



Good vs. Evil: AI Attack Mitigation Strategies

Steve McGeown, QAC Security Practice Leader

August 6, 2024



About the Session

Good vs. Evil: AI Attack Mitigation Strategies

Bad actors are more powerful now than ever before. Armed with AI-based technologies, unlimited compute power, encryption, password guessing tools, and substantial human and financial resources, organizations worldwide are struggling to combat their sophisticated and relentless pursuits. How do you fight an “invisible” enemy that hides behind untraceable IP addresses, based in countries on the other side of the world, and are seemingly unreachable due to the lack of internationally adopted legal protections? This session dives into five proven risk mitigation strategies to fight back against the criminals that are infiltrating our companies, stealing personal information, holding systems hostage, siphoning money, and more – all for personal gain, while putting property and lives at risk.

Learning objectives:

- 1) Identify your system’s weakest links making you susceptible to malicious infiltration
- 2) Understand the real, credible risks your organization faces from bad actors
- 3) Discover the best risk mitigation strategies to combat bad actors in an AI-driven landscape

About Me



PROFESSIONAL EXPERIENCE

- Executive Leadership
- Infrastructure Security
- Application Security
- Product Management
- Program Management
- Software Quality Assurance
- QA Governance
- Software Implementation
- System Integration
- Software Engineering and Testing
- Automation, Performance, Security, Usability, Accessibility, and Mobile Quality

Steve McGeown, Security Practice Leader

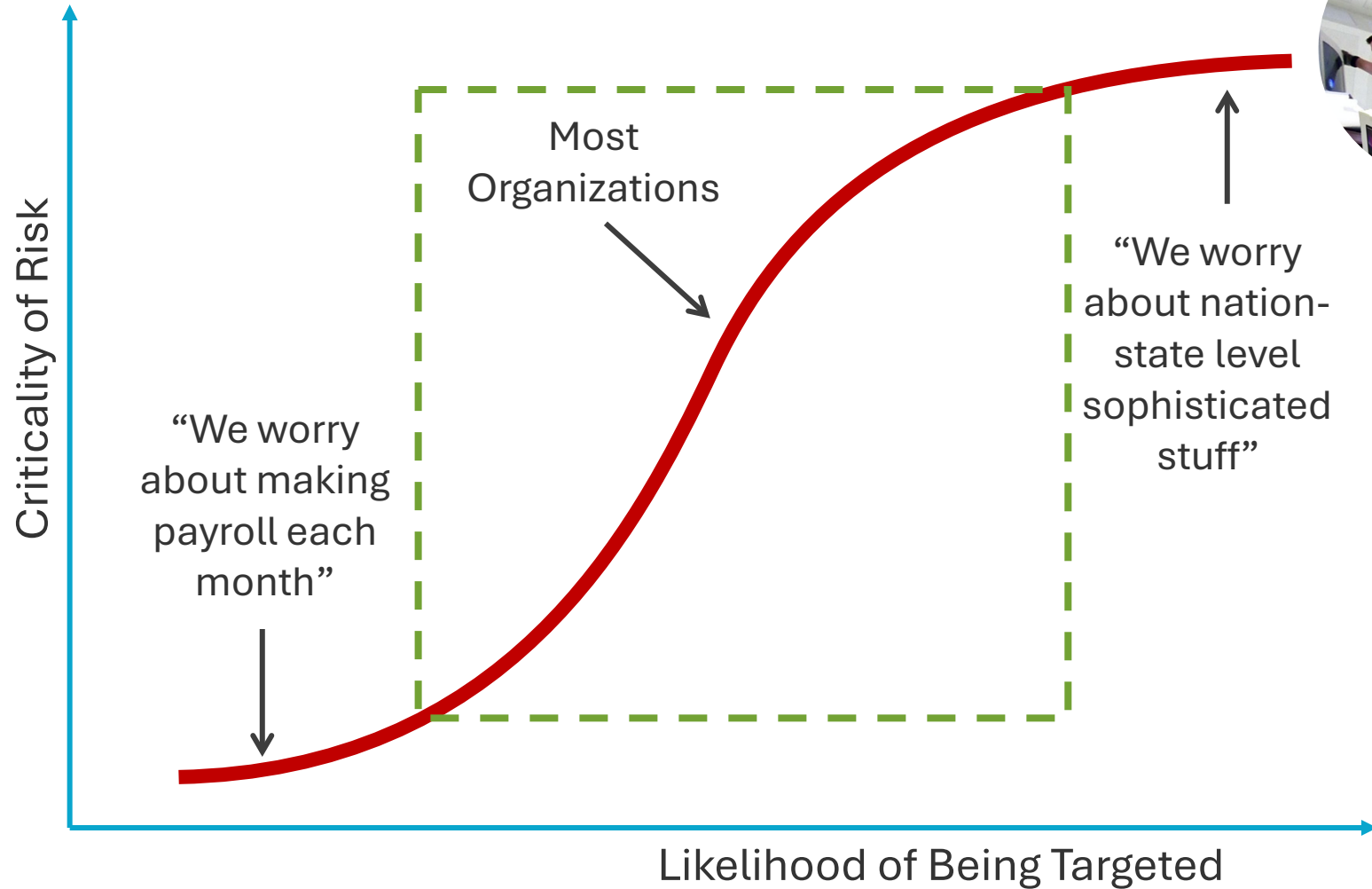
Steve leads all “Red Team” aspects of Security Professional Services in QA Consultants, as well as virtual CSO placements in other firms and has led all aspects of a Managed Security Services in virtual red and/or blue team teams. This includes supervision of the strategies, test methodologies, test execution, static code analysis, dynamic tooling and reporting techniques. Steve has performed management and execution of vulnerability assessments / penetration tests at both the application and infrastructure layers of highly complex and security-sensitive environments such as:

- Military grade applications
- Energy utility companies
- Financial institutions
- Software development companies
- Government institutions
- Specialty applications (kiosks / physical po-ups)
- and many others

Steve often speaks at various conferences on a diverse range of IT security topics. He has led all the QAC security projects listed in the case studies and customer redacted reports, outlining our experience and capabilities.



Who is This Targeted At? Risk Profile



A Seminal Moment in Cybersecurity History...

- Not the FIRST... but a 2016 Colonel Kurtz Moment...
- Years of Car Telemetry Data Encrypted / Denied
 - Estimated worth >\$2M, but to whom?
 - \$500 Ransom Paid



Fast Forward... We Learned So Much...

