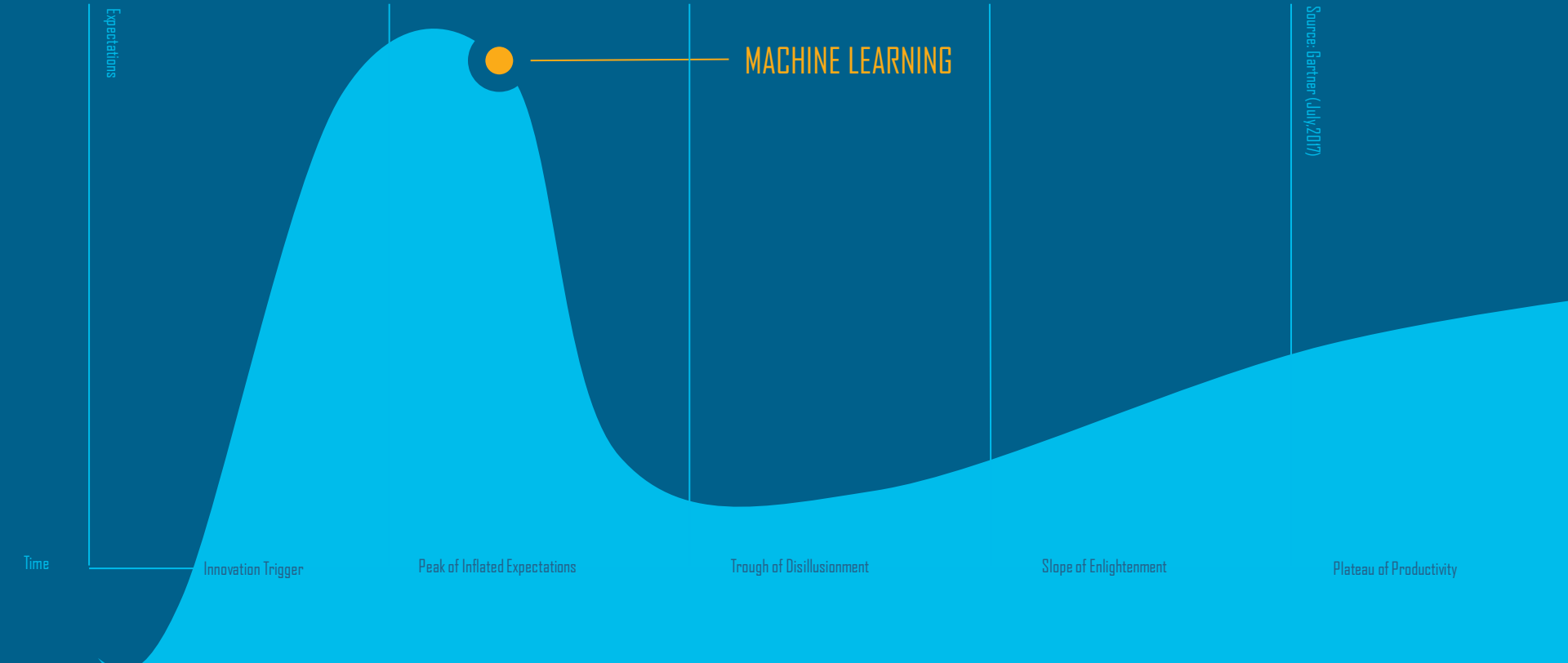
pbeyleve@cisco.com
@RegardingPaul
/paulbeyleveld

Machine Learning

for Cybersecurity in the Digital Era

Paul Beyleveld, CISSP
Consulting Systems Engineer

Gartner Hype Cycle for Emerging Technologies | 2017



Investment in and adoption of ML/AI is dramatically increasing.

$10B

VC investment in ML/AI companies in 2017

PITCHBOOK

75%

Of dev teams building AI functionality

IDC

30%

CIO top 5 investment priority

GARTNER

Today's Situation

Attackers Advantages
Expansive adversary tool box
Exhaustive nonstop attacks
High velocity

Environmental Challenges
Explosive growth of endpoints
Exponential growth of network traffic
Complex application environment

Highly skilled and determined attackers + Ever-growing attack surface

Defender Challenges
Constrained security budgets
Overwhelming volume of alerts
Unrelenting velocity of alerts

Defender Realities
35% report $ as #1 constraint
56% of alerts investigated
46% of legitimate alerts remediated

Constrained Budget, Time, Talent & Ability to respond to threats

Today's Situation

Attackers Advantages
Expansive adversary tool box
Exhaustive nonstop attacks
High velocity

Environmental Challenges
Explosive growth of endpoints
Exponential growth of network traffic
Complex application environment

Highly skilled and determined attackers + Ever-growing attack surface

RISK

Defender Challenges
Constrained security budget
Overwhelming volume of alerts
Unrelenting velocity of alerts

Defender Realities
35% report $ as #1 constraint
56% of alerts investigated
46% of legitimate alerts remediated

