RSAConference2019

San Francisco | March 4 – 8 | Moscone Center

BETTER.

SESSION ID: **SPO3-T06**

Harnessing the Law of Data Gravity: Cyber Defense and the Hybrid Cloud

Sian John MBE

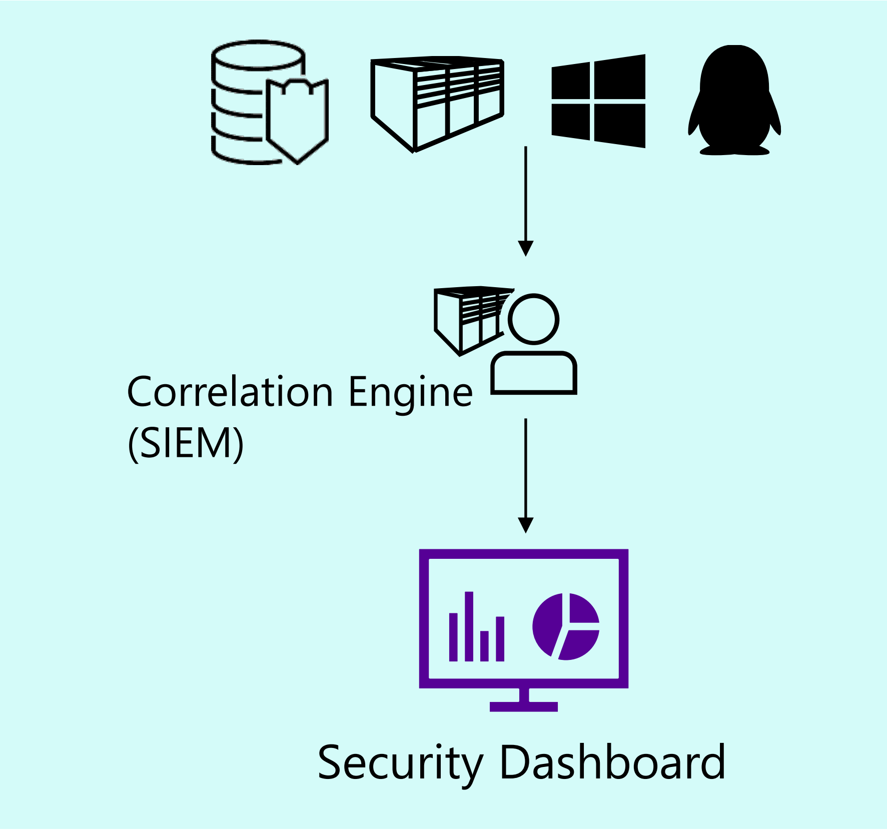
Chief Security Advisor
Microsoft, Cybersecurity Solutions Group
@sbj24

Diana Kelley

Cybersecurity Field CTO
Microsoft, Cybersecurity Solutions Group
@dianakelley14

The SOC has evolved

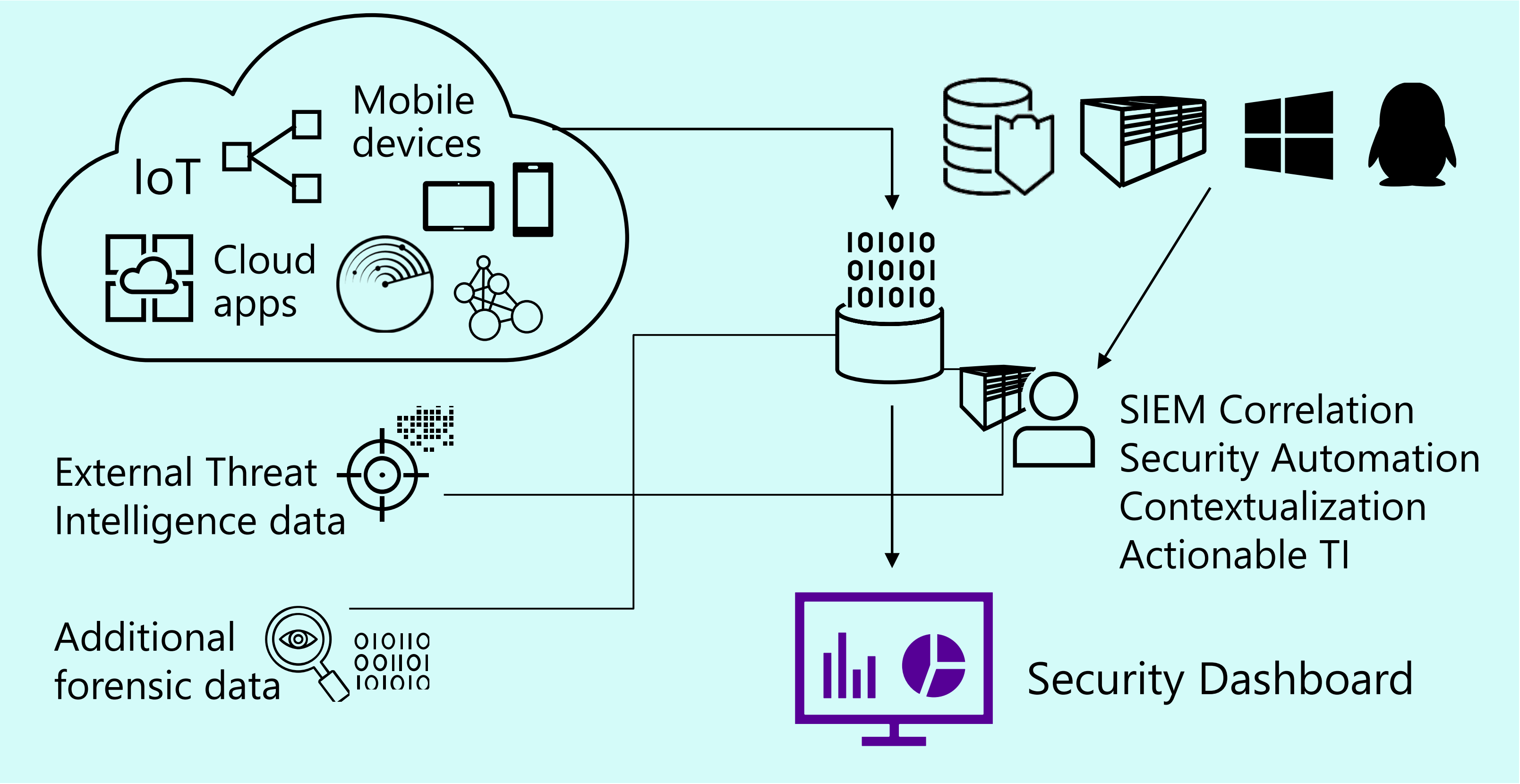
Then...