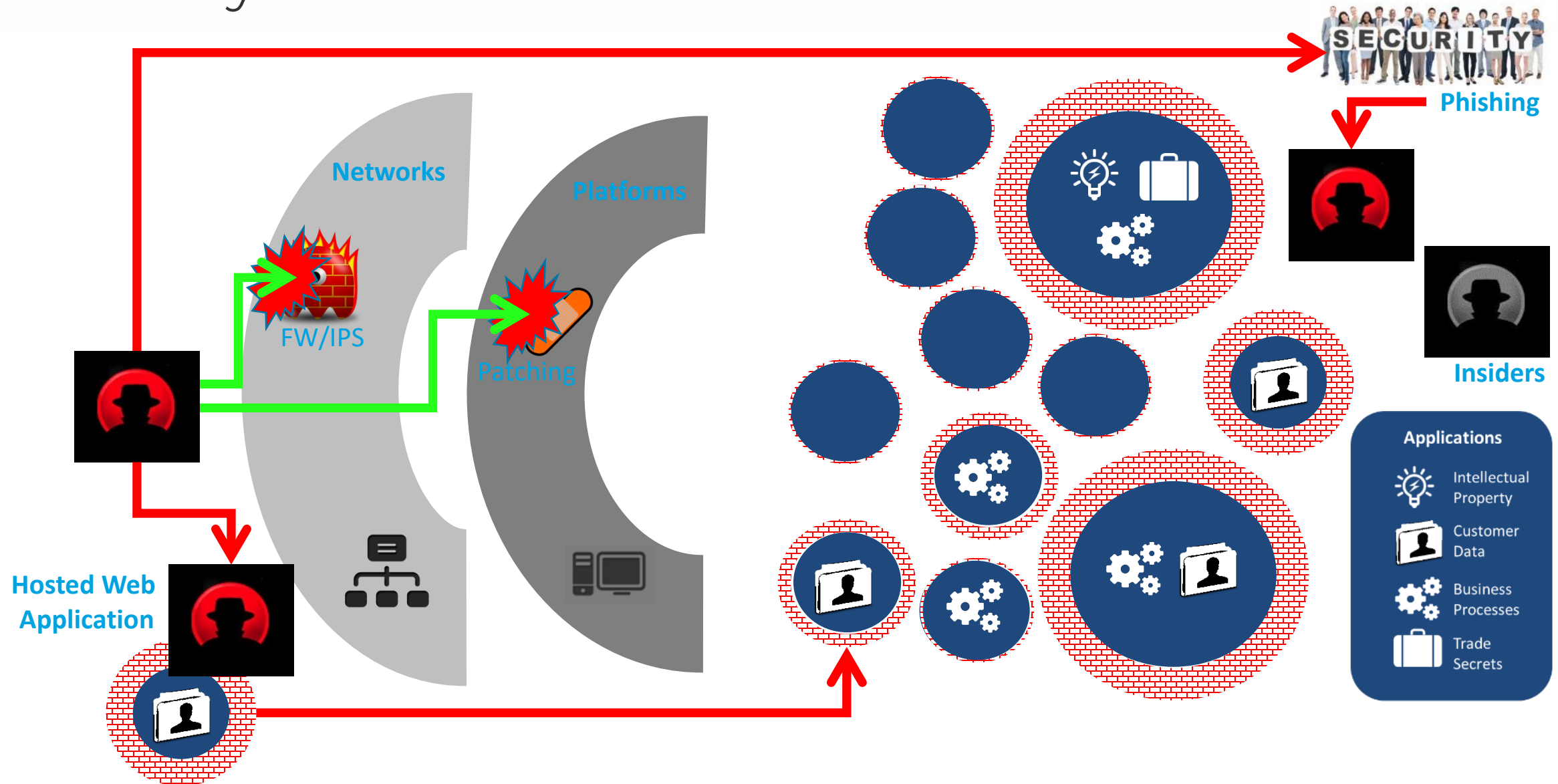
- It all feels so familiar somehow (June 2024)
 - 15,000 auto dealerships nationwide
 - \$25M (Rumor) Ransom Paid



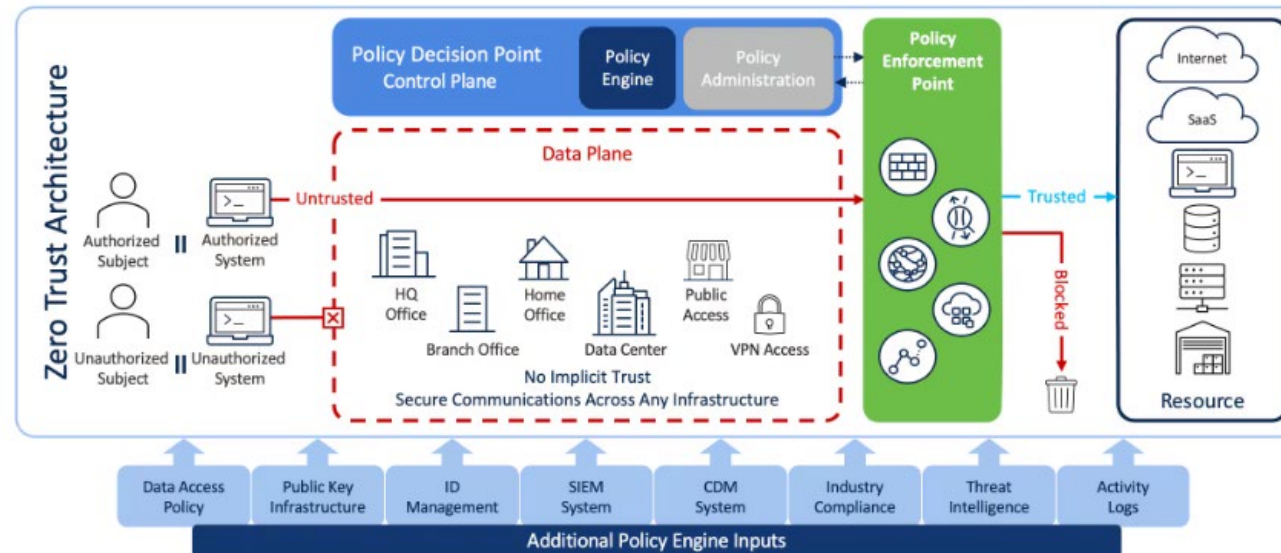
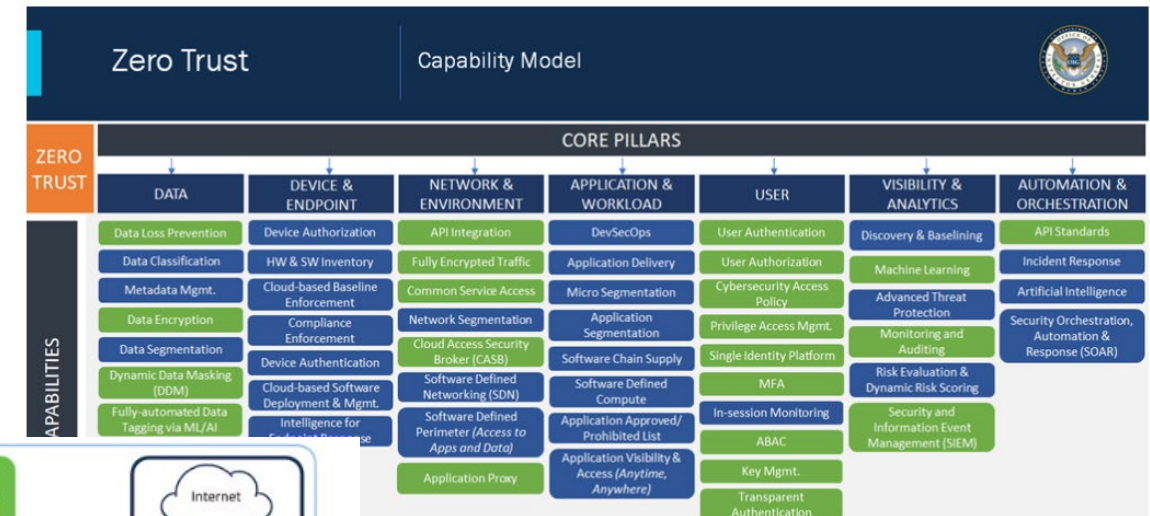
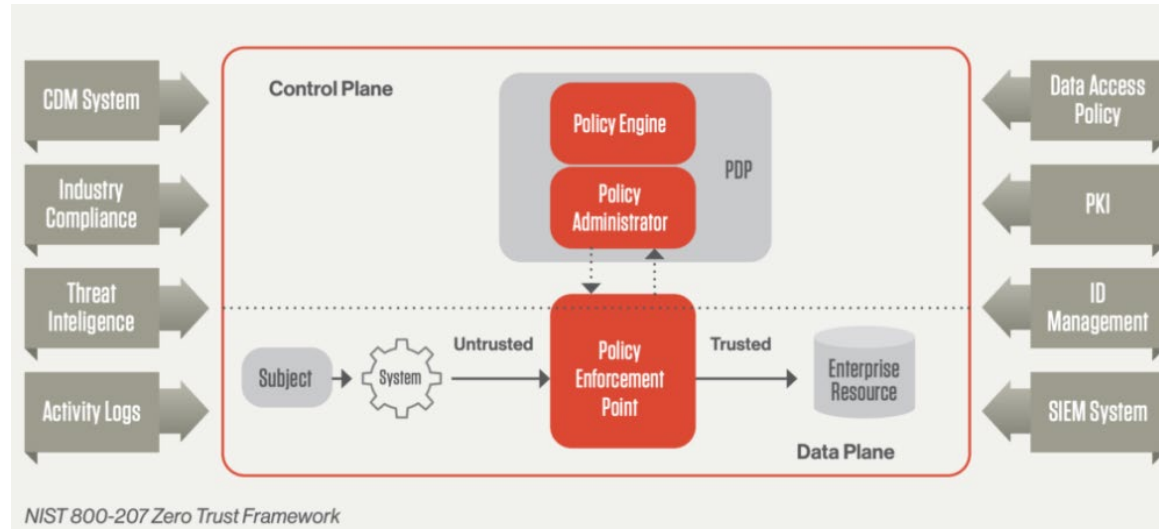