Constrained Budget, Time, Talent & Ability to respond to threats

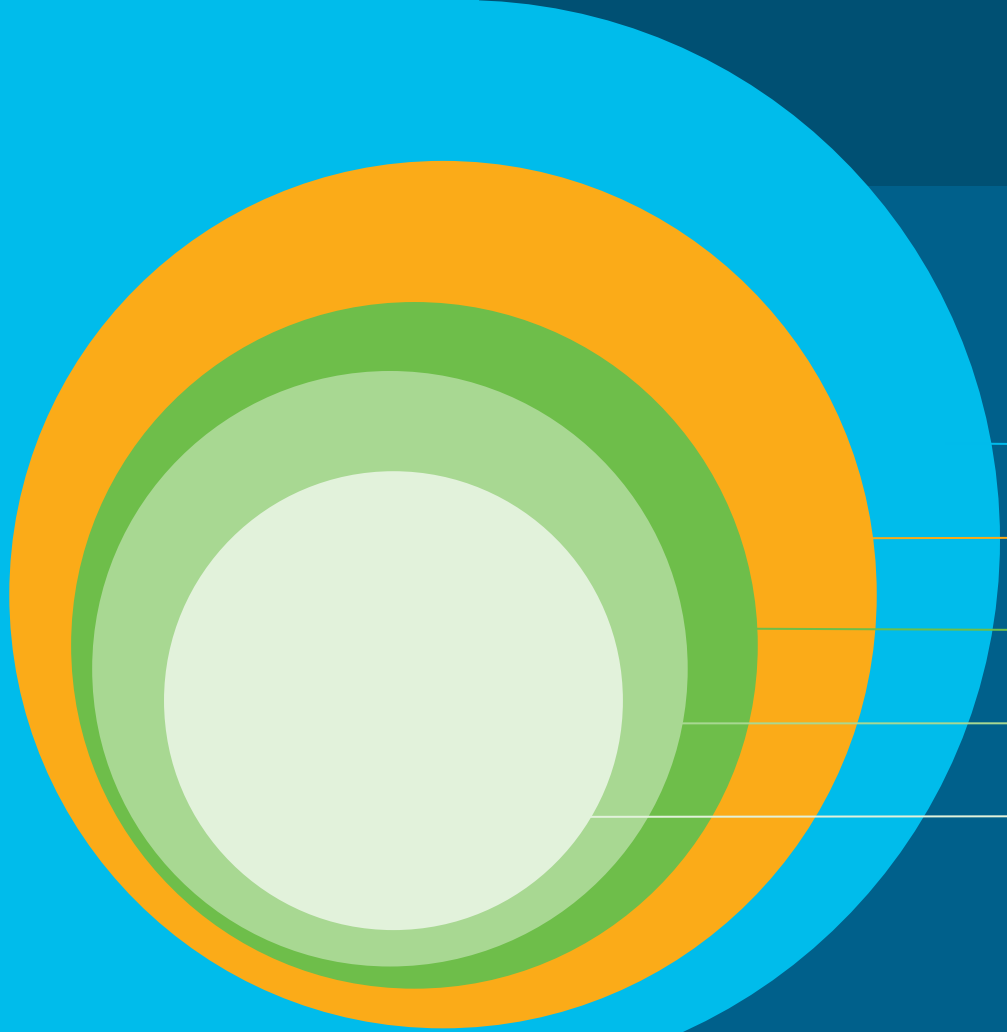
Operationalizing Security

…a **business-driven cybersecurity model**—one that can provide resiliency to increasingly flexible, open enterprises even in the face of highly capable and determined malevolent actors—is starting to emerge. - McKinsey&Company

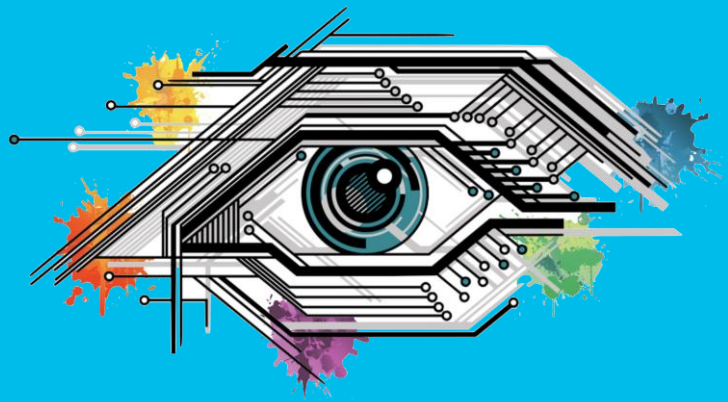
Reduce Adversary's Operational Space

Machine Learning

The Big Picture



- Artificial Intelligence
- Machine Learning
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning



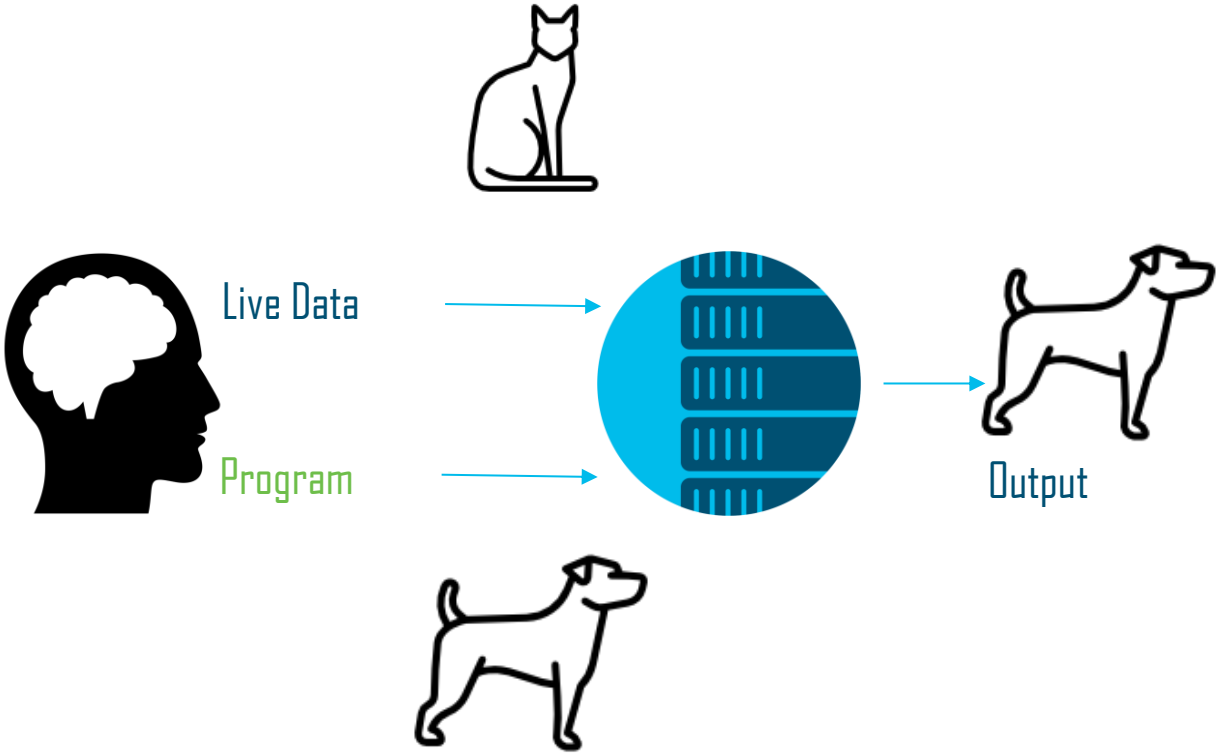
Machine Learning

So what is it?

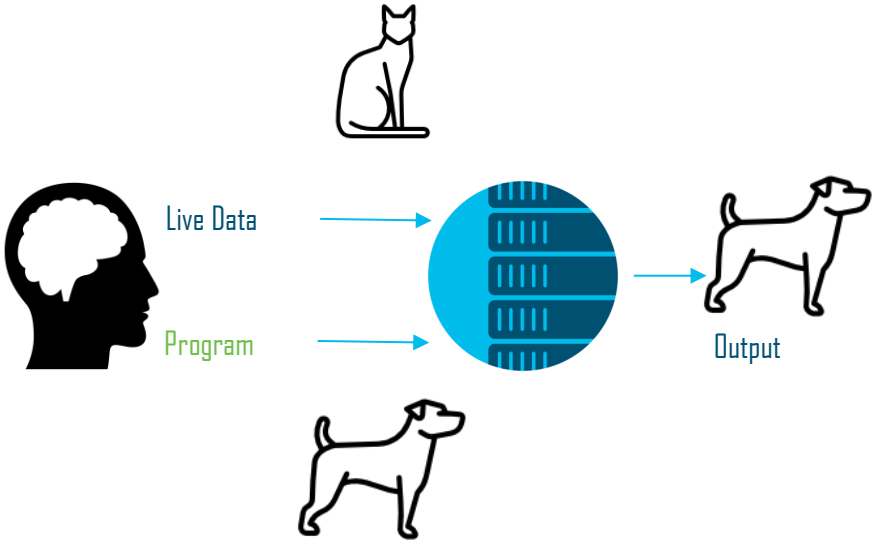
“Field of study that gives computers the ability to learn ***without being explicitly programmed.”***

