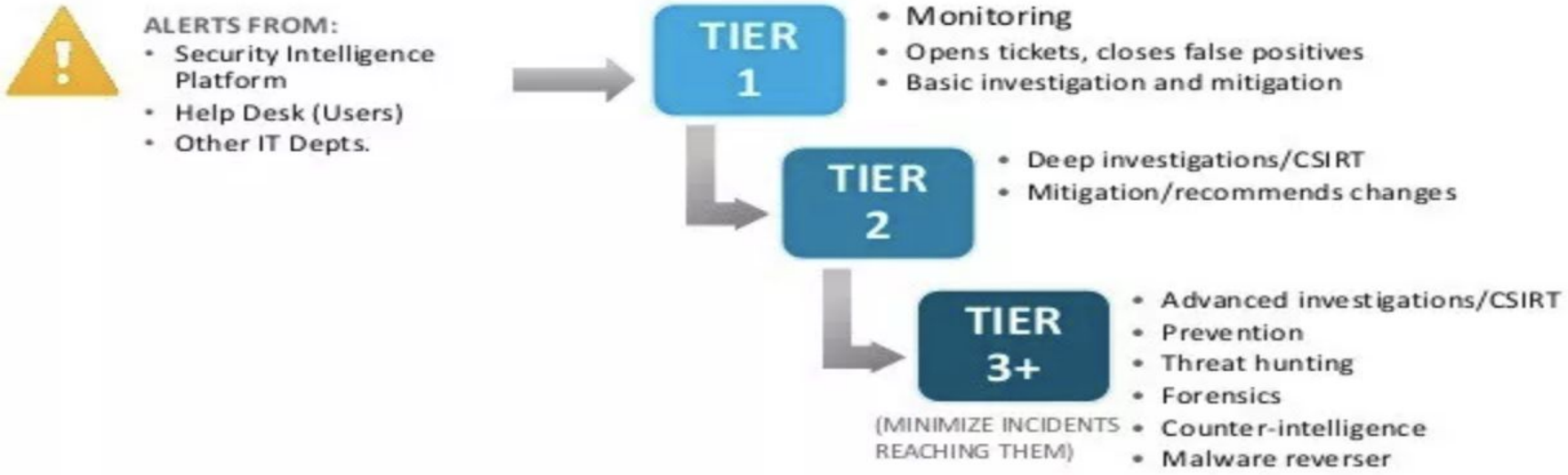
SOC types

- **Operational → Are things running OK**

- **Threat Centric → Are we being attacked? Anything malicious/unusual**

- **Compliance → Monitoring the compliance of your systems against reference configuration templates or standard system builds**

Security Operation Centers



Source: GBHackers

Security Operations Center

Where do all the information comes from?

Enterprise information systems (web sites, applications, databases, data centers and servers, networks, desktops and other endpoints.

So...

Data collection, aggregation, detection, and analytics of all these gathered information are fundamental for a SOC to function.

The problem with SOC

- Most SOCs rely on SIEM

  What is SIEM?

  “security information and event management software products and services combine security information management and security event management.”

  SIEM is good until you find yourself and your analysts dealing with...

Big Data Technologies

Current big data technologies provide an additional complication

- Too much data
- Too many disparate sources (Firewalls, EDR, logs, cloud, etc)
- Every technology requires training (QRadar, ArcSight, Splunk)
- Mainstream SIEM tech is proprietary and requires extensive investment
- Costs of putting all these together (Some companies EVEN charge for indexing/storage)

Security is not humanly scalable

- It takes 6 months to 1 year to have a trained SOC 1
- High turnover
- Current defense technologies inadequate & inaccurate
- Even if analyst are trained, the flow of data is overwhelming.
- Large number of false positives
- Can’t train enough people in all disparate techs in due time.

Security is not humanly scalable

- US companies took an average of 206 days to detect a data breach. - 2017
- The average cost of identifying a breach within this time was $5.99 million, but for breaches that took longer to identify, the average cost rose to $8.70 million

*Ponemon institute

What is SOC Automation

- SOC Automation is the application of technology to streamline the ingestion, collection, analysis and actions of a Security Operations Center in a self-regulating/automatic manner.

- This application of technology covers all the phases of the SOC workflow until information is presented in actionable, contextualized manner to the analyst. SOC automation does NOT replace human, but it does replace many things that SOC1 analysts do.

