IBM Security

AI for Cyber Security and Adversarial AI

Domenico Raguseo
@domenicoraguseo

June 2019

IBM

Let’s focus on the most critical security use cases

Outcome-driven security

Prove Compliance

- Get Ahead of Compliance
- Enhance Security Hygiene
- Govern Users and Identities

Stop Threats

- Detect & Stop Advanced Threats
- Orchestrate Incident Response
- Master Threat Hunting

Grow Business

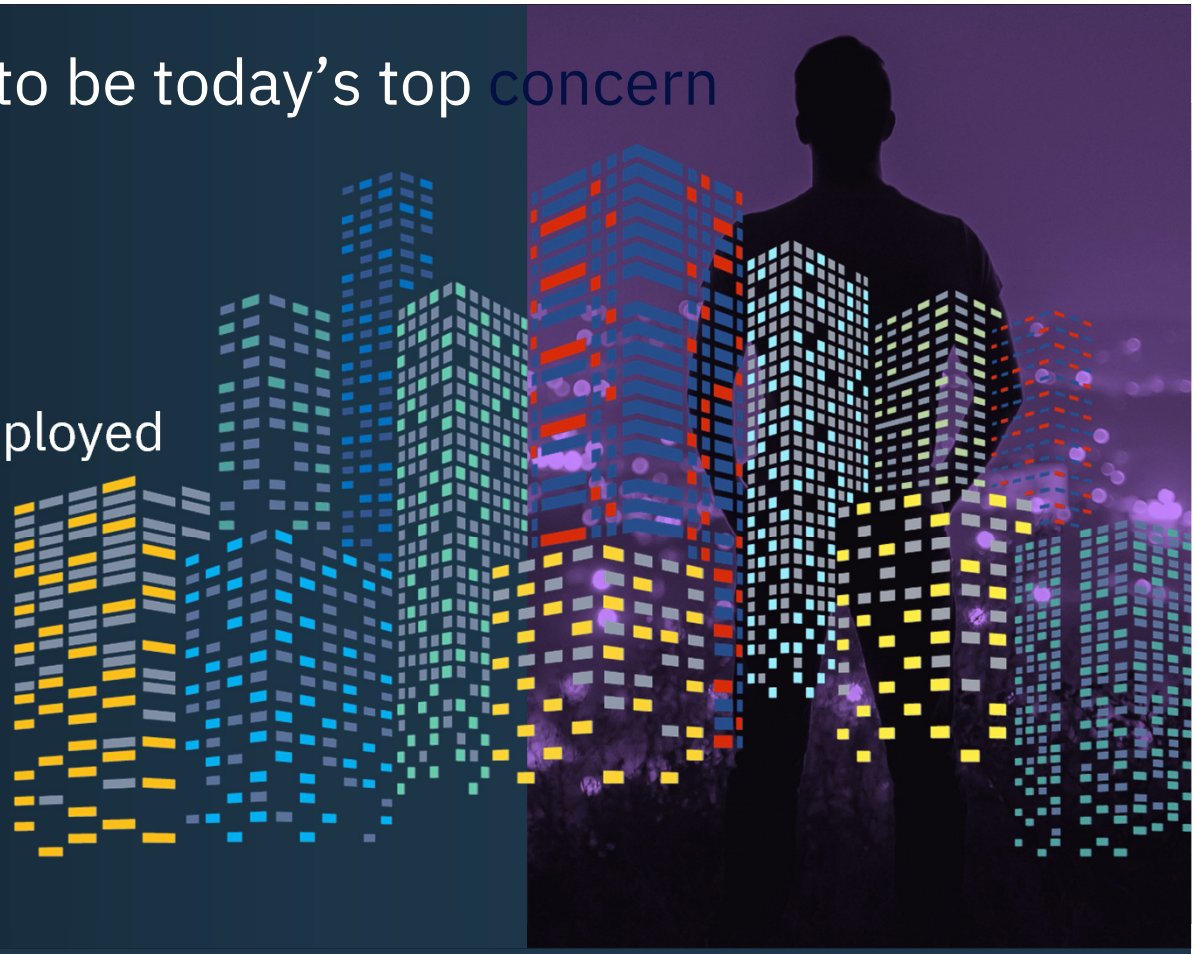
- Secure Hybrid Cloud
- Protect Critical Assets
- Prevent Advanced Fraud