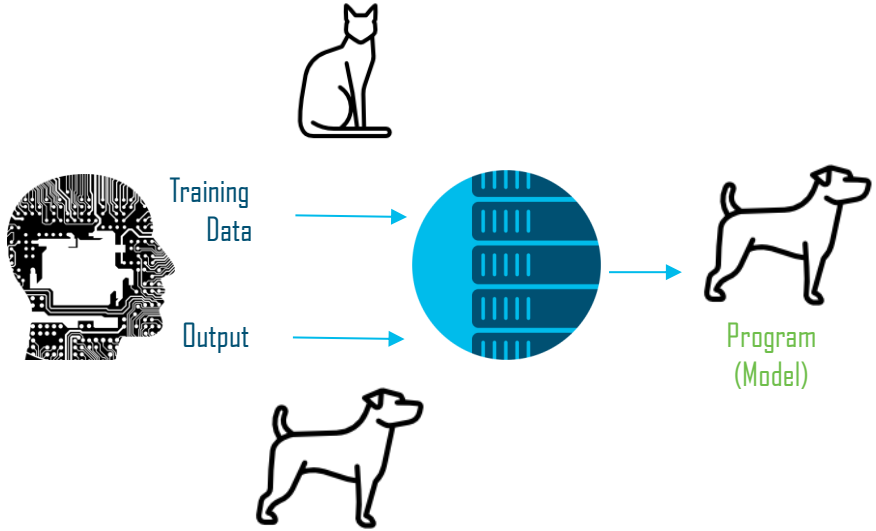
Arthur Samuel’s definition of machine learning in 1959

Traditional programming



Traditional programming vs. machine learning



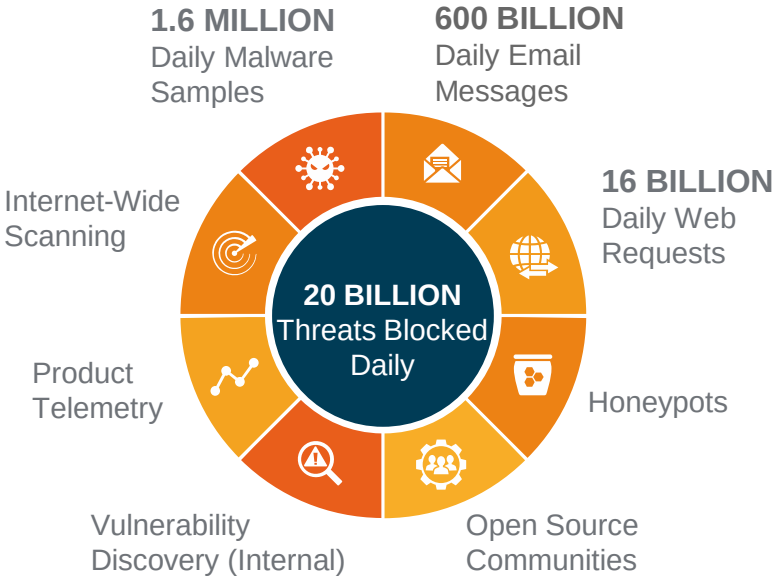


Natural conclusion: Training data is important and we have lots and lots of it.

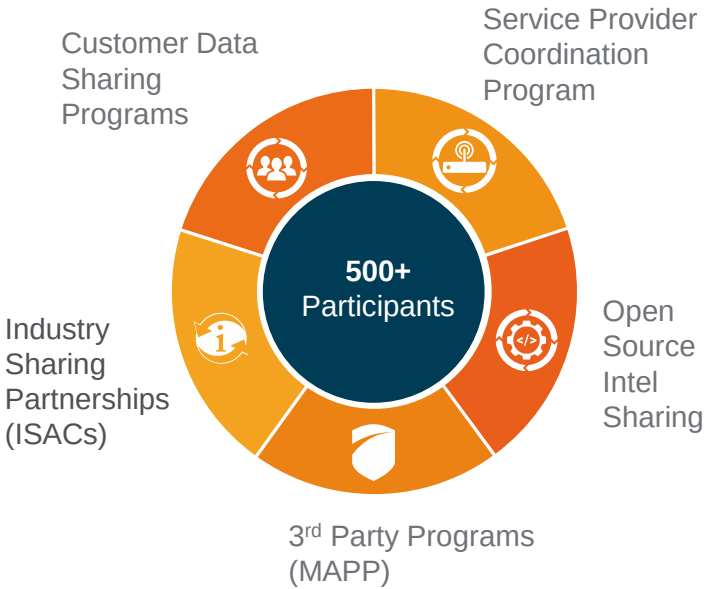
Identifying Encrypted Malware Traffic with Contextual Flow Data
Blake Anderson and David McGrew

Talos INTEL BREAKDOWN

THREAT INTEL



INTEL SHARING



300+
Full Time Threat Intel Researchers

MILLIONS
Of Telemetry Agents

4
Global Data Centers

100+
Threat Intelligence Partners

1100+
Threat Traps

Some of the most brilliant minds in the industry

Cisco Talos Team won round one of the Fake News Challenge

Collab team beat Google's own internal model in Google's TensorFlow Speech Recognition Contest

300+ Engineers working full time on AI/ML

Example: CTA Detection Technology

Which of these encrypted request is malware c&c?