Correlation Engine (SIEM)

Security Dashboard

Now...



IoT

Mobile devices

Cloud apps

External Threat Intelligence data

Additional forensic data

SIEM Correlation
Security Automation
Contextualization
Actionable TI

Security Dashboard

How can we find the signal in the noise?

Tie together disconnected systems?

Create meaningful insights?

Fully integrate the cloud w/traditional on-prem model?

Normalize/ harmonize logs and metadata?

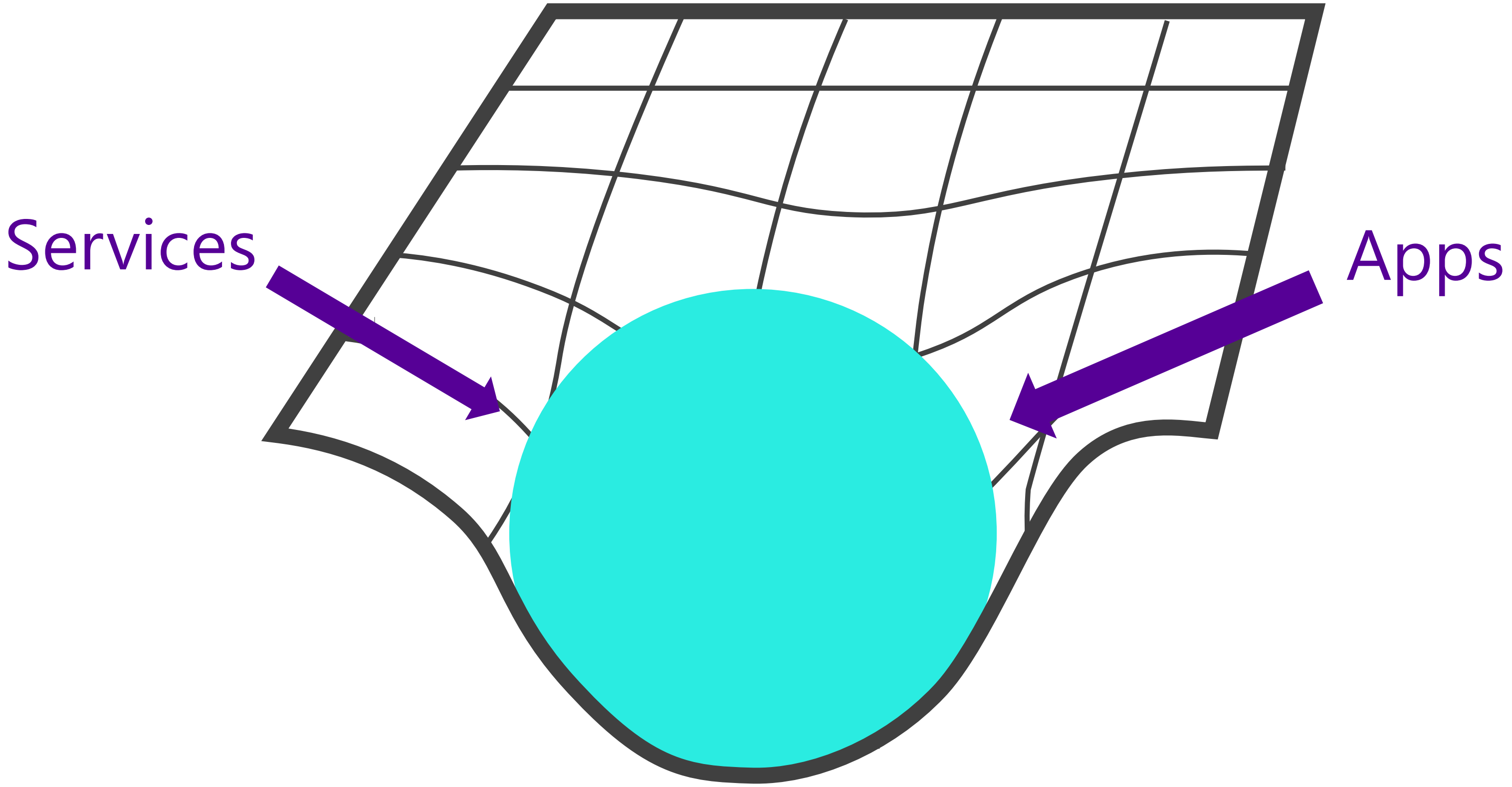
Break out of the brittle rule trap?