“YOU CLICKED??!!”

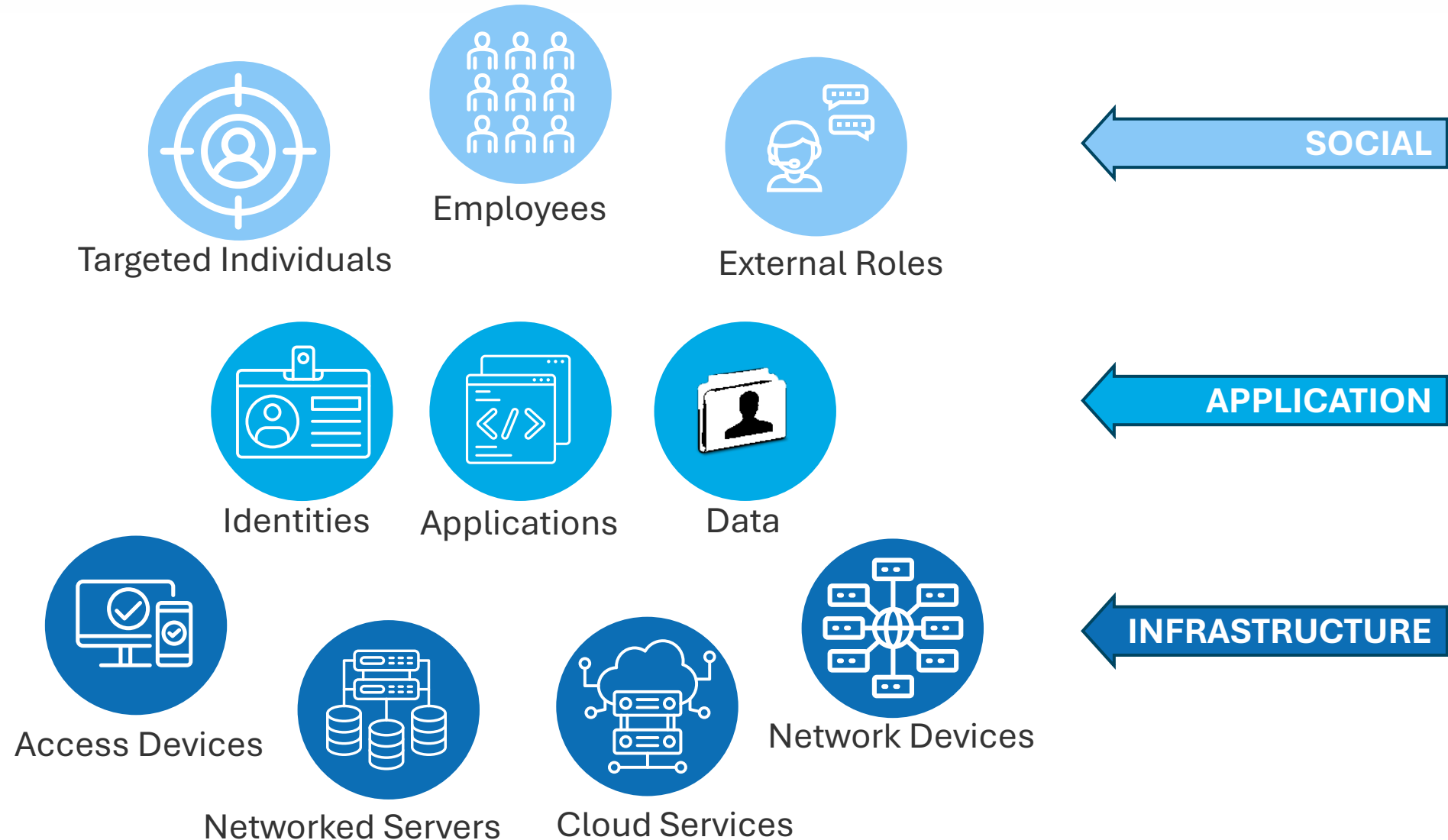
Really?? “Perimeter Mindset” Revisited...



So, What About This “Zero Trust” Thing?

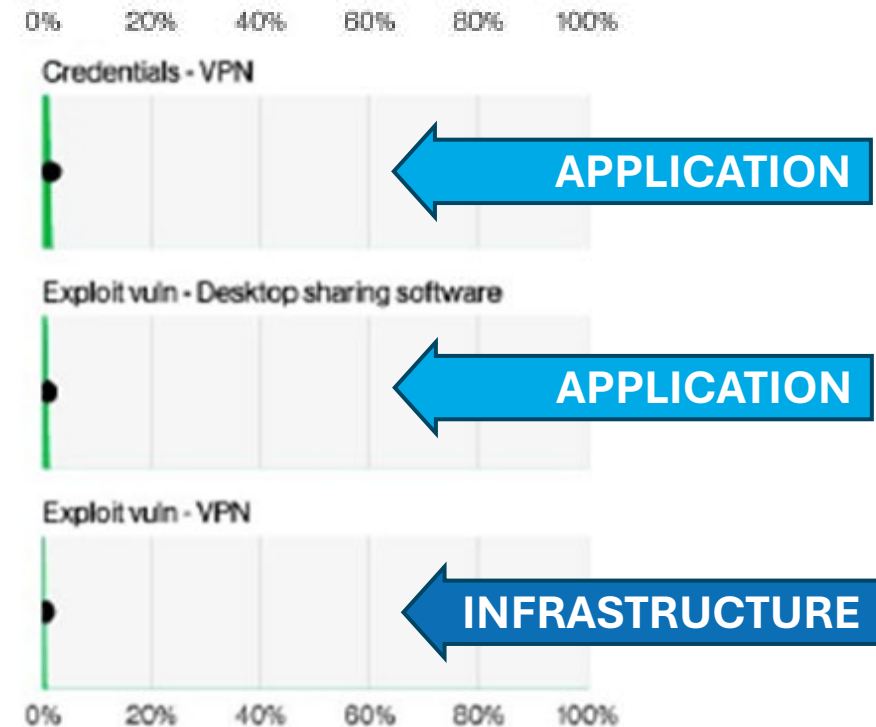
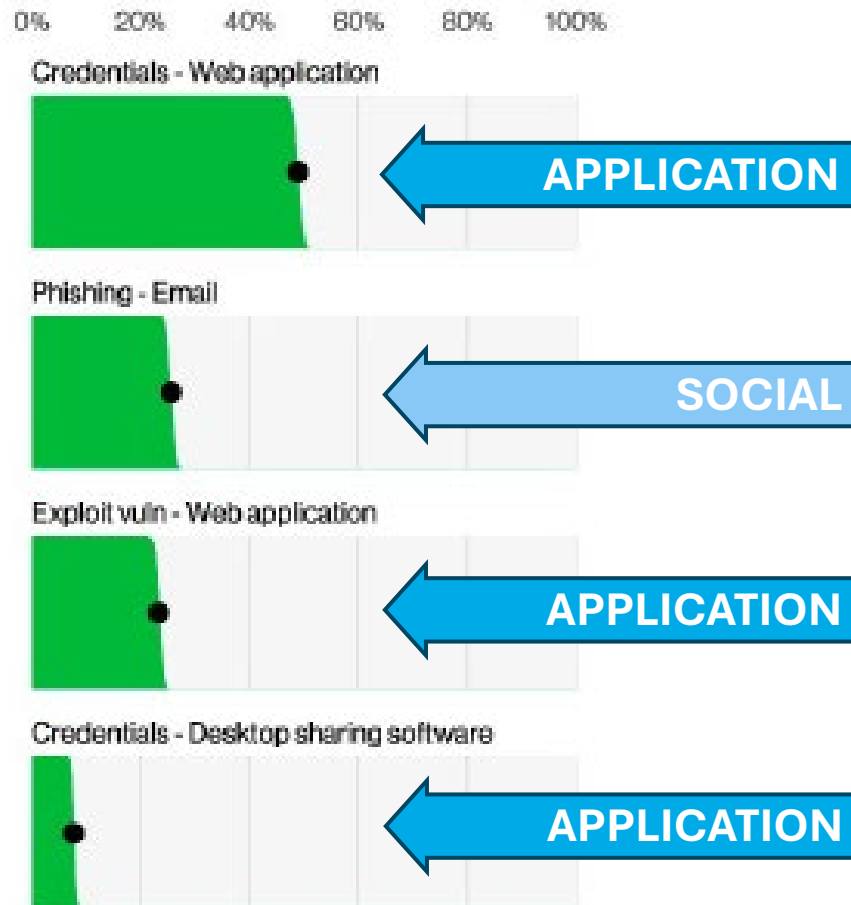