hxxps://136.243.4.68

hxxps://64.233.162.83

Which of these encrypted request is malware…

hxxps://136.243.4.68

- Malware c&c and exfiltration (encrypted)

hxxps://64.233.162.83

- Google Mail (encrypted)

Which of these is more legit?

hxxps://google.com

hxxps://llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogoch.co.uk

Which of these is more legit?

hxxps://google.com

- Used commonly by malware – connectivity check

hxxps://llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogoch.co.uk

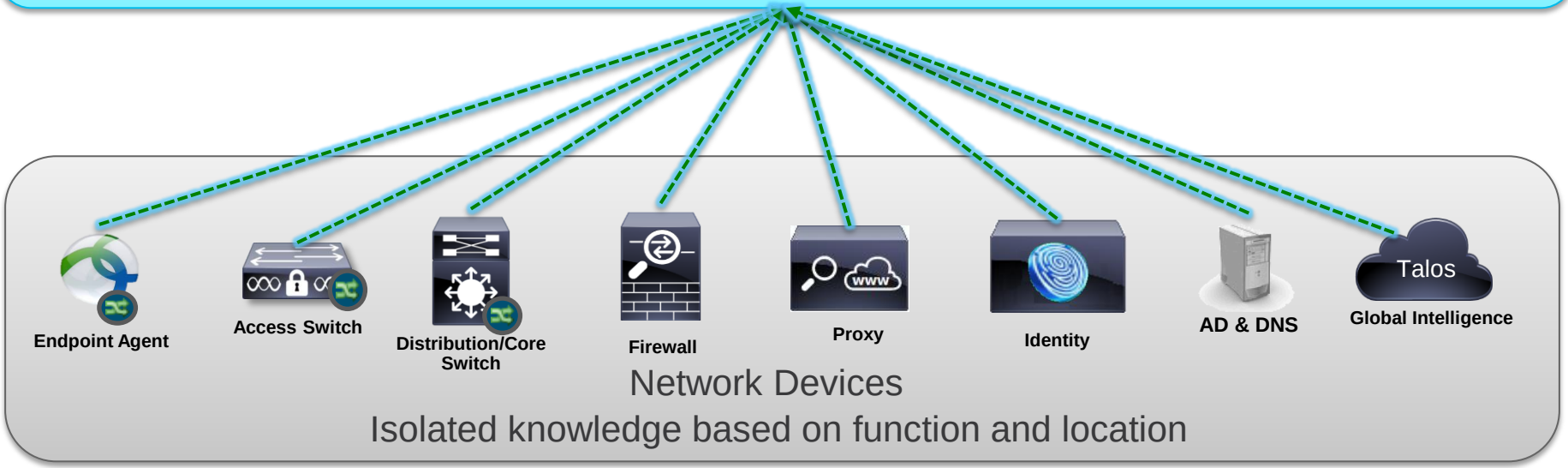
- Welsh village with the *longest* name in Britain

Value of AI and ML in Cybersecurity?

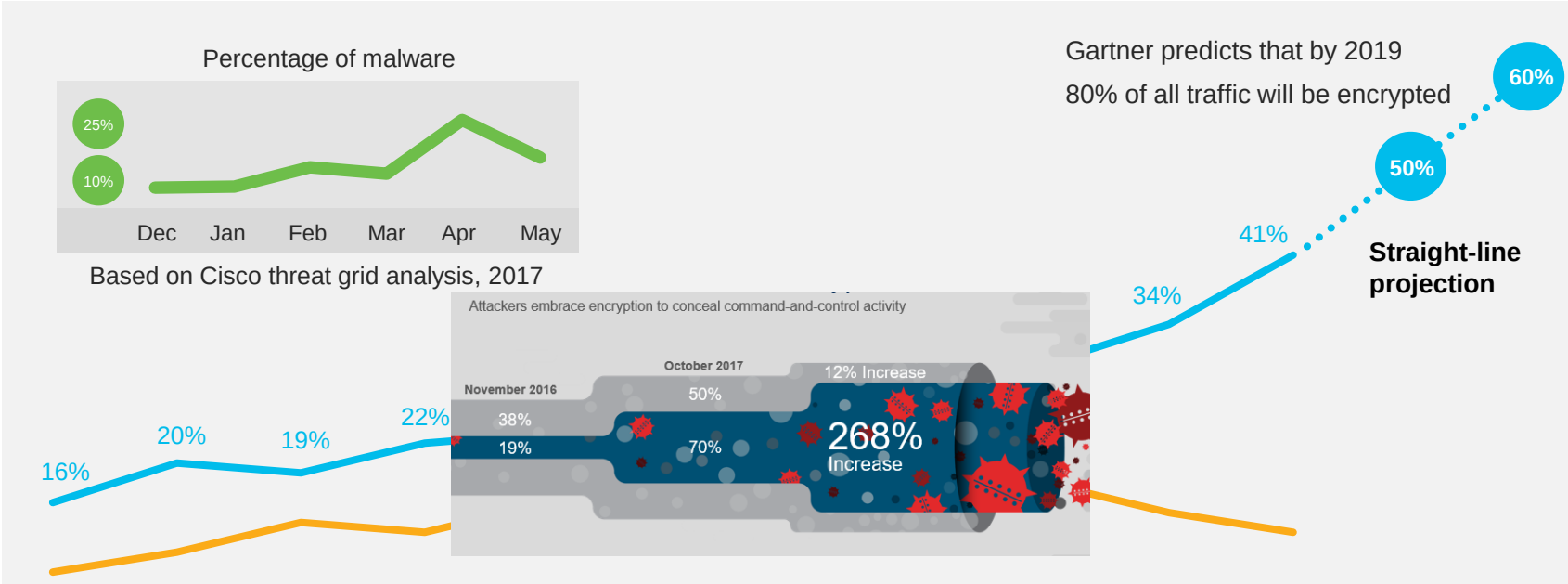
- AI/ML systems analyse internal security data, and correlate it with external threat data - **points human analysts to needles in the haystack**.
- **System adapts its monitoring and analysis based on human input**, optimizes likelihood of finding real cyber threats & minimizing false positives
- Enables Behavioural Analytics to **understand normal behavior** by analysing data over time to **identify anomalous and suspicious behaviors** that deviate from normal baseline
- **Enables analysts to focus on more advanced threat investigation** rather than performing tactical data crunching.