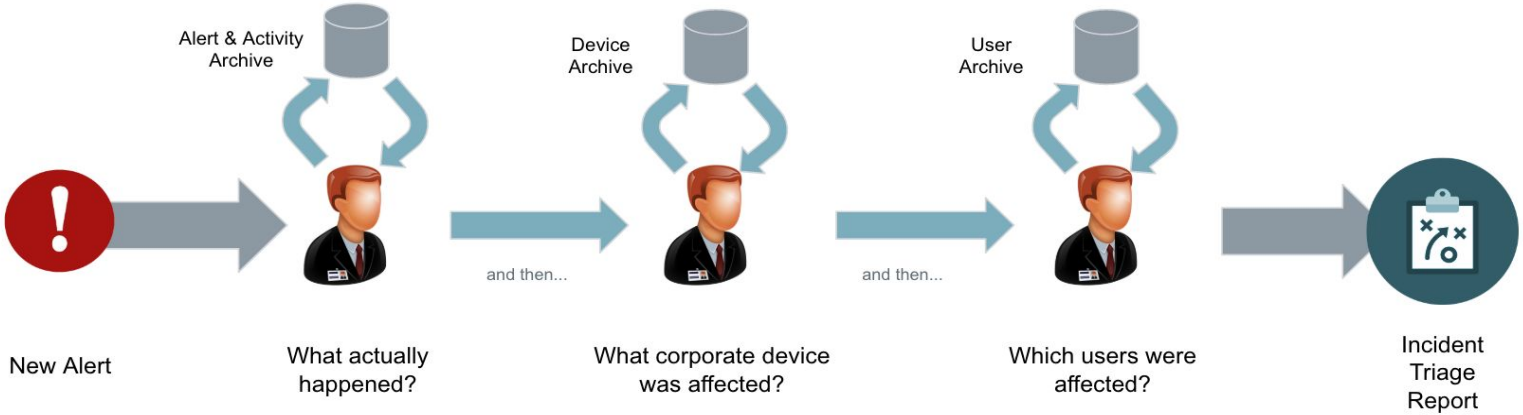
What is SOC Automation

Some of the technologies applied to SOC automation

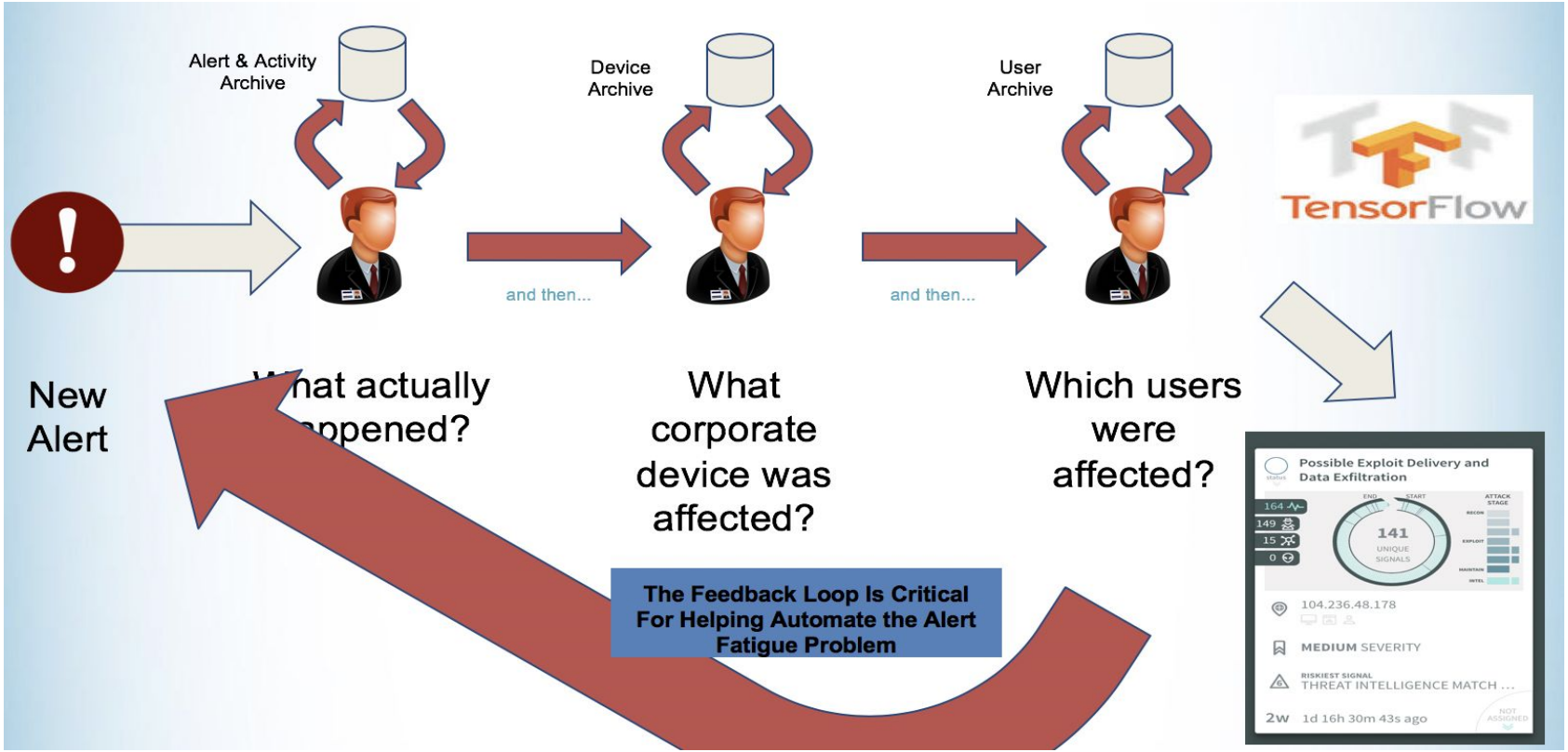
- Commoditized distributed processing technologies (Mainly cloud, Hadoop ecosystem)
- Machine Learning technologies (AWS, Google, Cloudera, Microsoft)
- Open source analytic, indexing storage technologies (ELK)

