RSAConference2018
San Francisco | April 16–20 | Moscone Center

SESSION ID: SPO3-T10

HELP ME NETWORK VISIBILITY AND AI; YOU'RE OUR ONLY HOPE

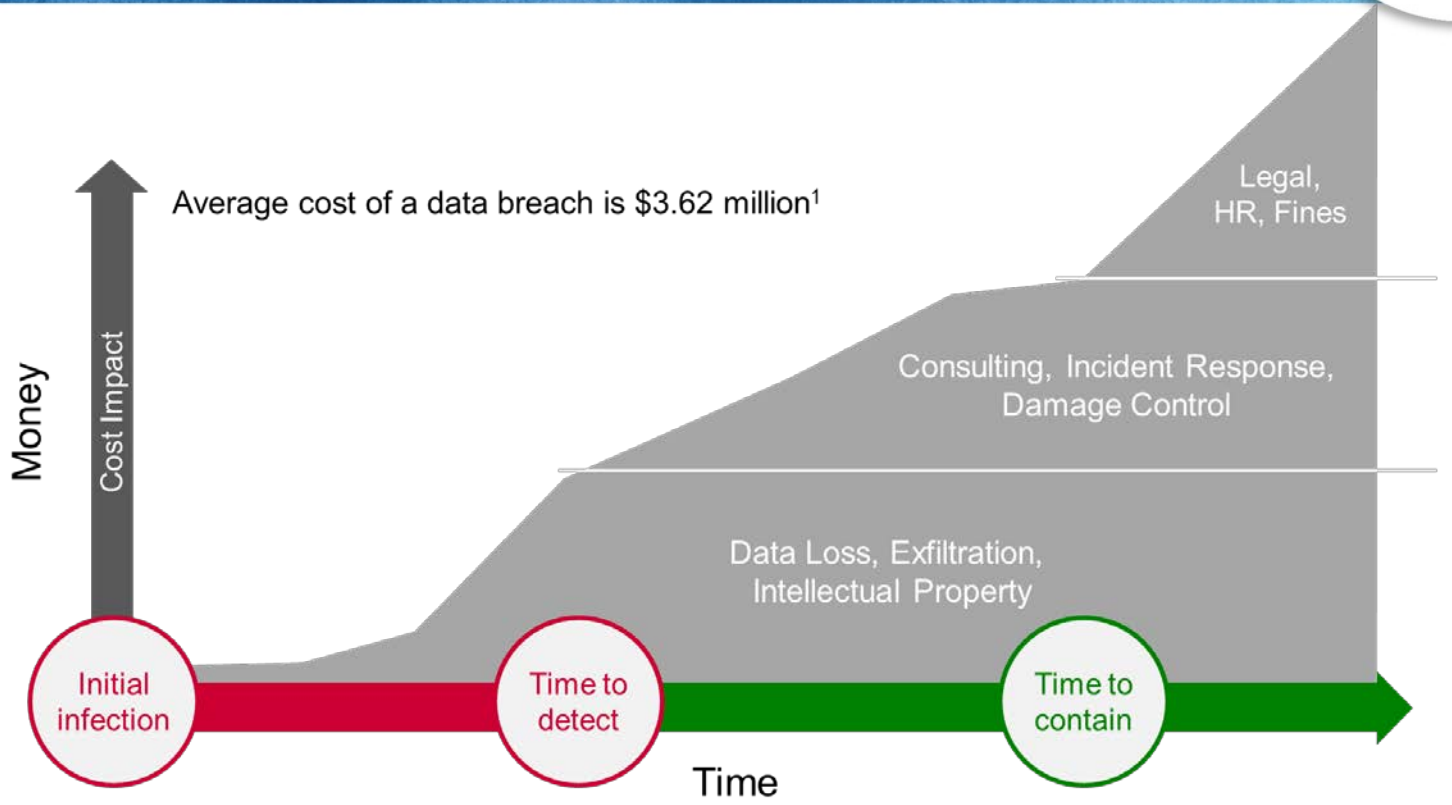
Chris Morales

Head of Security Analytics
Vectra Networks

Steve McGregory

Sr. Director, Threat Intelligence Research Center
Ixia, a Keysight Company
@stevemcgregory