The concept of data gravity

Data Gravity

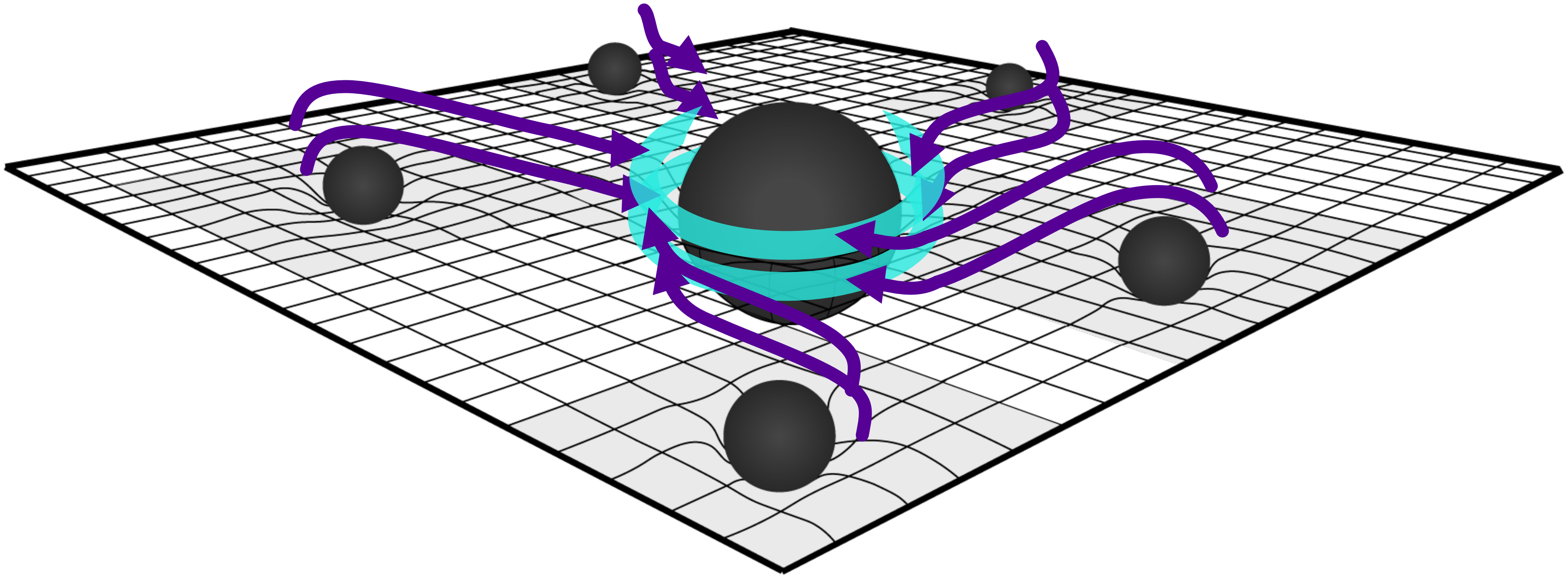
$$\frac{\left(\text{Data Mass} \times \text{Application Mass}\right) \times \text{Number of Requests per second}}{\left(\text{Latency in seconds} + \left(\text{Average Request Size in MBs} \,/\, \text{Bandwidth in MBs per second}\right)\right)^2}$$

McCrory’s Original Equation



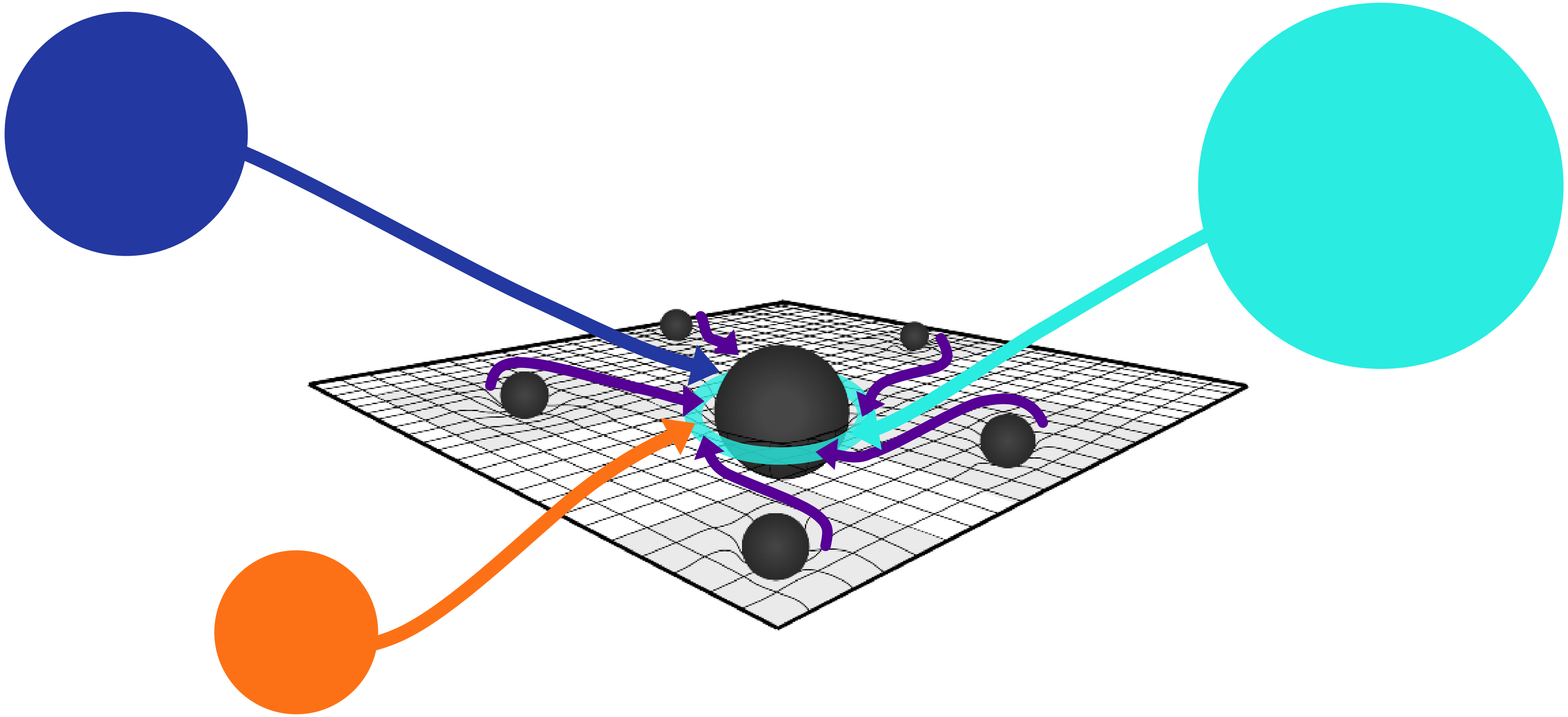
Data gravity in security

Analytics and monitoring
gravitates towards the data



Enabling SIEM transformation

Allows analysts to get insights and context across local and cloud hosted data gravity wells