Simpler Way to Look at “Zero Trust”...



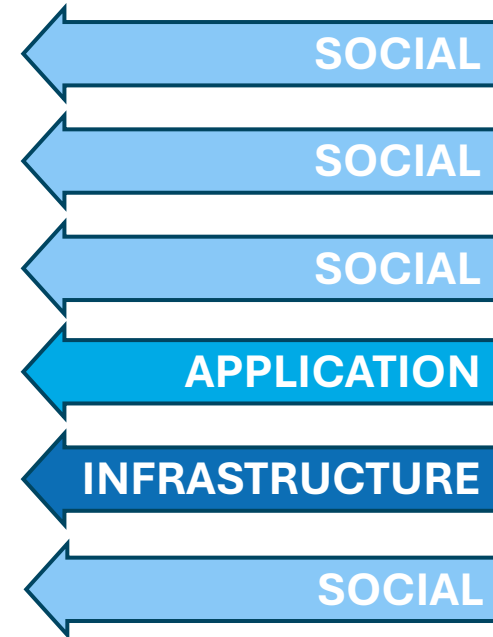
2024 Verizon Data Breach Investigation Report

Where Current Fun Is



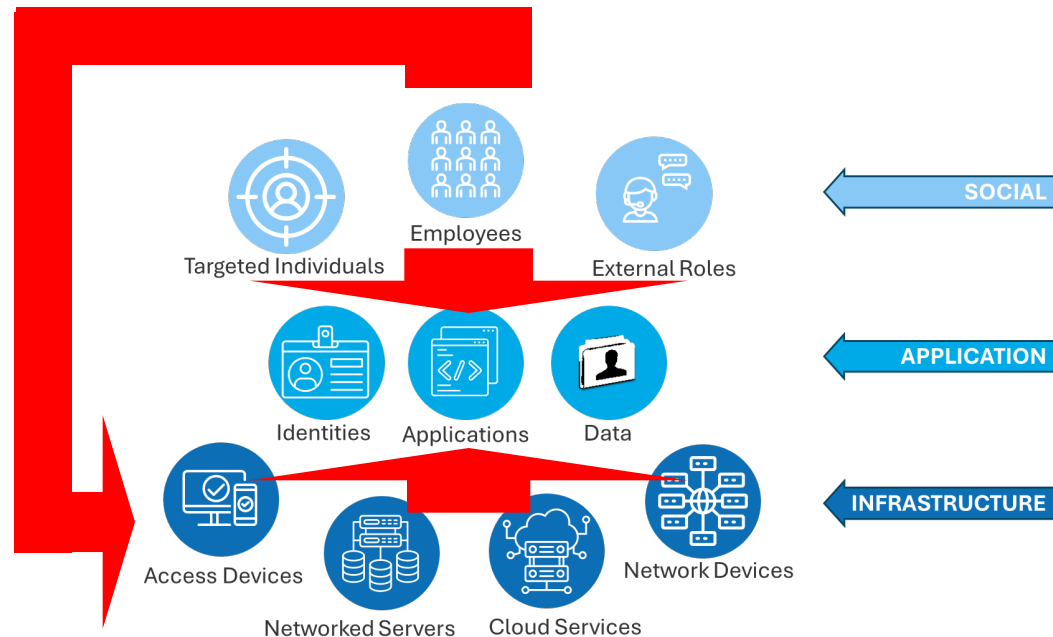
Where We Have Seen AI in Breaches

- Deepfakes
- Automated Phishing Email Targeting
- Ransomware Post Breach Exploration
- Credential Stuffing
- Exploiting Vulnerabilities
- Nation State Intense Data Gathering



Incoming AI Horror Show... Relevancy?

- Infrastructure Vector: Makes baseline more expensive
- Social Vector: Gets cheaper to target / be good at it
- Application Vector: Where the valuable stuff actually is....



Things AI Tends to Handle Well

- Efficiency & Automation (Repetitive Tasks & Speed)
- Accuracy & Precision (Error Reduction, Consistency)
- Large Dataset Handling (Predictive)
- Enhanced Tailoring / Personalization (Match to Target)