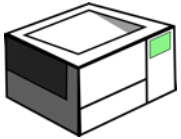
Time is money for security operations



[1] 2017 Ponemon Cost of Data Breach Study

Challenges of today's network

100 Employees



150 PC's/Laptops
30 Mobile Phones
100 BYOD's
120 Mailboxes
100 VoIP Phones

1 IT Guy

110,000

Millions

~1,000,000 Security Alerts Per Month

Steve's Sandbox

18K Workers | 12K Employees + 6K Contract | 5,000 Engineers | 250 IT Engineers
25K PC's | 6K Mobile Phones | 3K Linux Workstations | 550 Macs
22K Mailboxes | 20M Emails/Month | 4M WebEx Minutes/Month

200M Security Analytics Records/Month | 1.7M Block Phishing Attacks/Month
5K Malware Incidents/Months | 2 Major Security Incidents Per Day

150 Applications | 80 Corporate + 40 Manufacturing + 30 R&D
4M Customer Contacts | 2M Marketing Emails/Month | 1M Web Visits/Month
2M PLM Design Docs | 1.5M PLM Parts | 10K ERP Product Models

130 Sites | 37 R&D | 27 Customer Contact Centers
2.7 PetaBytes | 3,071 Servers (73% Virtualized) | 2,643 Managed Network Devices

Challenges of today's network

Can You Handle the Truth?

You Think You Know!
But You Don't…

How we know you don't?

On average, ONLY 29% of alerts received are investigated*

LEADING TO AN AVERAGE BREACH DETECTION TIME OF 170 DAYS.*

YOUR GOAL IS TO BRING 170 DOWN TO < 1

*Ponemon 2016 State of Malware Detection & Prevention

REAL WORLD CASE

Equifax

Equifax Cyber Threat Center

Malicious online security risk is evolving at an alarming rate. To address this issue, we created a Cyber Threat Center as a separate, dedicated group within Global Security. The core focus of this highly specialized team is to:

- Identify and mitigate active threats
- Model new and emerging threats to better understand future risk paths and trajectories
- Support investigations around a variety of situations such as insider threat, external bad actors, fraud and more

This group constantly asks the hard questions. What data do we have that's most valuable to others? Who has access to that data, and are our technical controls working? What's the baseline for "normal," what's changed and what doesn't look right?

The answers they get, paired with market-leading Equifax technology and automated analytics, dramatically expand our view of risk and provide greater, more predictive insight into current and future security issues.

In a typical day, our Cyber Threat Center:

2.5 Billion Logs Captured

50k Events/second Monitored

2,200 Security Device Health Checks

43k Domains Analyzed

250 Intel Forums Queried

Cyber Threat Center

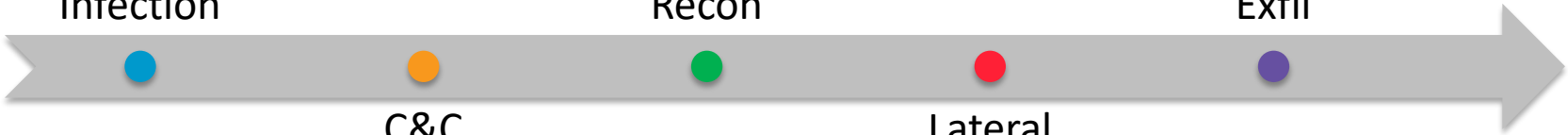
- Intelligence Gathering
- Vulnerability Management
- Countermeasures
- Cyber Security Strategy
- Security Operations

Equifax breach exhibited the attack lifecycle

March 10: Attackers exploit a vulnerability in the Apache Struts Web Framework to gain root access to online dispute web application

Attackers customize tools to efficiently exploit Equifax's software, and to query and analyze dozens of databases to decide which held the most valuable data (Port Sweep, Port Scan, Internal Darknet Scan, Kerberos Account Scan)

The trove of data the attackers collected was so large it had to be broken up into smaller pieces to avoid triggering as an anomalous behavior (Data Smuggler, Hidden HTTPS Tunnel)



Attackers set up about 30 web shells that were accessed from around 35 distinct public IP addresses – China Chopper (External Remote Access, Suspect Domain Activity)

May 13 – July 30: Attackers used hidden tunnels to bypass firewalls, analyzing and cracking one database after the next while stockpiling data on the company's own storage systems (Hidden HTTPS Tunnel, Suspicious Admin)

Why was Equifax so slow to respond?