What is SOC Automation

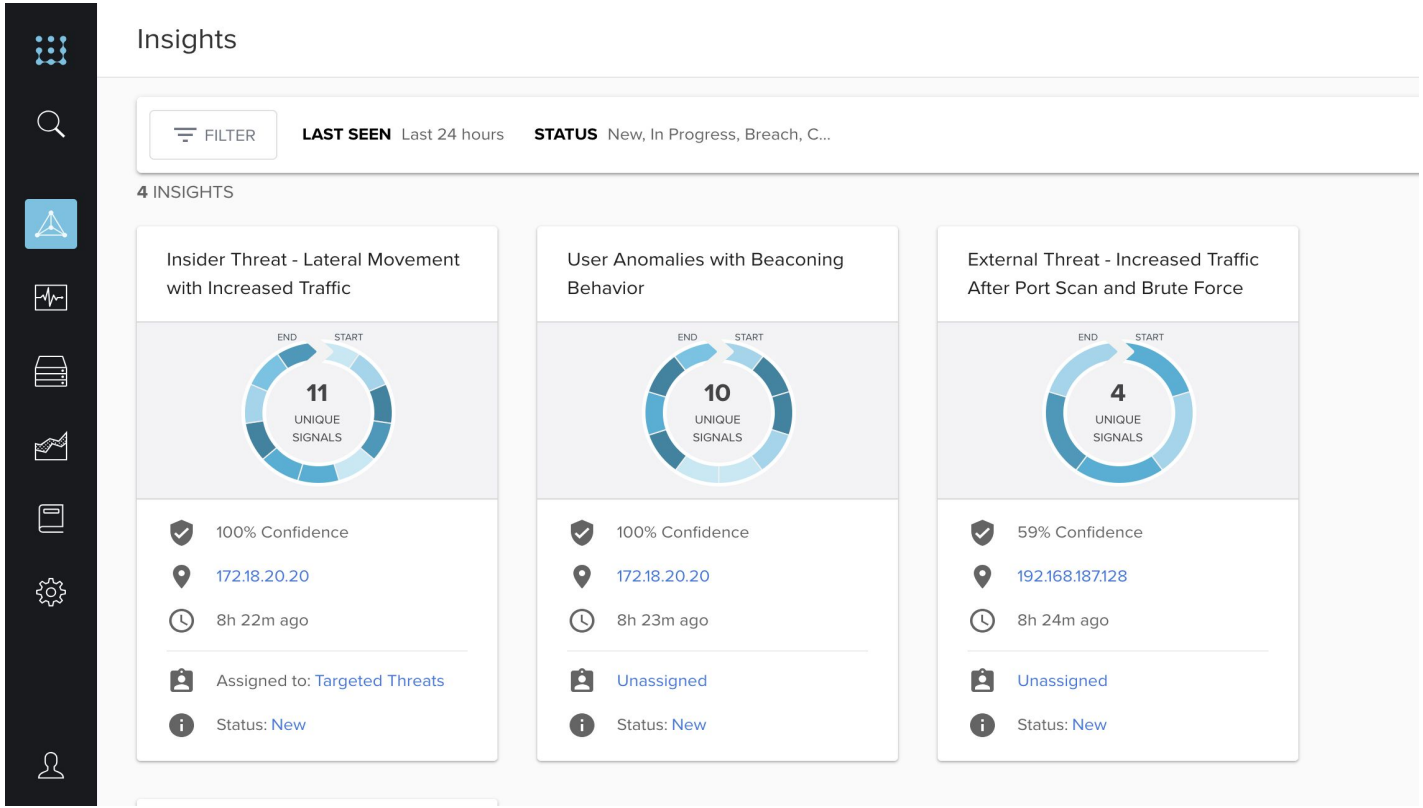
Current SOC Workflow



SOC Automation in action



Automated SOC workflow - example



What is SOC automation

Automated SOC workflow

- Decrease time and increment efficiency in the information presented to Analyst and actionable items.
    - Fewer Alerts
    - Greater context
    - Faster Triage

What is SOC Automation

Some of the benefits of SOC automation

- Autonomous parsing of events.
- Application of Machine Learning models to process and analyze very large amounts of data.
- Scalable based on smart/elastic technologies (use what you need when you need it).
- Crowdsource feedback from experienced analysts.
- Training of Algorithms based on such feedback.

What is SOC Automation

Automated SOC

- Automate basic, repetitive tasks
- Alleviate Alert Fatigue
- Automate ingestion, processing, indexing and alerting
- Depurate false positives and repetitive alerts
- Present only what is relevant to analyst (Analytics/ML)
- Streamline, speed up Triage (Reduced TTR)

Machine Learning backend

- Put together very large and distinct sources of data into a platform for analysis, interpretation and prediction.
- Go beyond of static signature based technologies.
- Creates an scenario where detection of threats based on dynamic and multi contextual indicators is possible.
- Enhances analyst ability to act on detected threats.

What about SOAR/SOAP?

SOAPA = Security Operations Analytics Platform Architecture

SOAR = Security Orchestration Automation and Response

Not a problem, can be integrated with SOC automation platforms. It is simply another piece in the puzzle. Most of the orchestration tools have APIs (Phantom, Demisto, etc)

Are we replacing the analyst?

NO

People vs People model has proven to be more effective as current threat detection/prevention technologies do not seem sufficient nor effective against malicious actors. The numbers speak for themselves.