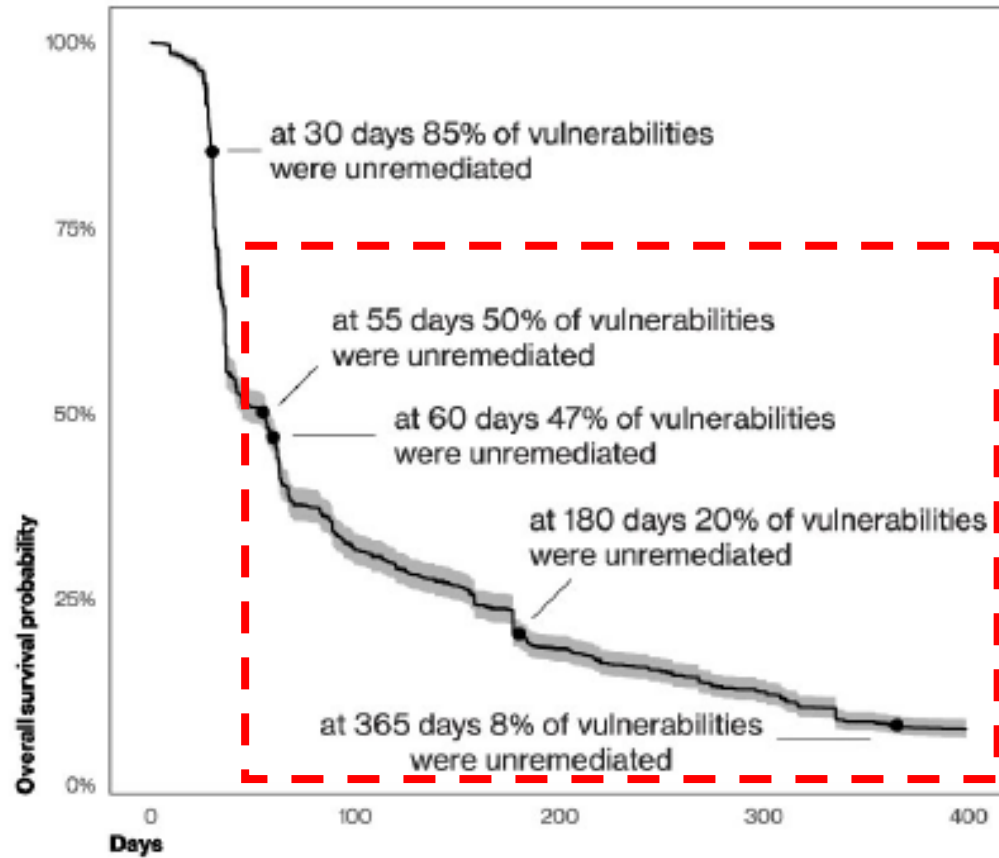


Things AI Tends to Not Handle Well

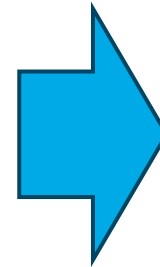
- Understanding Context
- Common Sense Reasoning
- Ambiguity
- Limited Learning From Few Examples



Bad News! Infrastructure Is Perfect For Bad Guys With AI To Attack / Recon ❌❌



Survival KNE (Known Exploited Vulnerability)



Good News! AI Infrastructure Defense Also Predictable, Cloud is Good at it ✓✓

- Enterprise
 - XDR platforms (i.e. devices)
 - AI threat detection & intel
 - MDR vendors
 - AI security can be expensive
- Outsourced Cloud
 - Small army of security assets employed
 - Pre-packaged services
 - Cloud can be expensive



Enterprise Cloud & Infrastructure

Bad News! Expert Social Engineering More Democratized/Intense With AI ❌❌❌

- Especially Highly Targeted “Spear Phishing”
- We’re Well Beyond “Click for a Free Starbucks Coffee”
- Call to Action -> Play on Emotion Remains the Same
 - Fear
 - Greed
 - Curiosity
 - Helpfulness
 - Urgency

Good News! Know How to Do This, Sort of...

- Social Policies Control Users
 - MFA enforcement -> good
 - Some can be overwhelming (i.e. Office 365 compliance examples)
 - At what cost to the business?
- Social Engineering Testing
 - Relevant to each person's task?
 - Updated with new content?
 - Can we keep up to new threats?
 - Spear phishing key personnel?
- Security Awareness Training
 - Timing cycle?
 - Just a checkmark or true value sought out?
 - Fun? Interesting?
 - Relevant to the business?

However, More... ❌❌



What Executives Want
to Believe Happens in
Security Awareness
Training....

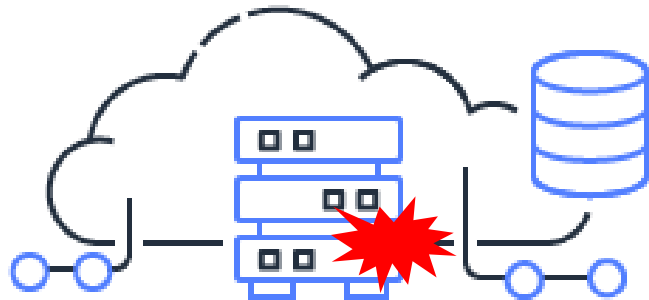


What Employees /
CISOs Know Happens
in Security Awareness
Training....

Every Application Is Its Own Snowflake...