Time to detect and respond

- Time to patch: 138 Days
- Time to notice compromise: 78 Days
- Time to notify public: 117 Days

In spite of…

- Investing millions in security measures
- Running a dedicated operations center
- Deploying a suite of expensive anti-intrusion software

Security effectiveness compromised by

manual processes and the departure of key personnel

Alert Fatigue

Agenda

- ~~Introductions~~
- ~~State some obvious problems~~
- The Goals for Today
- Definition
  - What is Network Visibility?
  - What is ML?
- Application
  - How to apply ML and Network Visibility to Network Security
  - BTW, it works for other stuff like network performance, downtime prediction

The Goals for Today

#BeatTheBreach

START TO BRING 170 DAYS DOWN TO < 1

What is Network Visibility?

Not Just Perimeter Visibility

Not Just Core Visibility

Not Just Cloud Visibility

Not Just Device Visibility

Not Just Application Visibility

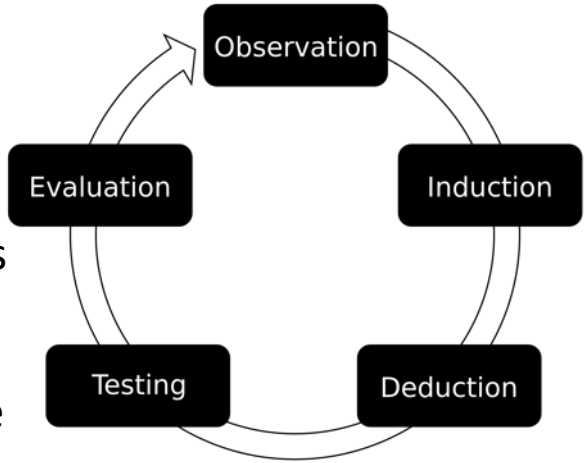
What is Network Visibility?

Network Visibility is the enabler of actionable intelligence.