Network Telemetry

Telemetry: an automated communications process by which measurements and other data are collected at remote or inaccessible points and transmitted to receiving equipment for monitoring.
https://en.wikipedia.org/wiki/Telemetry



Encryption is changing the threat landscape



Source: Thales and Vormetric

What is Mobile World Congress?

GSMA MOBILE WORLD CONGRESS

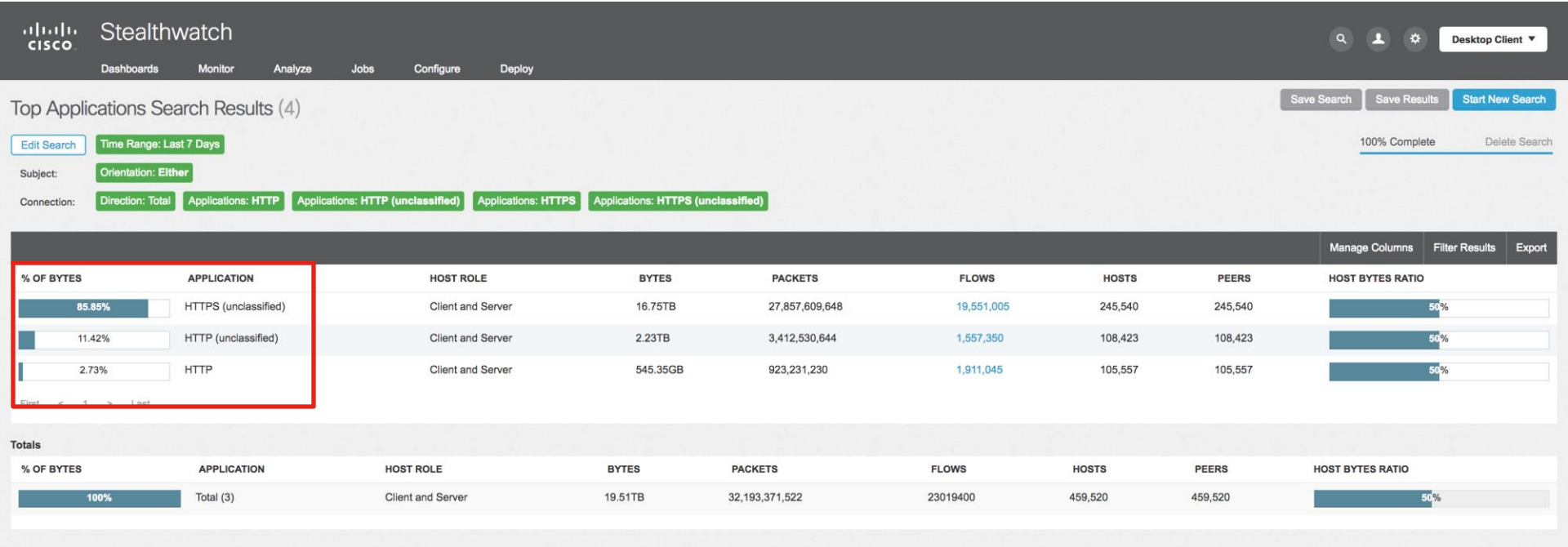
BARCELONA 26 FEB-1 MAR 2018

More than 107,000 visitors from 205 countries and territories

Over 55% of attendees held senior-level positions, including more than 7,700 CEOs