Future SOC
Data gravity + machine learning

Strong governance across all layers of operations

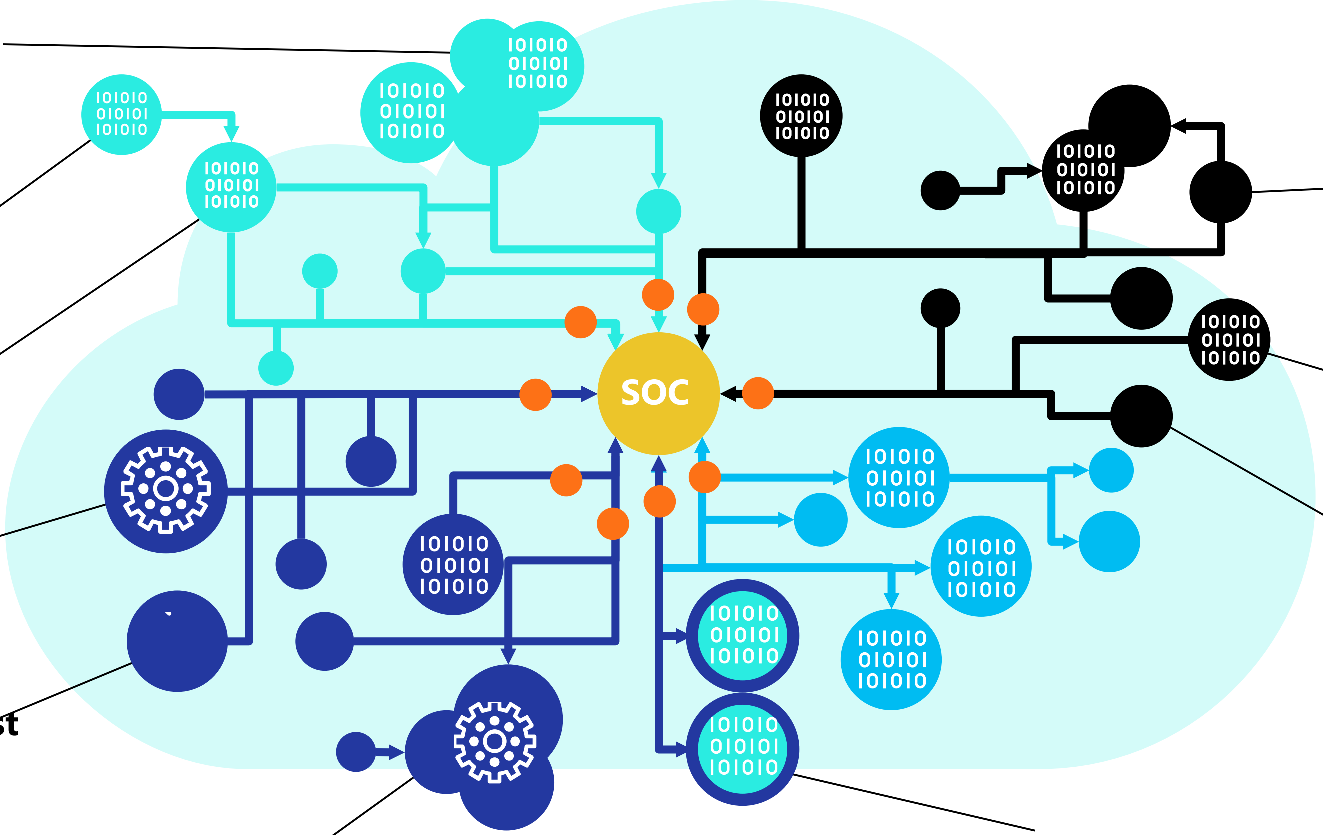
Advanced **contextualization** based analytics on wide range of security **data and telemetry**

Matrix driven approach for **operations performance measurement**

Capability to process **threat intelligence** data into **actionable** content, reduce noise