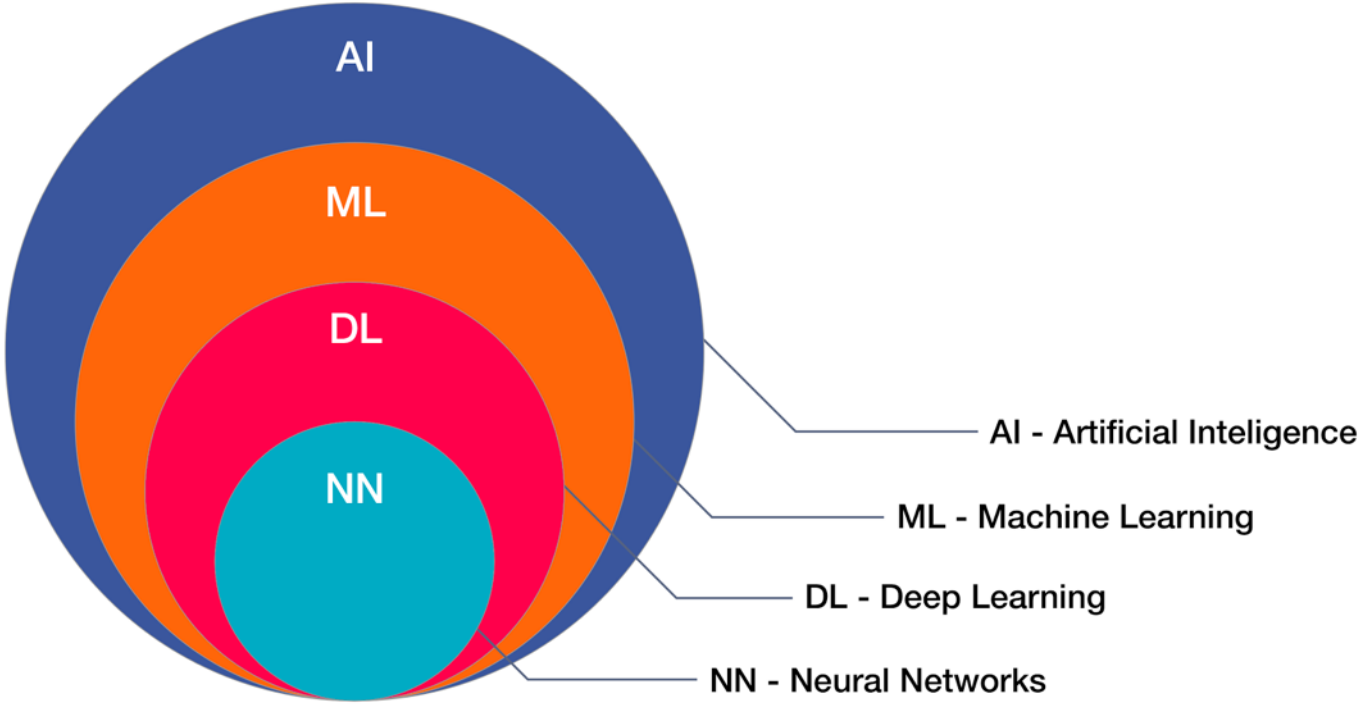
Empirical research

Applied to Network Traffic Leads to ML Success

Leads to you knowing everything, all the time



What is Machine Learning?



What is Machine Learning?

Give “computers the ability to learn without being explicitly programmed”

(Arthur Samuel, 1959)

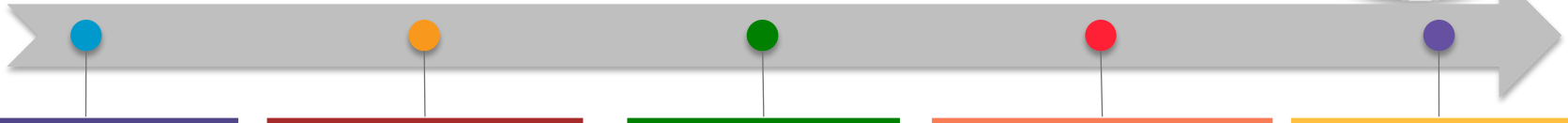
Using algorithms, you can allow the computer to learn from observed patterns and recognize them in future trials.

It is not AI, but it is a step toward AI

It does provide us with ”A New Hope”

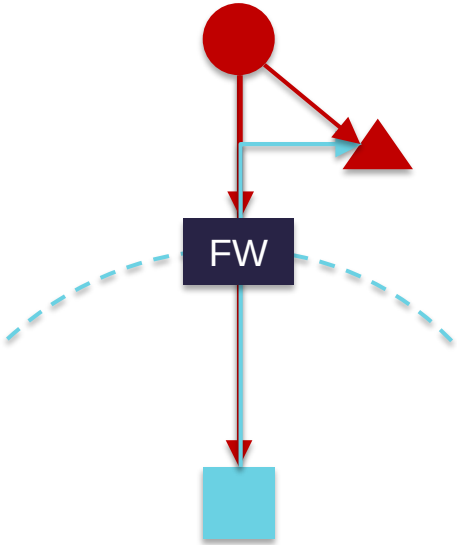
LET'S APPLY VISIBILITY AND ML TO THE REAL WORLD, THINGS THAT IMPACTED EQUIFAX

Attacker behaviors:



Botnet Monetization	Command and Control	Reconnaissance	Lateral Movement	Exfiltration
Abnormal Web Activity	External Remote Access	Internal Darknet Scan	Automated Replication	Data Smuggler
Bitcoin Mining	Hidden DNS Tunnel	Kerberos Account Scan	Brute-Force Attack	Hidden DNS Tunnel
Brute-Force Attack	Hidden HTTP/S Tunnel	Port Scan	SMB Brute-Force	Hidden HTTP/S Tunnel
Outbound DoS	Malware Update	Port Sweep	Suspicious Kerberos Client	Smash and Grab
Outbound Port Sweep	Peer-to-Peer	File Share Enumeration	Suspicious Kerberos Account	Staged Transfer
Outbound Spam	Pulling Instructions	SMB Account Scan	Kerberos Server Activity	TOR Activity
	Suspect Domain Activity	*RDP Recon*	Ransomware File Activity	*USB Smuggler*
	Suspicious HTTP	*LDAP Recon*	Shell Knocker	*Slow Bleed*
	TOR Activity	*RPC Recon*	Suspicious Admin	
			Suspicious Remote Exec	
			Suspicious Remote Desktop	
			SQL Injection Activity	
			Key Data Gathering	

External Remote Access



- Time series data
- Deep learning model
- Discovers the human on the outside taking control

Attacker wants to establish manual control over asset inside the network

Firewalls block most inbound connection attempts

So compromised internal asset calls out to “meeting point” and attacker takes over