Criminals learn just like algorithms… so they bypass them as well. Automation enhances analysts operations. Machines cannot replace an experienced analyst. However they can do most of the tasks of an inexperienced entry level tier 1 operator...

Some of the JASK top research ml use cases

USE CASE 1:
SPEAR PHISHING

Phishing campaigns represent one of the most common attack vectors still successful in cybercrime today. These attacks involve sending a malicious attachment or link to an email address in hopes of getting a user to execute malicious content.

HOW TO DETECT

Finding creative ways to capture metadata from users sender/receiver patterns the key idea for analytic detection of phishing. Taking header and subsampling of body data into account it is possible to build a graph of 'sender trust' based on user activity over time.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase1_SpearPhishing

USE CASE 2:
WATERING HOLE

In addition to email phishing, watering hole attacks is another example of the most common techniques used to deceive and manipulate enterprise victims. A watering hole is usually a website or web application service that has been compromised or built as a fake service in order to compromise unsuspecting victims.

HOW TO DETECT

Baselining web application services with path traversal statistics about common interactions is important for data driven detection. Also monitoring for the behavior of rare redirect patterns in tandem with referrer chains is an additional indicator of risk.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase2_WateringHole

Some of the JASK top research ml use cases

USE CASE 3:
LATERAL MOVEMENT

These forms of attacks represent an escalation of risk in an overall kill chain, particularly when targets in the attack campaign go from low level user machines to VIP assets.