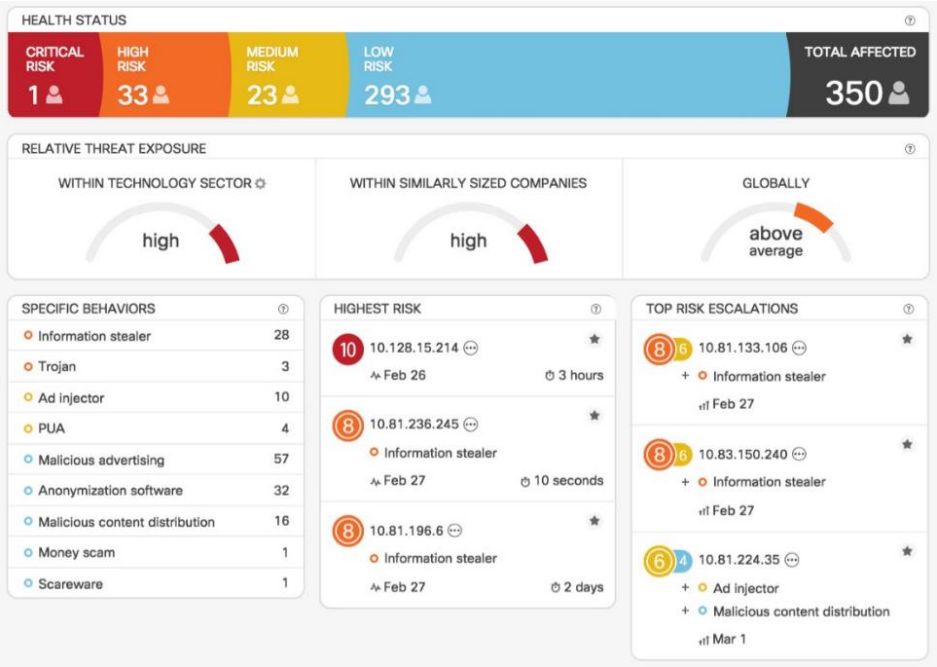
Stealthwatch monitored all the wireless traffic to and from the Internet with Encrypted Traffic Analytics

MWC 2018: HTTPS:HTTP on the show floor



82% HTTPS vs 18% HTTP traffic mix

Detection on 2/26 and 2/27



Global Threat Analytics raised 350 events

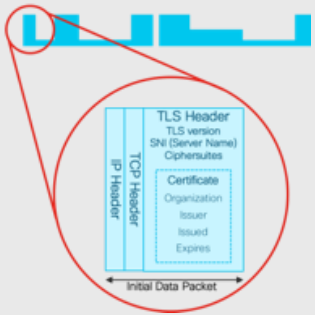
Cryptomining
Android Trojans (Android.spy, Boqx, infected firmware)
SALITY malware
SMB Service discovery malware
OSX Malware Genieo
Conficker
RevMob
Phishing
AdInjectors
Several Android mobile devices were identified to have an infected firmware

Malware Trojans were identified that were **using PowerShell to communicate to the C&C servers through HTTPS**.
Several malwares / potentially unwanted applications that **used Encrypted traffic**