Complexity continues to be today's top concern

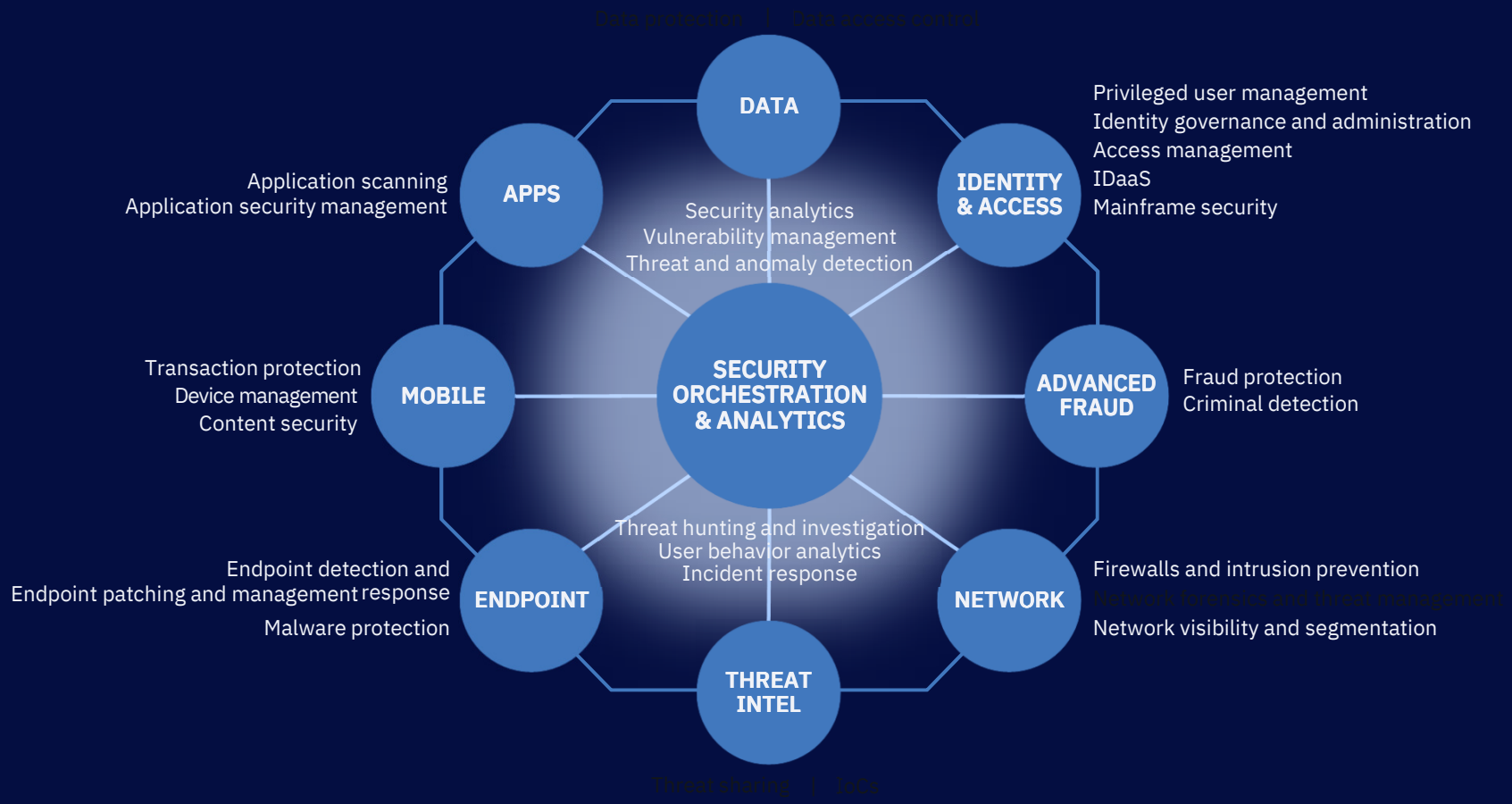
Reactive strategies driven by threats

Products and processes deployed in silos

Organizations and teams continue to work alone

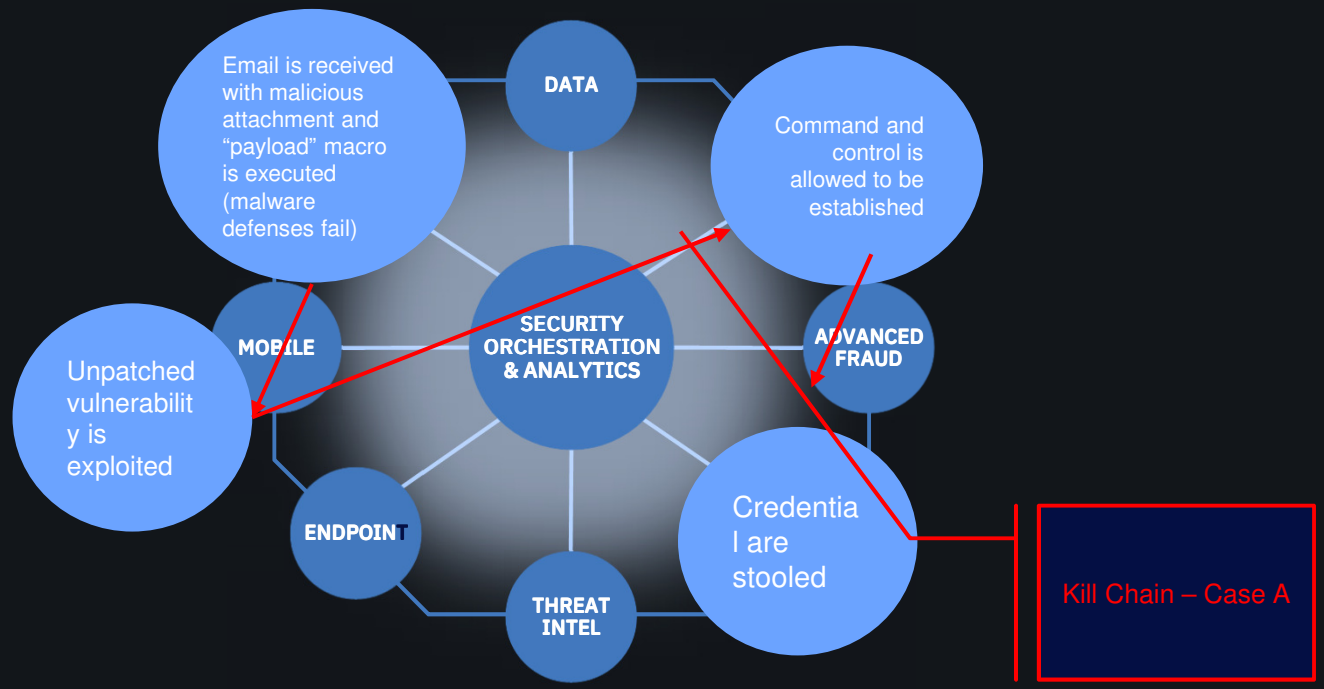


Build an integrated security immune system

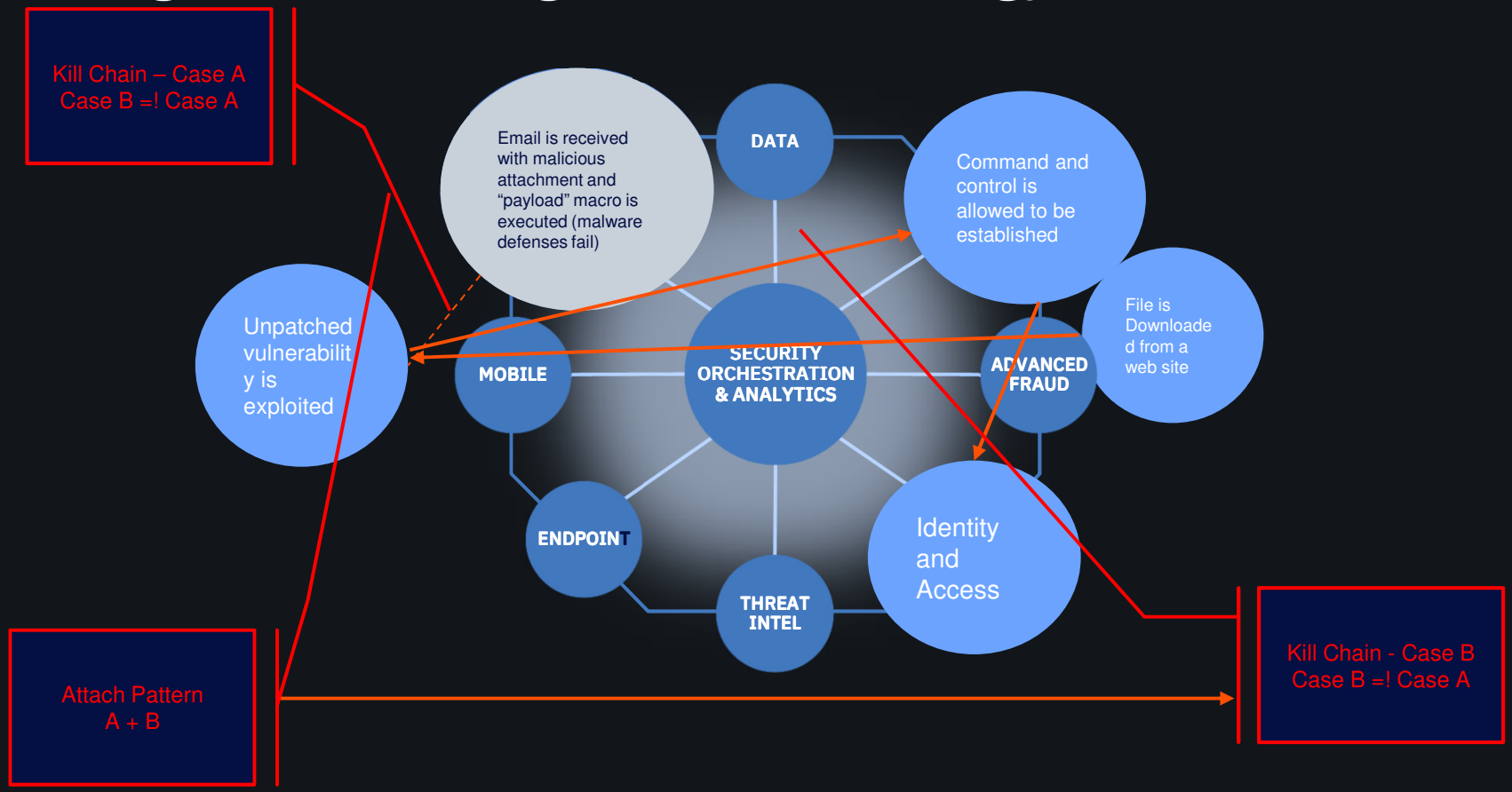


Analysis of an Incident