Ability to investigate **every single** alert generated

Ability to **detect most** of the attacks in the cyber kill chain



Leverage security automation and security orchestration

Automated and **intelligent workflow** to minimize human errors

Security **alert optimization** to reduce alert fatigue

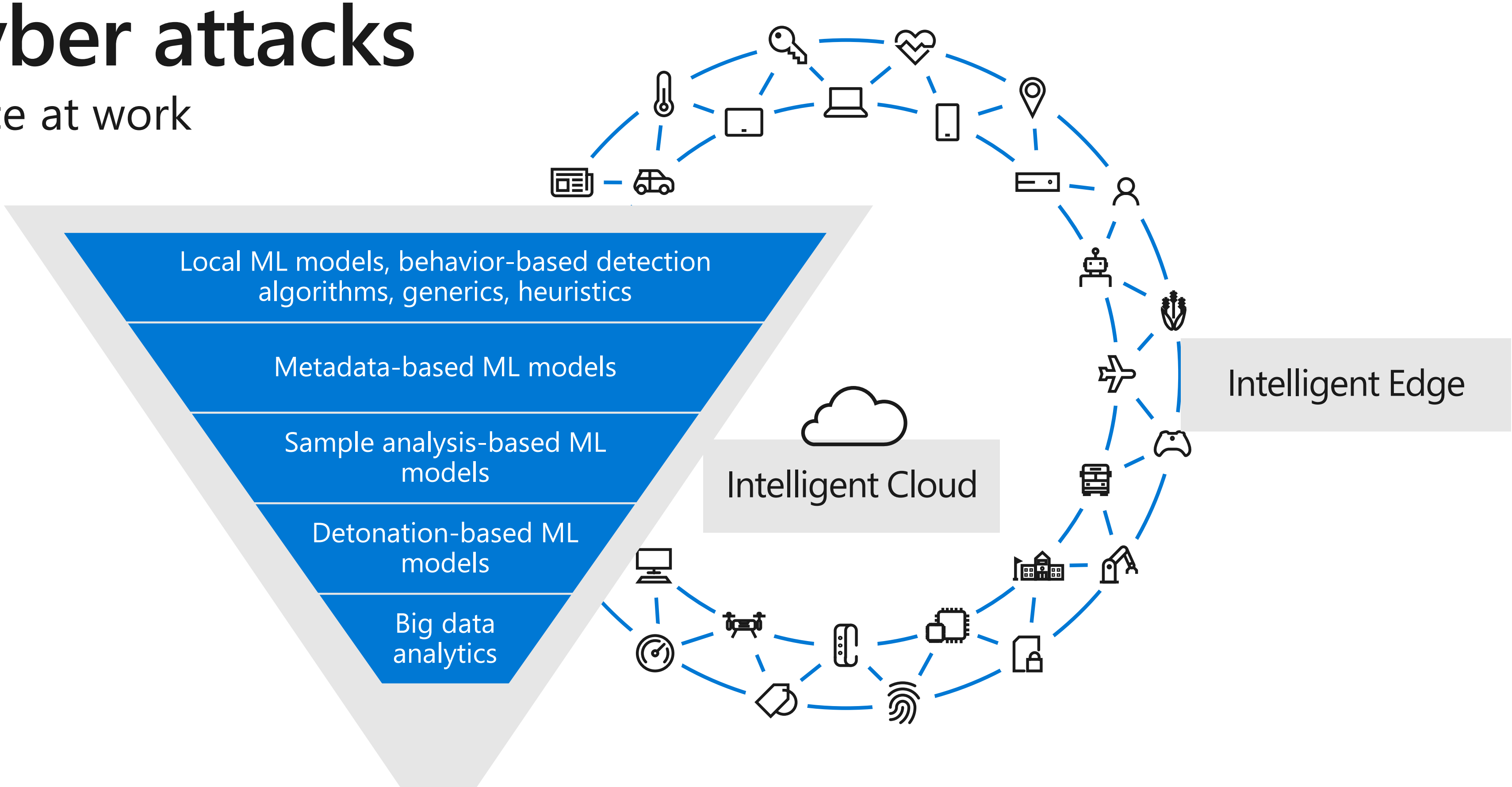
Proactive collection of **forensic evidence**, provide context to the alerts

Optimized operational capability by **continuous automation** and improvement

This Photo by Unknown Author is licensed under CC BY-SA

Stopping cyber attacks

Real-world intelligence at work


Local ML models, behavior-based detection algorithms, generics, heuristics
Metadata-based ML models
Sample analysis-based ML models
Detonation-based ML models
Big data analytics
Intelligent Cloud
Intelligent Edge

October 2017 – Cloud-based detonation ML models identified Bad Rabbit, protecting users 14 minutes after the first encounter.

February 3 – Client machine learning algorithms automatically stopped the malware attack Emotet in real time.

March 6 – Behavior-based detection algorithms blocked more than 400,000 instances of the Dofoil trojan.

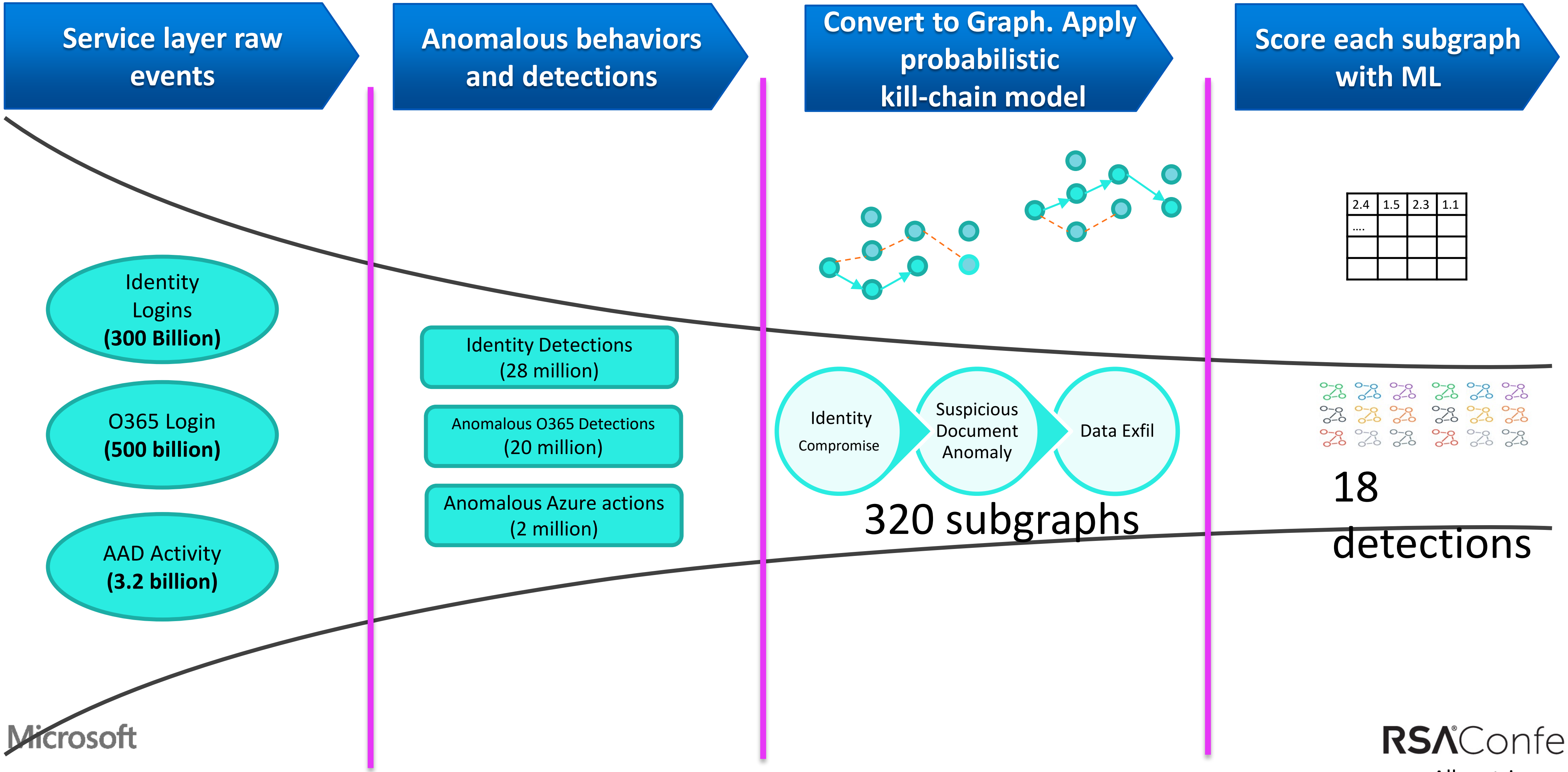
August 2018 – Cloud machine learning algorithms blocked a highly targeted campaign to deliver Ursnif malware to under 200 targets