** Over 13,500 alarms in Stealthwatch on 2/26*
Over 18,500 alarms on 2/27

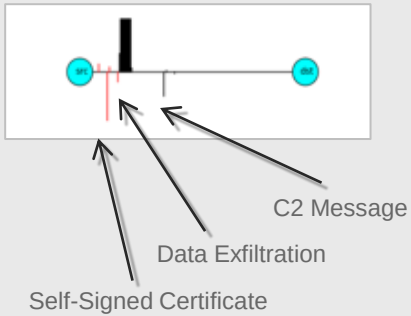
Detect malware in encrypted traffic

Initial data packet



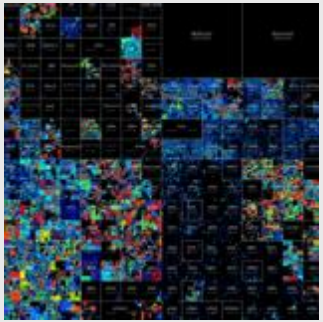
Make the most of the unencrypted fields

Sequence of packet lengths and times



Identify the content type through the size and timing of packets

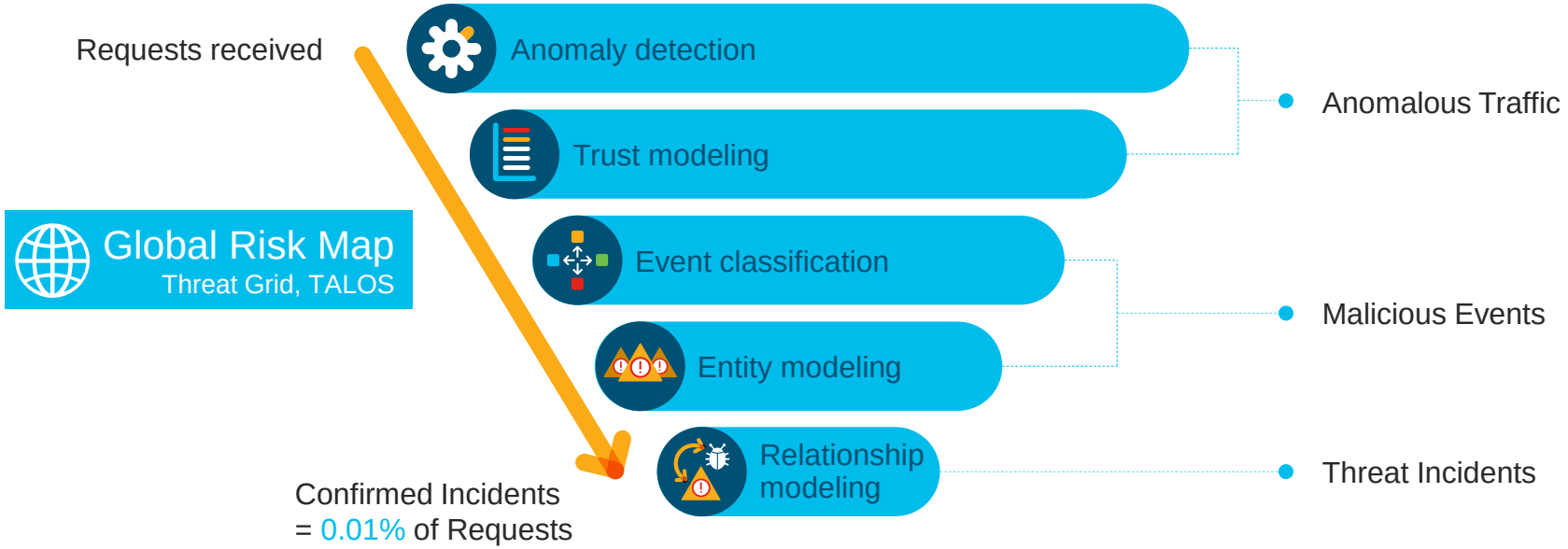
Global Risk Map



Know who's who of the Internet's dark side

Power of multi-layer machine learning

Increase fidelity of detection using best-in-class security analytics



Summary

- Machine learning is very effective for reducing incident count to high fidelity threat incidents
- Coupled with context increases ability to detect and reduce the time to respond
- Machine learning is not a silver bullet but will help to reduce advisory operational space
- **Artificial Intelligence** and **Machine Learning** deliver **intent-based networking** to enable the entire network to contribute to the cybersecurity fight

Questions?

pbeyleve@cisco.com

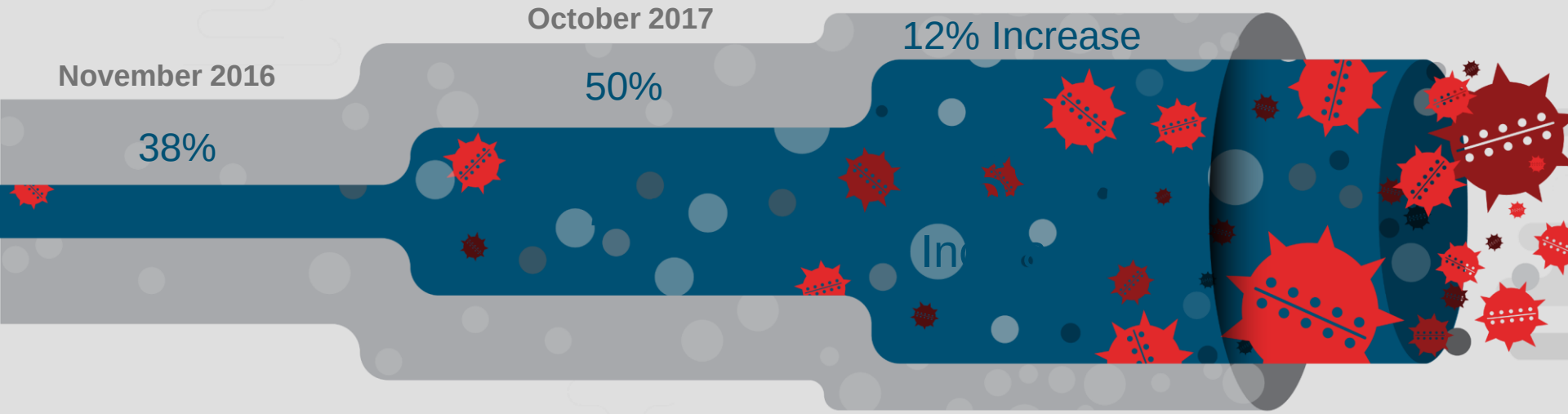
@RegardingPaul

/paulbeyleveld


CISCO

Malicious Activity and Encryption

Attackers embrace encryption to conceal command-and-control activity



Global Encrypted Web Traffic

Malicious Sandbox Binaries with Encryption