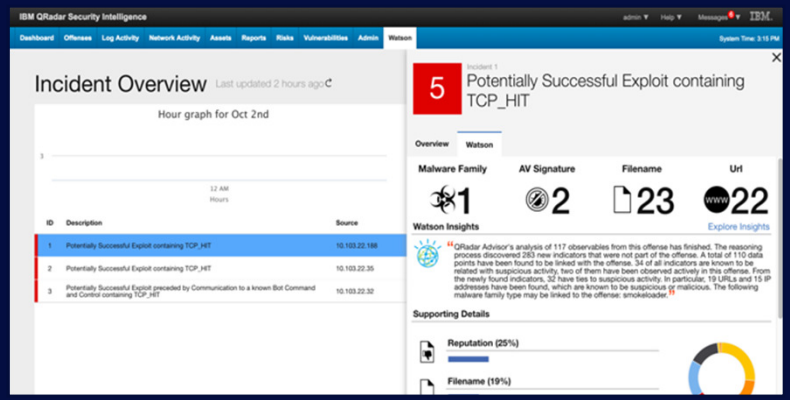
Activities performed during Business Email Compromise – Case A

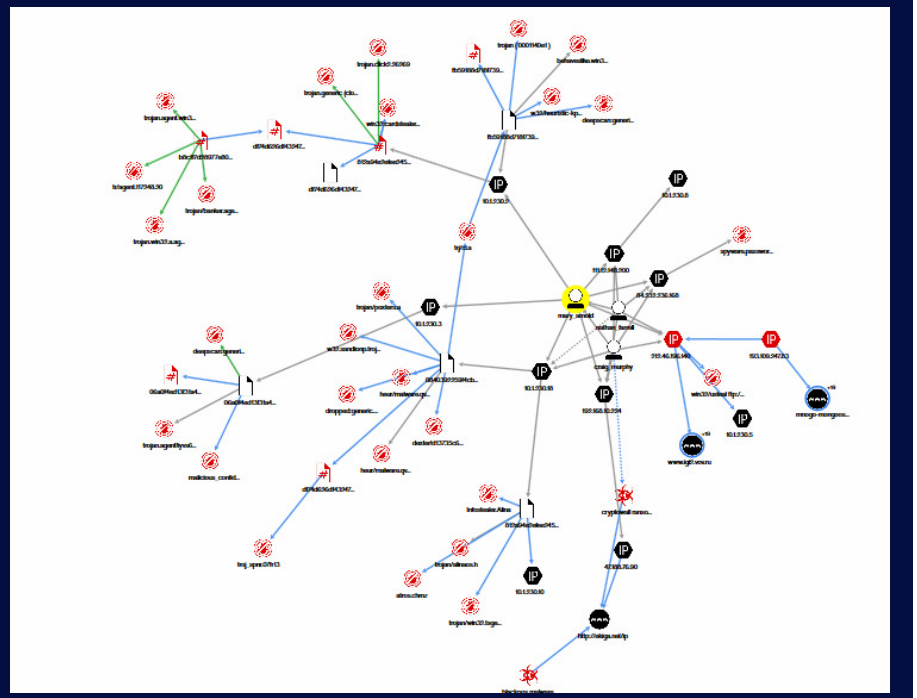


Watering hole .. A change in attach strategy . Case B



The future of Incident Analysis in Cyber Security is AI





Use AI to gain a head start

Automatically investigate incidents and anomalies to identify the most likely threats

- Quickly gather insights from millions of external sources
- Apply cognitive reasoning to build relationships

Incident Analysis

Intelligence gap

#1 most challenging area due to insufficient resources is threat research (65% selecting)

#3 highest cybersecurity challenge today is keeping current on new threats and vulnerabilities (40% selecting)

Speed gap

The top cybersecurity challenge today and tomorrow is **reducing average incident response and resolution time**

This is despite the fact that **80%** said their incident response speed is much faster than two years ago

Accuracy gap

#2 most challenging area today is optimizing accuracy alerts (too many false positives)