2017

2018

For a given scenario…

Compromise identity → Suspicious document → Exfiltrate data



Service layer raw events

Anomalous behaviors and detections

Convert to Graph. Apply probabilistic kill-chain model

Score each subgraph with ML

Identity Logins **(300 Billion)**

O365 Login **(500 billion)**

AAD Activity **(3.2 billion)**

Identity Detections (28 million)

Anomalous O365 Detections (20 million)

Anomalous Azure actions (2 million)

Identity Compromise

Suspicious Document Anomaly

Data Exfil

320 subgraphs

2.4	1.5	2.3	1.1
….			

18 detections

All metrics are for a month

Compound Detection

- Service/Component Detections
- Generic Anomaly Detection
- ML Algorithm
- Domain Knowledge
- Products/ Infrastructure

Azure Actions (3.2 billion) → Anomalous AAD Admin Actions (2 million)

AAD Admin Actions (4.1 billion) → Anomalous Azure Actions (20 millon)

AAD Logins (300 billion) → Identity Protection Detections Low + Med + High (28 million)

Storage logs (60 billion) → Storage Anomaly Detections Low + Med + High (…)

KeyVault Logs (21 billion) → KeyVault Detections Low + Med + High (…)

AIP Logs, client & server (24 million) → Azure Information Protection Detections, Med + High (30K)

Windows events (…) → Windows Detections (…)

O365 logs (API) (…) → Anomalous O365 Actions

… →

Anomalous Behaviors/ Service layer Detections

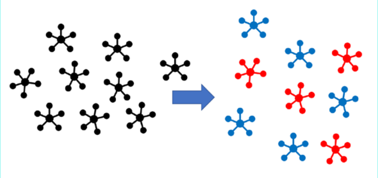
- Anomalous AAD Admin Actions (2 million)
- Anomalous Azure Actions (20 millon)
- Identity Protection Detections Low + Med + High (28 million)
- Storage Anomaly Detections Low + Med + High (…)
- KeyVault Detections Low + Med + High (…)
- Azure Information Protection Detections, Med + High (30K)
- Windows Detections (…)
- Anomalous O365 Actions

Graph Building and Probabilistic Scoring

Full graph of cross service detections & behaviors

- Map of detections and anomalies to kill chain
- Connectivity Calculation Probabilistic Graph Walks
- Sub-graph generation for every attack scenario (in thousands)

Sub-graph scoring



- Unsupervised Graph Scoring (Hundreds)
- Supervised Graph Scoring (dozens of detections)