Your Cloud & Infrastructure



Everyone Else's Cloud & Infrastructure



Your Application



Your Other Application

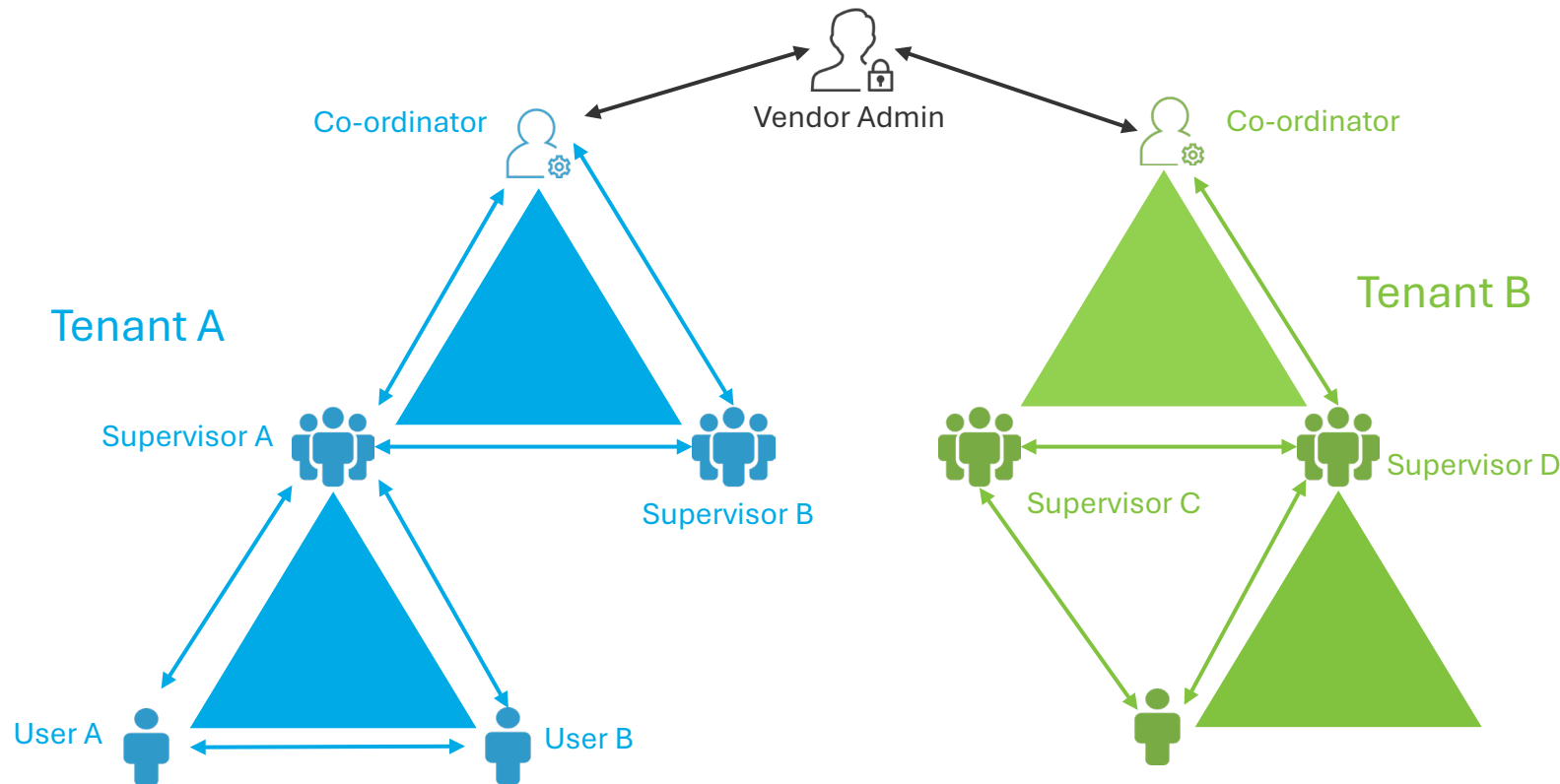


Bad News! Old App Code Must Live! ❌

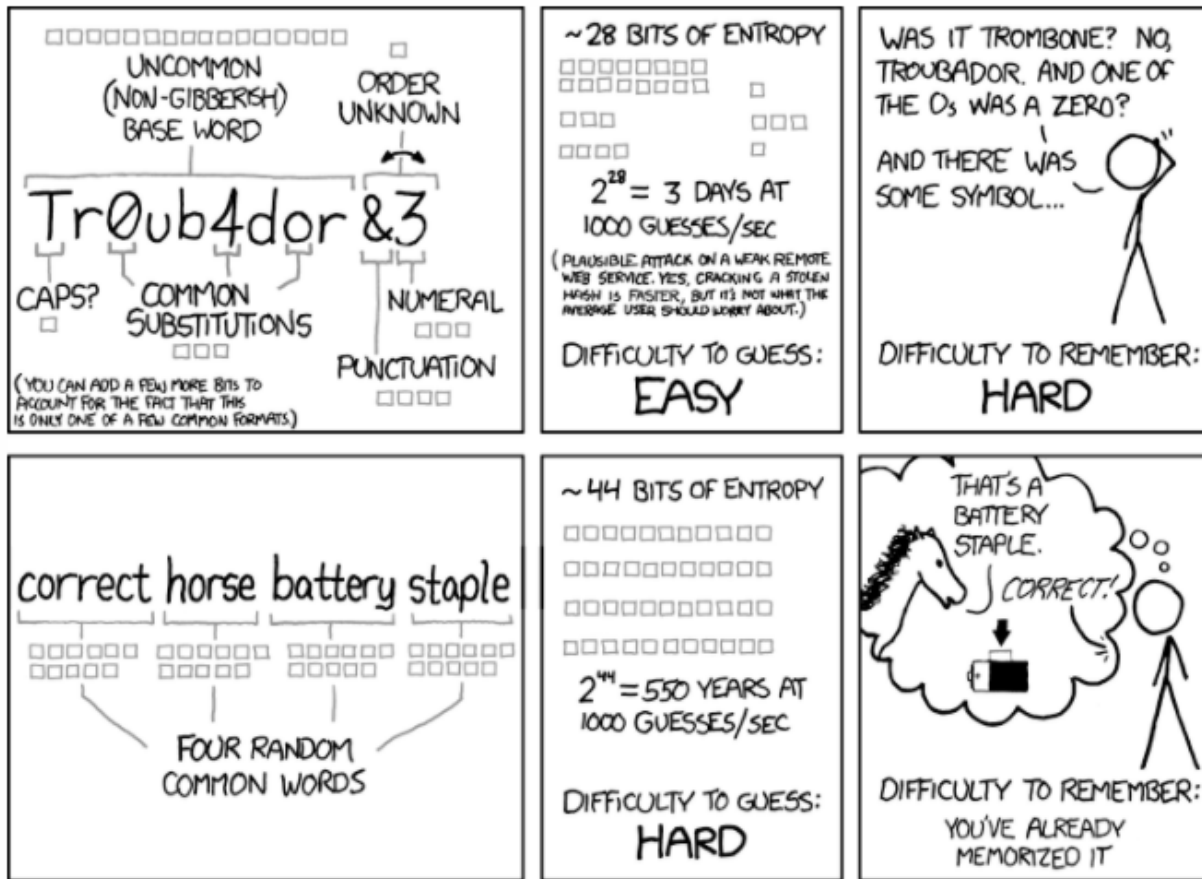
- Old Frameworks / Mixed Frameworks
- Sometimes Band-Aids Make More Sense Than Fixes
- Old Exploits and “Cruft” Are Excellent Targets
- Complex Patchwork of Open Source, etc.

Neutral.. Authorization of Roles is Hard

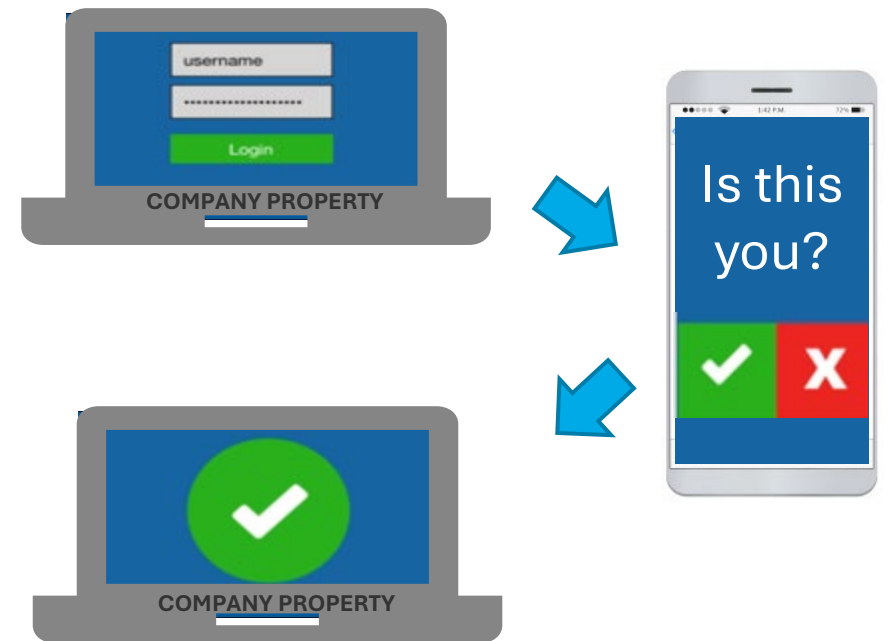
- Context is Hard To Automate With Little Training
- AI of Credentialed Apps Hides Training Data To Learn.