#3 most challenging area due to insufficient resources is threat identification, monitoring and escalating potential incidents (61% selecting)

Addressing gaps while managing cost and ROI pressures

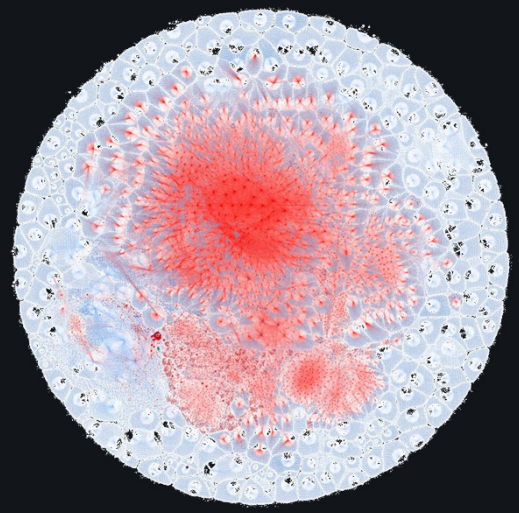
Watson for cybersecurity **unlocks a tremendous amount of security knowledge** enabling rapid and comprehensive investigation insights

Human Generated Security Knowledge and IBM Research

- Maintains the currency of
- Leverages the power of collaboration and crowdsourcing of threat intelligence and activity for more accurate insights
- Security Knowledge
- Learns new threat relationships and behaviors
- Performs cognitive exploration of suspicious activities and behaviors identifying root cause and additional indicators

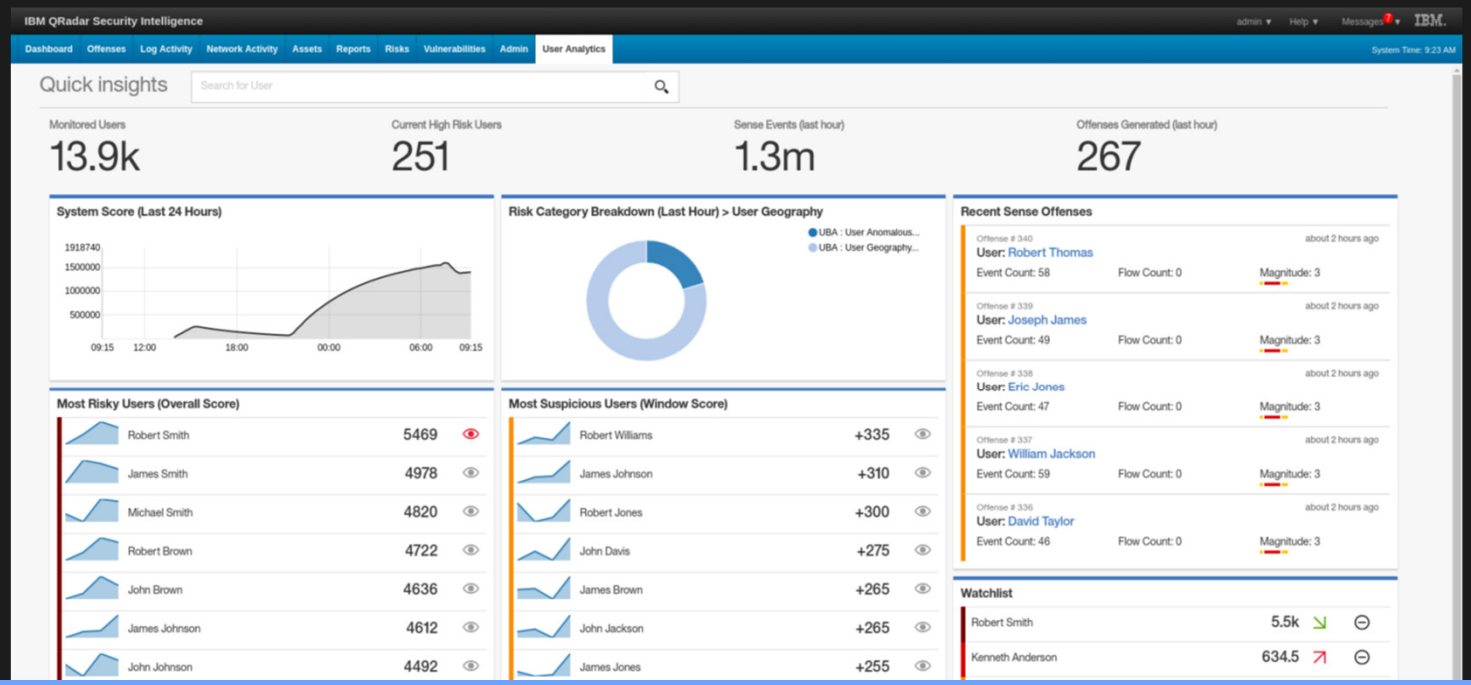
The Corpus of Watson for CyberSecurity in action