Azure Sentinel

Microsoft

RSAConference2019

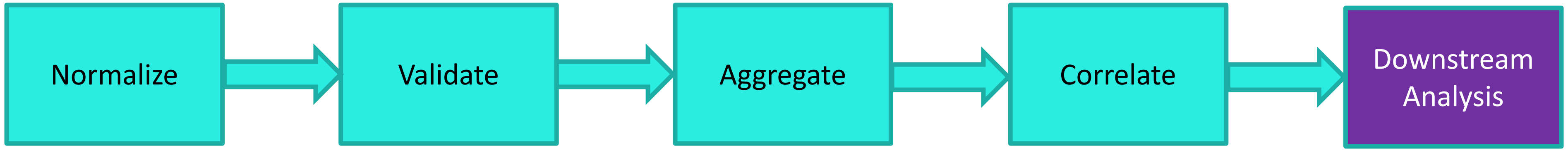
All stats are for a month

Traditional Rule Based Engines

🔴 High Risk

🟡 Medium Risk

🟢 Low Risk



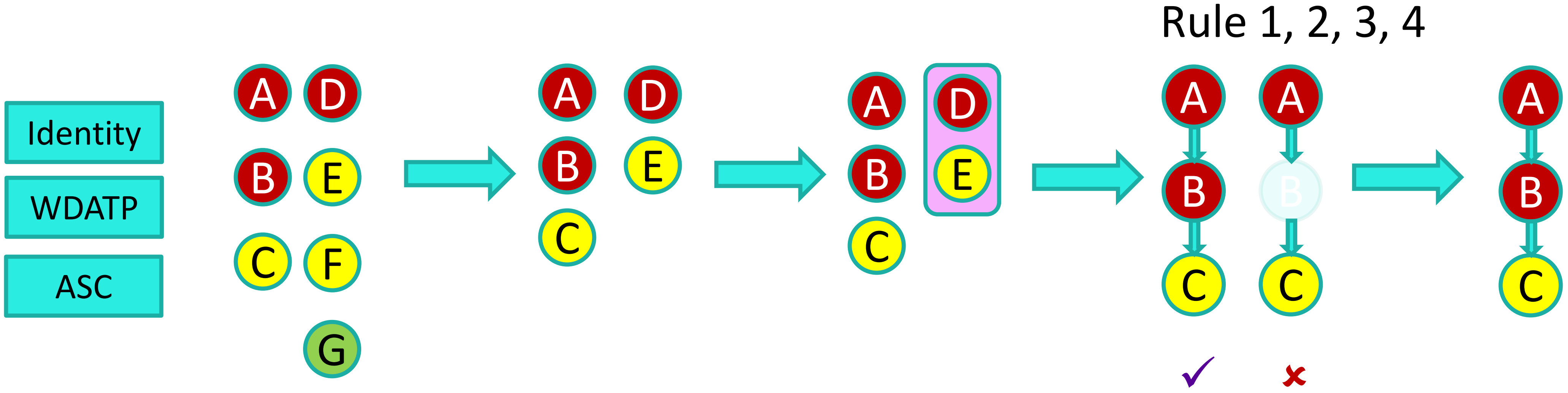
- Standardized schema

- Custom logic for handling false positives, throttling data load

- Reduce complexity of space, eliminate common events

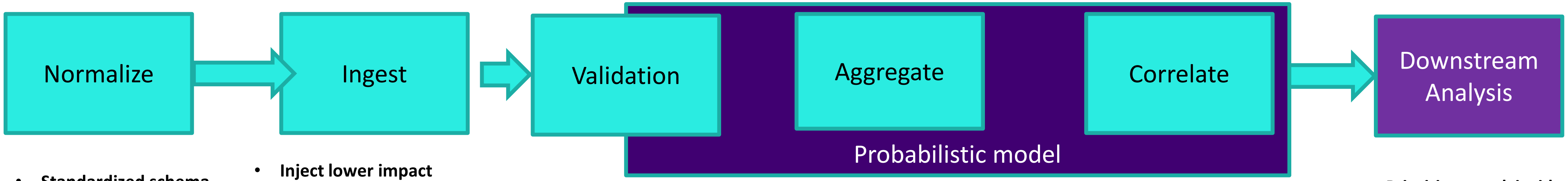
- Rule based approach for joining known data

- Prioritize attack incidents
- Predict Known attacks
- Detect novel attack strategy
- Find missing attacks
- Find similar attacks



Graph based Machine Learning

- ● High Risk
- ● Medium Risk
- ● Low Risk
- ● High Impact Activities

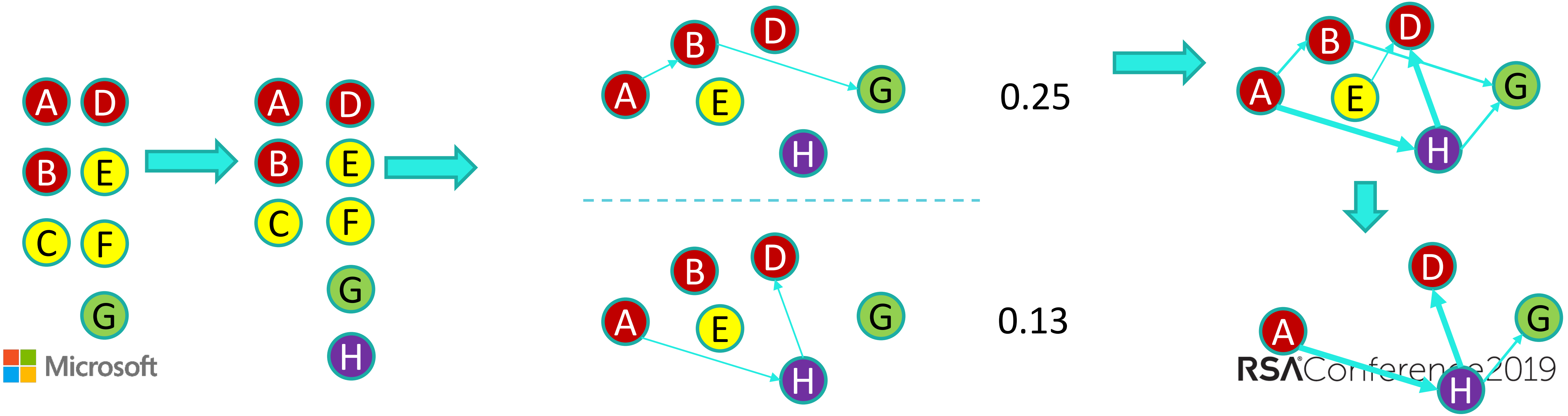


- Standardized schema

- Inject lower impact events, and other high impact activities
- Validate HIA lightly, push other validation downstream

- Probabilistic Bayesian inference, to validate, aggregate, and correlation behavior between alerts and High Impact activity
- Iterative Reversible Jump MCMC algorithm to expand, contract graph, calculate probability between events, and aggregated events
- Output continuous probabilities between edges, and hyper

- Prioritize attack incidents (Probabilistic)
- Predict Known attacks (Probabilistic)
- Detect novel attack strategy
- Find missing attacks
- Find similar attacks