HOW TO DETECT

Machine Learning informed data driven contextualization can be achieved with input logs of network traffic. Automated contextualization of different asset profiles lets us get a dynamic view of normal behavior in order to perform change point detection.

https://www.github.com/jasklabs//blackhat2017/datasets/UseCase3_LateralMovement/

USE CASE 4:
COVERT CHANNEL DETECTION

Covert channels are used by attackers to maintain control of compromised assets and to carry out tactics over time through hidden channels in the network.

HOW TO DETECT

Calculating risk for all rare domains with simple statistics is the key to breaking the overall visibility that command-and-control relies on as a covert channel.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase4_CovertChannelDetection

Some of the JASK top research ml use cases

USE CASE 5:
RANSOMWARE

Ransomware is malware that uses drive wiping and encryption techniques to hold devices and infected machines ransom in exchange for an encryption key.

HOW TO DETECT

Micro behaviors related to the ransomware phenomena is a challenging use case because of the lack of network evidence. Focus on the initial infection payload along with entropy statistics on processes interacting with the entire file system.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase5_Ransomware

USE CASE 6:
INJECTION ATTACKS

The OWASP top 10 shows, yet again, injection as the number one vulnerability for web applications attacks. Injection attacks involved supplying untrusted input data to a targeted service.

HOW TO DETECT

DB logs combined with higher layer application data can be used to build statistical profiles of each group of users and how they access individual applications in the enterprise.

https://www.github.com/jasklabs//blackhat2017/datasets/UseCase6_InjectionAttacks/

Some of the JASK top research ml use cases

USE CASE 7: RECONNAISSANCE

An often overlooked use case in analytics technology is the problem of recon, whether it is at the perimeter or within the LAN. Recon is usually left to signature matching or other simple point solutions that will often times create a high volume of noisy alarms.

HOW TO DETECT

Derive a graph of all parts of the topology of a network through netflow data. The algorithms breadth (depth) first search lets us identify the spread of new graph patterns in rapid time correlating to recon tactics.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase7_Reconnaissance

USE CASE 8: WEBSHELL

A webshell is a piece of malicious code on a web server that allows execution of commands, database data dump, file transfers, install of software and other system functions.

HOW TO DETECT

Focusing on statistics that identify normal shopping cart activity from reverse connections, and rare indicators on the file system such as search file and content paths.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase8_Webshell

Some of the JASK top research ml use cases

USE CASE 9:
CREDENTIAL THEFT

Credential misuse including VPN comprise have been at the heart of some of most high profile attacks in history.

HOW TO DETECT

Modeling users typical login patterns alongside correlating their location and time of logins are the first step in adding an behavior driven detection for this use case.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase9_CredentialTheft

USE CASE 10:
REMOTE EXPLOITATION

This attack pattern involves a series of malicious events where a system is targeted and then delivered an exploit which may successfully drop a payload (malicious binary) and then execute code at the targeted system.

HOW TO DETECT

Machine Learning can learn to identify rare sequences in which the sequential behavior manifests as correlations involving the delivery of the exploitation payload.

https://github.com/jasklabs/blackhat2017/tree/master/datasets/UseCase10_RemoteExploitation

Some open source projects you can play with...

AKTAION v2

https://github.com/jzadeh/aktaion2

Aktaion V2 is a python3 project for detecting exploits (and more generally attack behaviors). The project is meant to be a learning/teaching tool on how to blend multiple security signals and behaviors into an expressive framework for intrusion detection.

Some open source projects you can play with...

CHIRON

https://github.com/jzadeh/chiron-elk

CHIRON is a home analytics based on ELK stack combined with Machine Learning threat detection framework AKTAION. CHIRON parses and displays data from P0f, Nmap, and BRO IDS. CHIRON is designed for home use and will give great visibility into home internet devices (IOT, Computers, Cellphones, Tablets, etc).

SOC Automation

Conclusion

- Allows orgs to address and manage challenging amounts of security data ingestion, processing, indexing and analysis.
- Analyst focus on their core tasks (Reduced TTR)
- Better and improved visibility (Noise to signal ratio)
- Combat skill shortage and analyst turnover
- Automate actions on detection (Orchestration)