Good News! Application Identity / Authentication Can Be Very AI Resistant

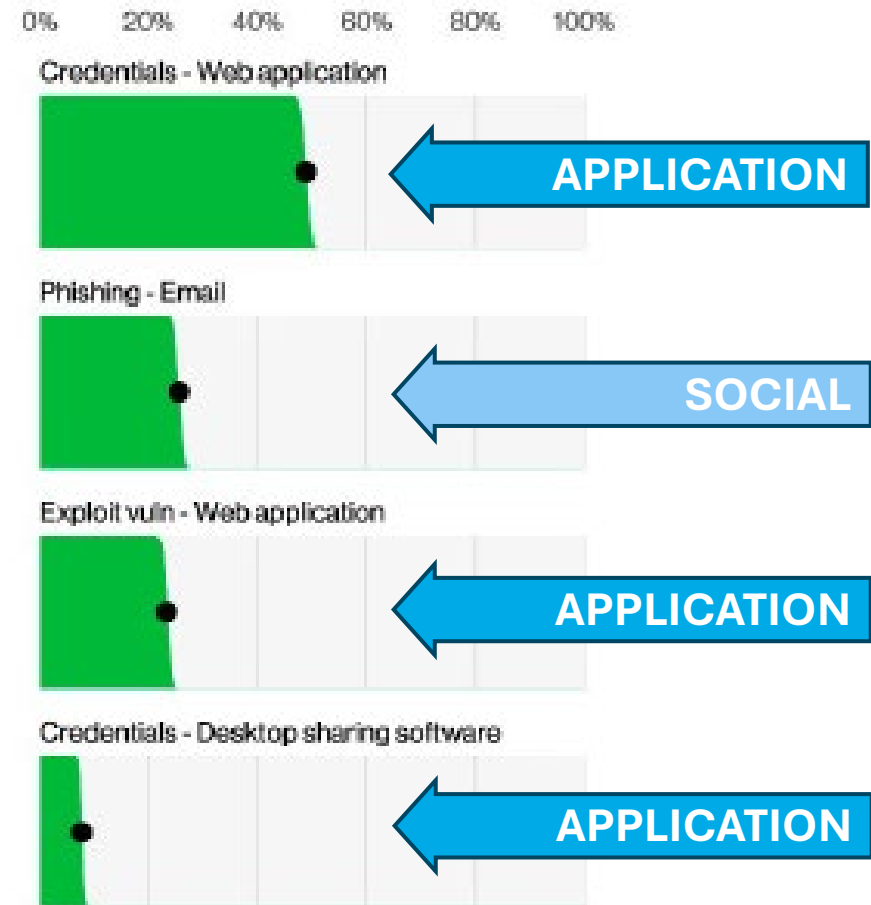


THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

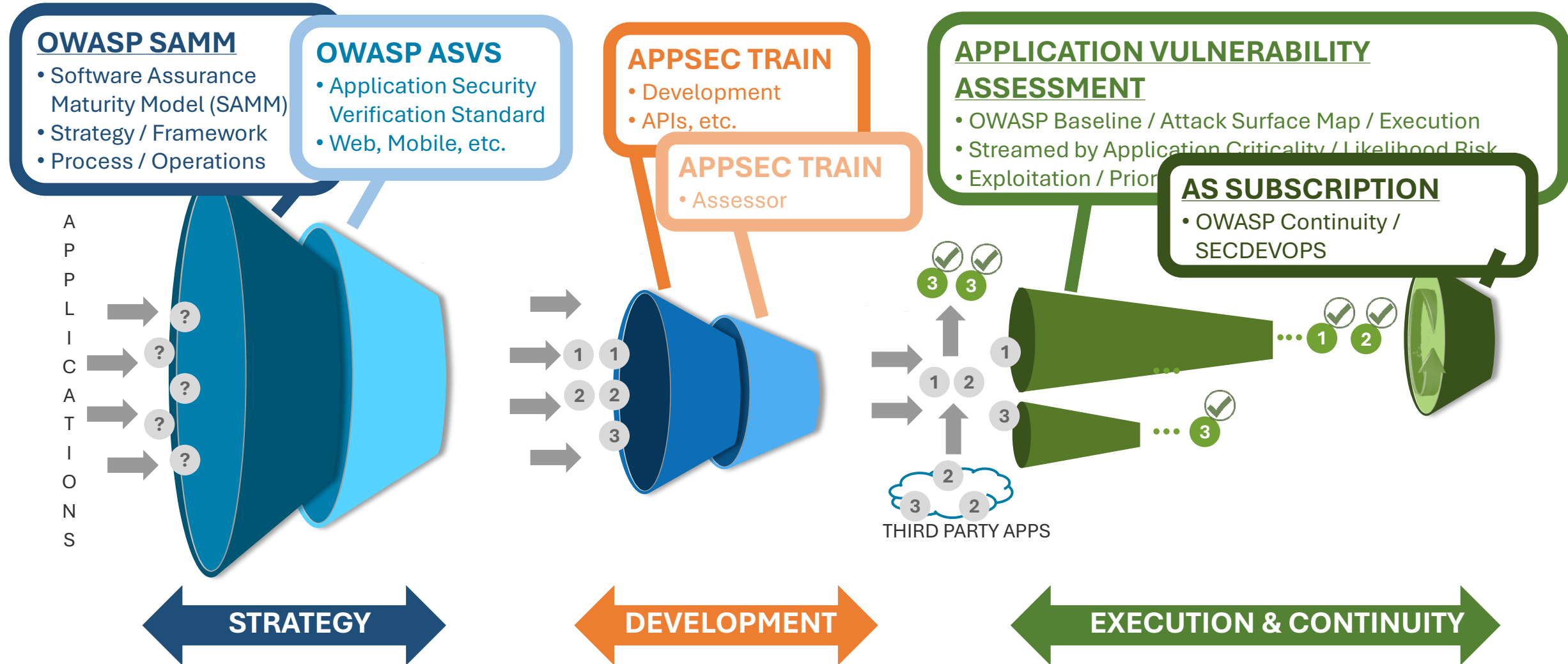


Good News! AI is Very Limited In Application Breaches... So Far... ✓✓

- Brute force is not really AI
- Some credential stuffing maybe, eventually
- Credentialed apps make understanding roles hard



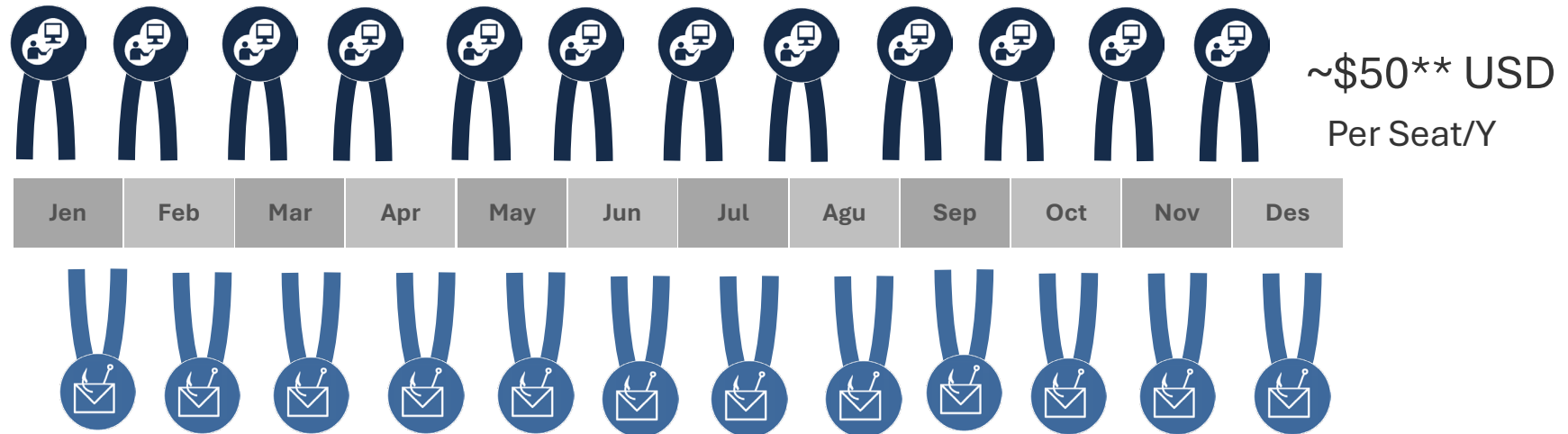
So, What to Do? Plan Application Security In