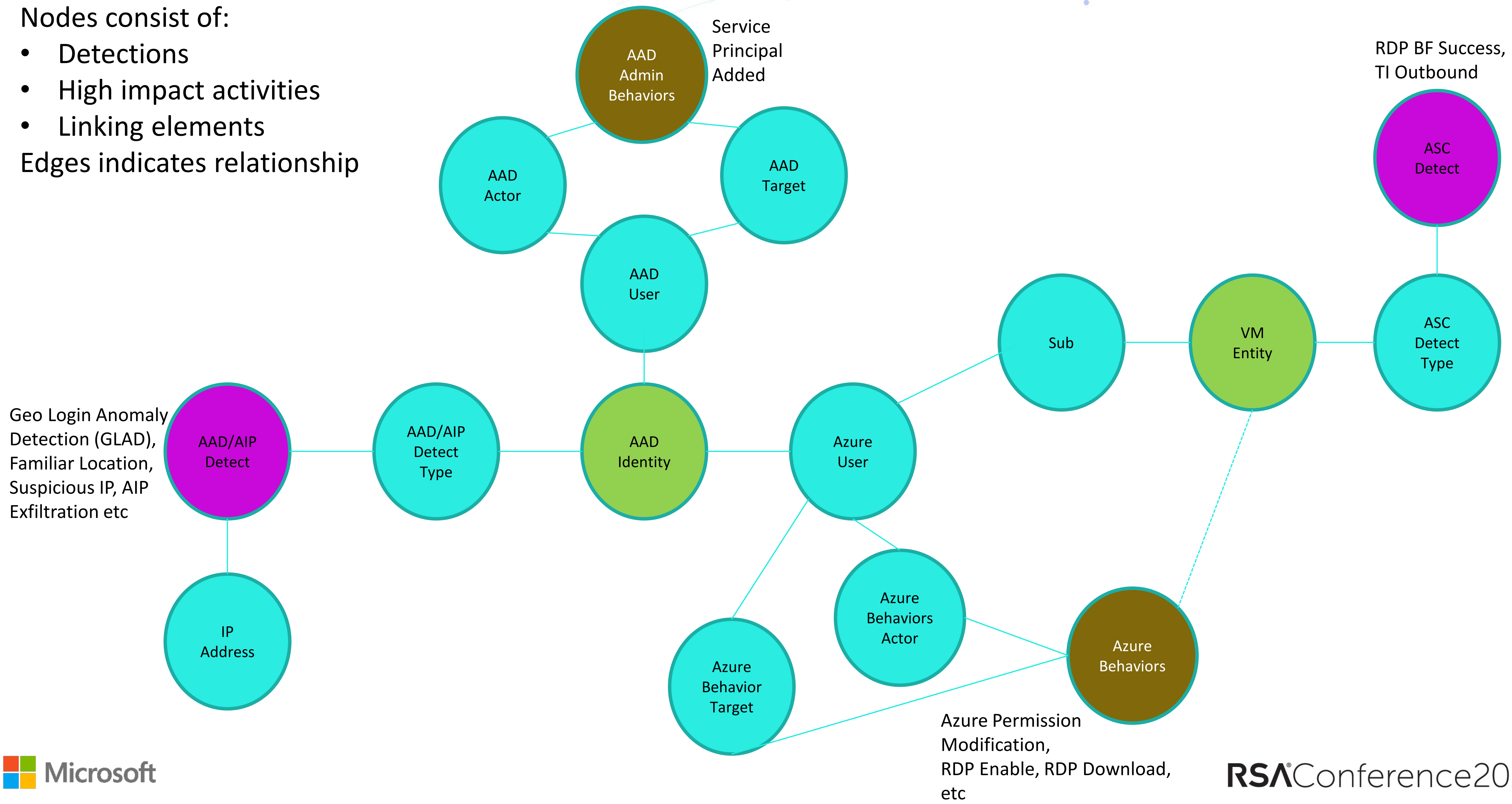


Building a Cross Service Graph of Detections and Behaviors

Nodes consist of:

- Detections
- High impact activities
- Linking elements

Edges indicates relationship



Real World Proof

Noisy results

- Company proxy
- Cell phone networks
- Vacations/travel

A former rules-based Microsoft system scored

2.8% of logins as suspicious

1 billion logins per day =

280 million

"suspicious" logins

After applying **machine learning with rules,** the rate dropped to less than

0.01%

Work by Mace et. al, Microsoft

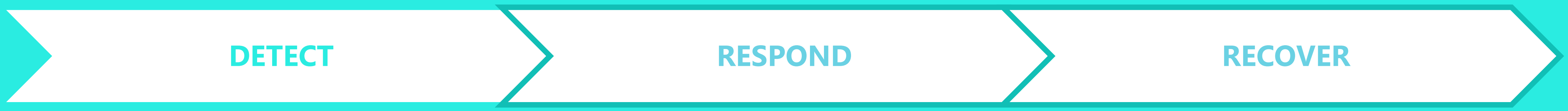
Benefits

Maximize visibility

Reduce manual steps and errors

Maximize human impact

SPEED THE MTTI/MTTC



Observe Orient Decide Act

Summary

- SIEM and traditional SOCs can’t keep up
- We need to reimagine our response
- Harnessing the law of data gravity helps move us to a CDOC model
- Informed and augmented with layered ML

Apply What You Have Learned Today

- Next week you should:
  - Assess your current SOC, can it keep up?
- In the first three months following this presentation you should:
  - Determine SOC requirements for the next 1-3 years
    - Data collection, multi-cloud, multi-partner, containers & functions
    - Consider applying Data Gravity concept to evolved SOC planning
- Within six months you should:
  - Build the strategy for Future SOC
  - Deploy in functional buckets, single cloud before broader roll-out

RSAConference2019

Thank you!