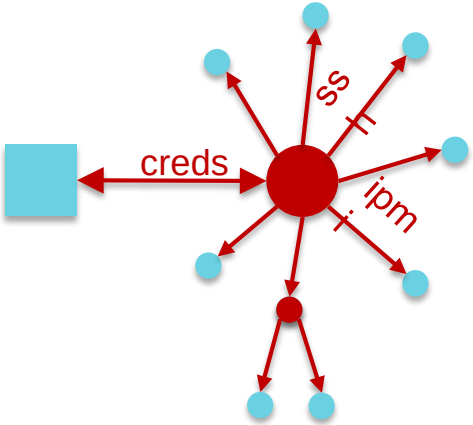
Blackshades
Poison Ivy
NOPEN (Shadow Brokers)
WebEx
TeamViewer

Simply recognizing individual tools’ marks won’t suffice

Data science approach

- Supervised
  - Big data problem
- Key Leverage Points
  - Statistically unique sequences
  - Generalize across tools
  - Offline data store
- Algorithm Implementation
  - Traffic to a multi-dimensional time series
  - Long short-term memory neural network
  - Utilize deep learning to featurize the data flow – not the human dependent

Suspicious Admin



Once administrative credentials have been acquired, attackers like to use administrative network protocols to move laterally in the environment

Administrative protocols include flexible protocols such as SSH, RDP and VNC, but also low-level protocols such as IPMI and AMT

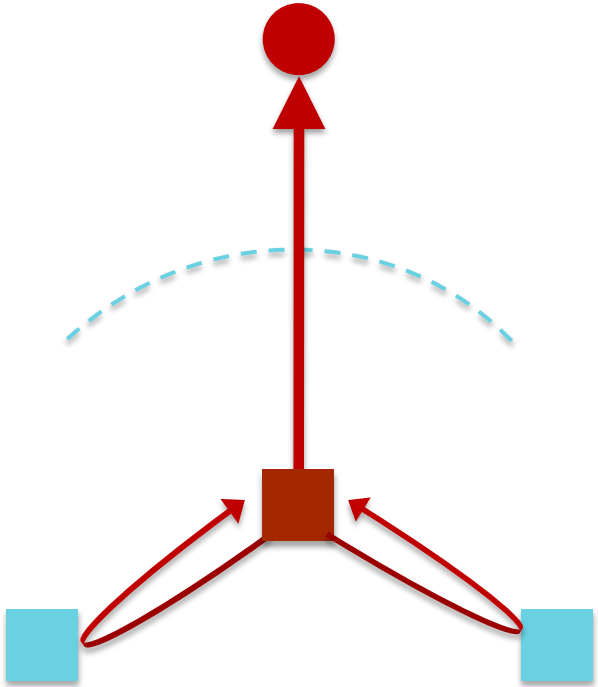
These protocols are attractive because either because they are super-privileged (IPMI) or like Swiss Army knives (SSH)

- Data Source: Admin protocol connection graph over time
- Unsupervised learning model learns admins, servers, realms
- Discovers attacker or malicious insider use of admin capabilities

Data science approach

- Unsupervised
  - No two admins in the same network behave the same
- Key Leverage Points
  - Admins and their "realm" can be represented as a graph
  - Vectra's host-identification system leads to a stable framework
- Algorithm Implementation
  - Create time evolving graphs for "manager-like" connections
  - Classify central nodes are as admins, common resources, scanners, etc.
  - Associate machines to realms managed by an admin
  - Detect when there has been an intrusion into the managed group

Data Smuggler



Attackers will collect data from deep inside the network to a staging area
after which they will exfiltrate the data to outside the network

Exfiltration can be done over any port and many different protocols

“Benign” destinations and new destinations can be used for the same goal

- Time series data
- Correlation model
- Discovers exfiltration events and affected systems

Data science approach

- Unsupervised
  - Black list destinations are not stable
  - Common websites can be used for exfiltration
- Key Leverage Points
  - Correlation of the data pull and data push in time
- Algorithm Implementation
  - Maintain rolling history for all data collected by internal hosts
  - Maintain rolling history for all data sent outside the network by all internal hosts
  - Correlate the the histories in realtime

Apply scoring to detect mayhem

Summary

- You need complete visibility to find the threats
- ML techniques are good at finding attack behaviors that would otherwise go unnoticed in the noise
- AI will reduce the workload of analysts, accelerate threat hunting and incident response

- The hope can be realized, and you can get to < 1

#BeatTheBreach

QUESTIONS?

Thank you