- Continually growing and adapting through the absorption of new security knowledge
- Performs cognitive exploration of suspicious activities and behaviors identifying root cause and additional indicators
- Creates and finds paths and linkages easily missed by humans
- Learns, adapts and doesn’t forget



Anomaly Detection

Detect and stop advanced threats



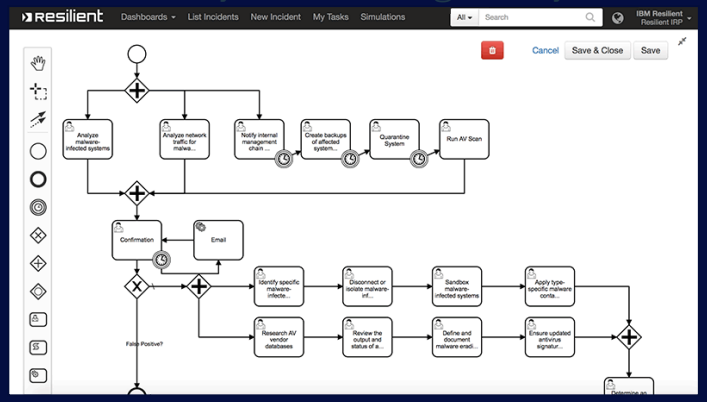
Advanced analytics for advanced threat detection and response across the enterprise

The User Behavior Analytics dashboard is an integrated part of the QRadar console

Intelligence Response

AI and Orchestration

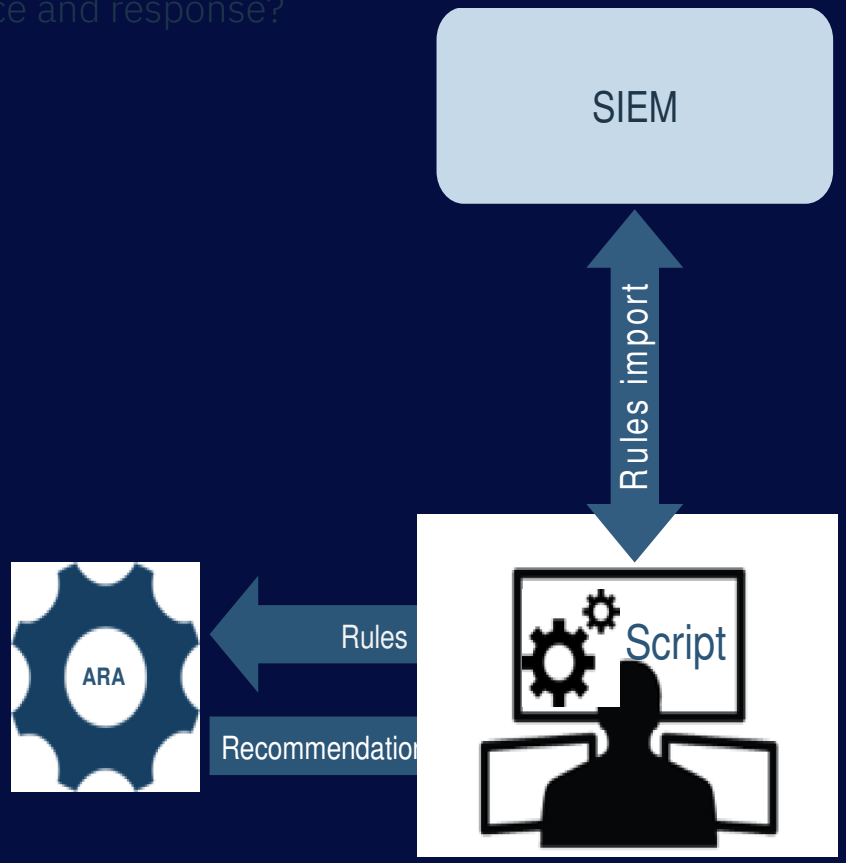
What if you could augment your teams' intelligence and response?