So, What to Do? Plan People Security In

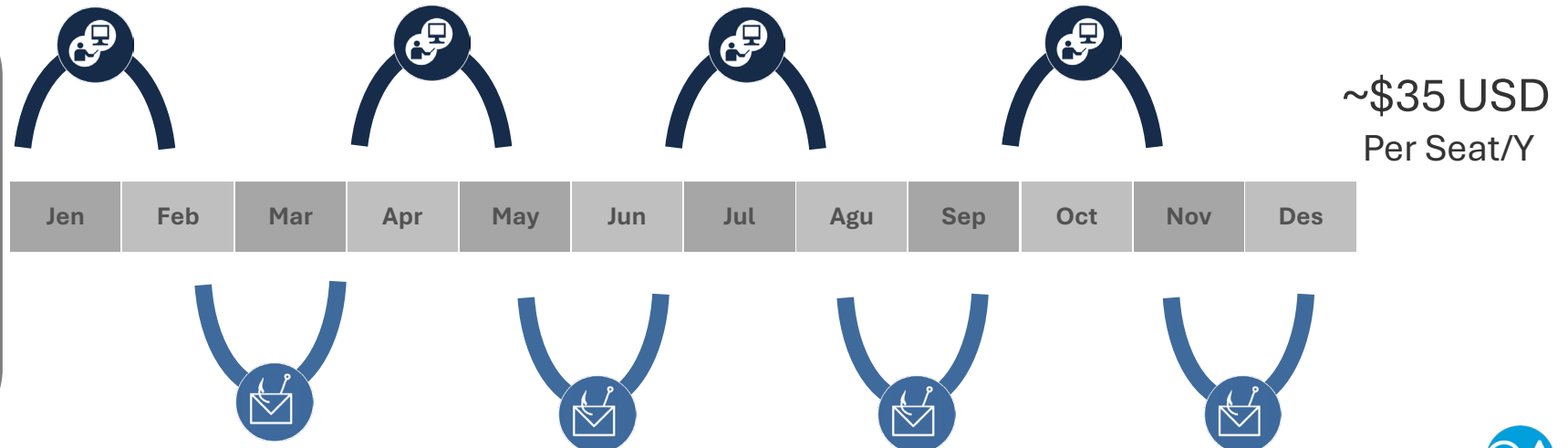
Bank Model

- **Cycling Content**
 - High Risk Organization
 - Technology Focus
 - Company size / flux
- **Content Message**
 - Formal Education Normal
 - Heterogeneous Skills



Tech Company Model





- **Cycling Content**
 - Lower Risk Organization
 - Lower Staff Turnover
- **Content Message**
 - Higher Tech Skills
 - Homogeneous Skills
 - Abstraction Capable



So, What to Do? Application Security Execute

Step 1: Select the Level

The desirable level of application security.








ASVS LEVELS	MEANING	COLOR	NOTES
Level 3	Advanced		The highest level. t
Level 2	Standard		Standard. "Adequa
Level 1	Opportunistic		You still have no se
Level 0	Cursory		"Flexible", "custom

1

1. V1. Architecture, design and threat modeling
2. V2. Authentication
3. V3. Session management
4. V4. Access control
5. V5. Malicious input handling
6. V7. Cryptography at rest
7. V8. Error handling and l
8. V9. Data protection
9. V10. Communications
10. V11. HTTP security confi
11. V13. Malicious controls
12. V15. Business logic
13. V16. File and resources
14. V17. Mobile
15. V18. Web services
16. V19. Configuration

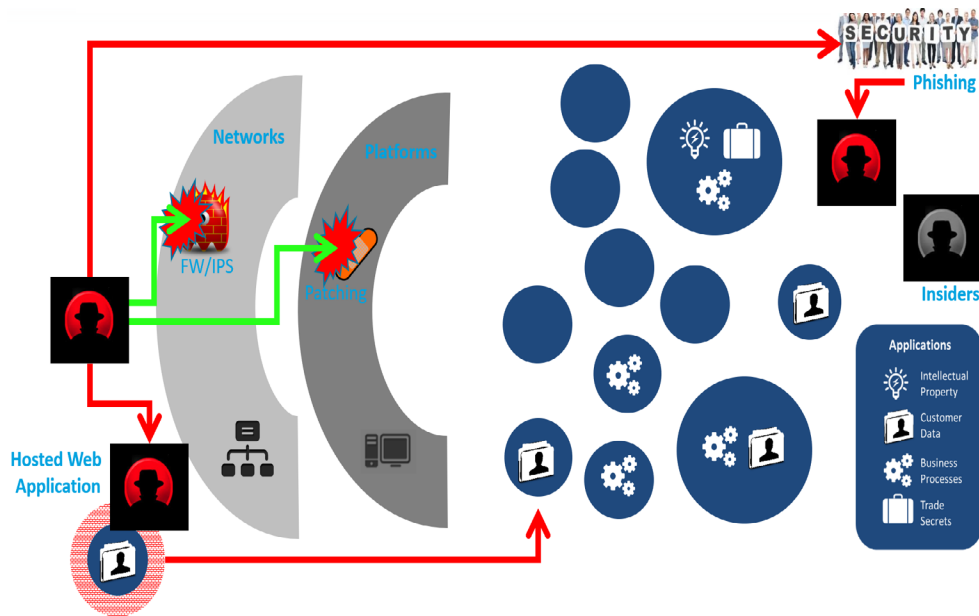
2

3

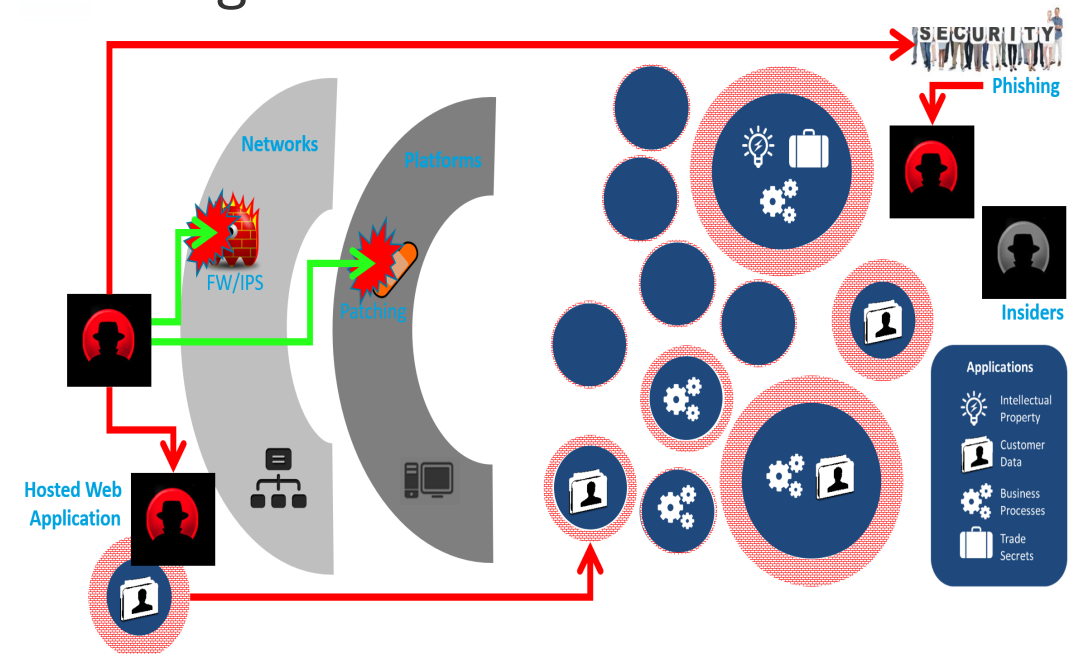
#	Description	1	2	3	Since
1.1	Verify that all application components are identified and are known to be needed.				1.0
1.2	Verify that all components, such as libraries, modules, and external systems, that are not part of the application but that the application relies on to operate are identified.				1.0
1.3	Verify that a high-level architecture for the application has been defined.				1.0

So, What to Do? Get Here, Before AI Turns Up

So Sure, Make This “Hard”.... Layers Matter



But, Get Here....
Imagine a Perimeter-less World



Summary: AI Cyber Security Landscape

- Infrastructure Layer Not Huge AI Technical Problem, But Given Where Breaches Occur, Budget Priority Danger?
- Social Layer Not Always Working Now, Only Going to Get Worse With Advanced Generative AI Techniques
- Application Layer Can & Should Converge to “Zero Trust”
- Application Layer Doesn’t Expose Context Easily to Generative AI Techniques
- Pretend Perimeter / Social Engineering Bludgeoning Didn’t Exist, Re-Think This Entirely Around Applications



Consultants
an ALTEN company

Thank You! Questions?

Steve McGeown
smcgeown@qac.com