Respond quickly with confidence

Orchestrate a complete and dynamic response, enabling faster, more intelligent remediation

- Create dynamic playbooks built on NIST / CERT / SANS
- Deploy response procedures and expertise



Attacker’s Use of AI Today

Attacker's Use of AI Today

AI Powered Attacks

- **Generate:** DeepHack tool learned SQL injection [DEFCON'17]
- **Automate:** generate targeted phishing attacks on Twitter [Zerofox Blackhat'16]
- **Refine:** Neural network powered password crackers
- **Evade:** Generative adversarial networks learn novel steganographic channels


Attacking AI

- **Poison:** Microsoft Tay chatbot poisoning via Twitter (and Watson "poisoning" from Urban Dictionary) **[Po]**
- **Evade:** Real-world attacks on computer vision for facial recognition biometrics [CCS'16] and autonomous vehicles [OpenAI] **[Ev]**
- **Harden:** Genetic algorithms and reinforcement learning (OpenAI Gym) to evade malware detectors [Blackhat/DEFCON'17] **[Ev]**


Theft of AI

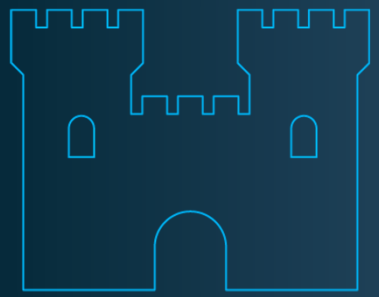
- **Theft:** Stealing machine learning models via public APIs [USENIX'16] **[DE]**
- **Transferability:** Practical black-box attacks learn surrogate models for transfer attacks [ASIACCS'17] **[ME, Ev]**
- **Privacy:** Model inversion attacks steal training data [CCS'15] **[DE]**


ME: Model Extraction
DE: Data Extraction
Ev: Model Evasion
Po: Model Poisoning

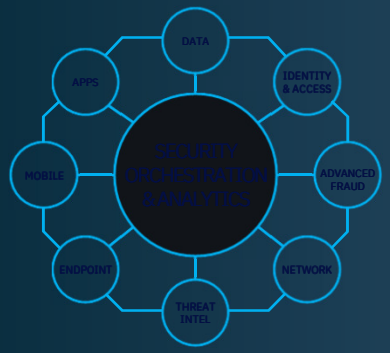
Reduce Complexity

Our continued journey

Before 2011
Security for an IT project



2011-2018
Security connected across the enterprise



2019+
Security at the Speed of Cloud

IBM Security

THANK YOU

FOLLOW US ON:

- ibm.com/security
- securityintelligence.com
- ibm.com/security/community
- xforce.ibmcloud.com
- @ibmsecurity
- youtube/user/ibmsecuritysolutions

IBM