Regional Workshop for French-Speaking Africa

Addressing the Nexus of Artificial Intelligence and Insider Threat Mitigation

Typhon Intrusion – Atomic Regulatory Agency (ARA)

Use of Artificial Intelligence to identify Insider Threats

INS International Nuclear Security
Reducing Risk of Nuclear Terrorism
NNSA National Nuclear Security Administration

The Atomic Regulatory Authority

Overview: The Atomic Regulatory Agency (ARA)

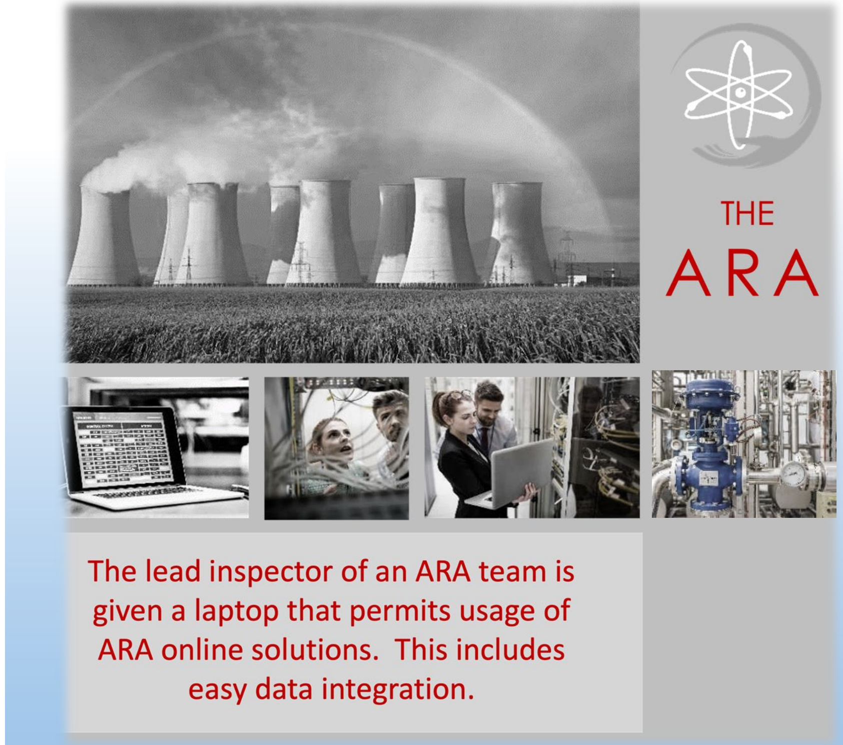
The Atomic Regulatory Agency (ARA) is the Competent Authority for your country.

- Licenses and regulates the Nation’s use of radioactive and nuclear materials.
- Creates and enforces safety, security, and reliability regulations
- Inspectors collect data during an audit and archive inspection activities.

As part of this effort, the regulator has developed online tools to enable ARA inspectors to relay, process, and store data acquired during an audit or assessment.

The Security Challenge:

- Inspectors have privileged access to sensitive nuclear facility data
- Online tools create new digital vulnerabilities and data exposure risks
- Traditional security measures may not detect subtle behavioral changes
- Need for proactive identification of potential insider threats
- **Today's Exercise:** Working in groups, you will explore how artificial intelligence can strengthen ARA's insider threat detection capabilities while maintaining operational efficiency.



The lead inspector of an ARA team is given a laptop that permits usage of ARA online solutions. This includes easy data integration.

Group Exercise: AI for Insider Threat Detection

Overview: The Atomic Regulatory Agency (ARA)

The Mission: Your team has been tasked by ARA leadership to identify how artificial intelligence can strengthen our insider threat detection capabilities. You'll need to think like both security professionals and inspectors to find solutions that enhance protection without hindering operations.

Exercise Instructions:

- Form groups of 4-5 participants
- Work through each element of the scenario and complete the assignments.

  Your goals include:

  - *Identify AI capabilities that could detect insider threats*
  - *Consider various threat scenarios and detection methods*
  - *Develop a prioritized list of AI solutions*
  - *Prepare a 5-minute group presentation*

March 4th, 2024: The ARA is Attacked: Typhon Ransomware

Incident Details

Typhon : What is it?

Typhon is a Ransomware-as-a-Service (RaaS) platform that uses the double extortion model (Exfiltrates AND encrypts).

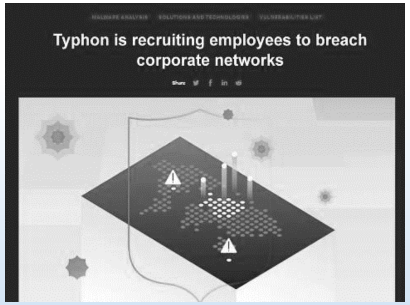
- **Exfiltrates**: Steals sensitive data and threatens to publish it publicly
- **Encrypts**: Locks files making them inaccessible until ransom is paid
- Attackers demand payment to: (1) decrypt files AND (2) not leak stolen data
- Even if ransom is paid, there's no guarantee data won't be sold or leaked later

Typhon is Everywhere in ARA!

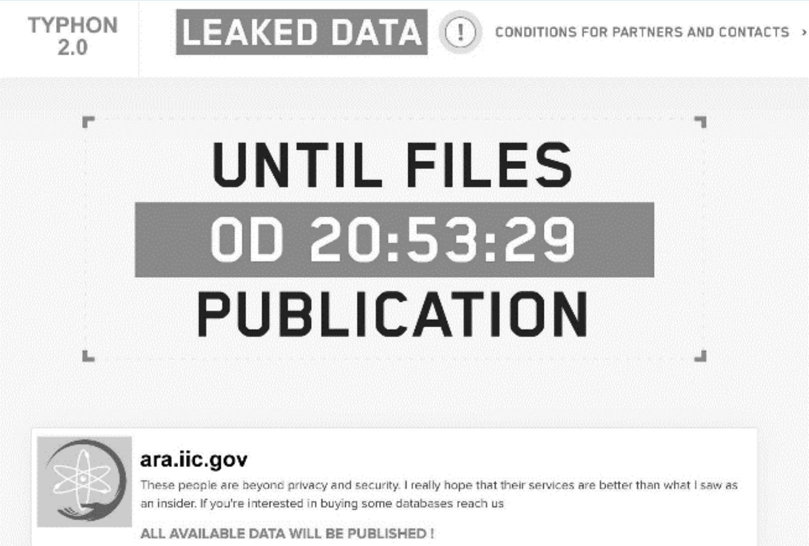
Typhon installed onto multiple ARA internal servers:

- All databases containing Sensitive Nuclear Information are encrypted
- All data backups are encrypted
- Inspection tools and applications are encrypted.

Was this an Inside Job?



Typhon is recruiting employees to breach corporate networks



UNTIL FILES
0D 20:53:29
PUBLICATION

ara.iic.gov

These people are beyond privacy and security. I really hope that their services are better than what I saw as an insider. If you're interested in buying some databases reach us

ALL AVAILABLE DATA WILL BE PUBLISHED !

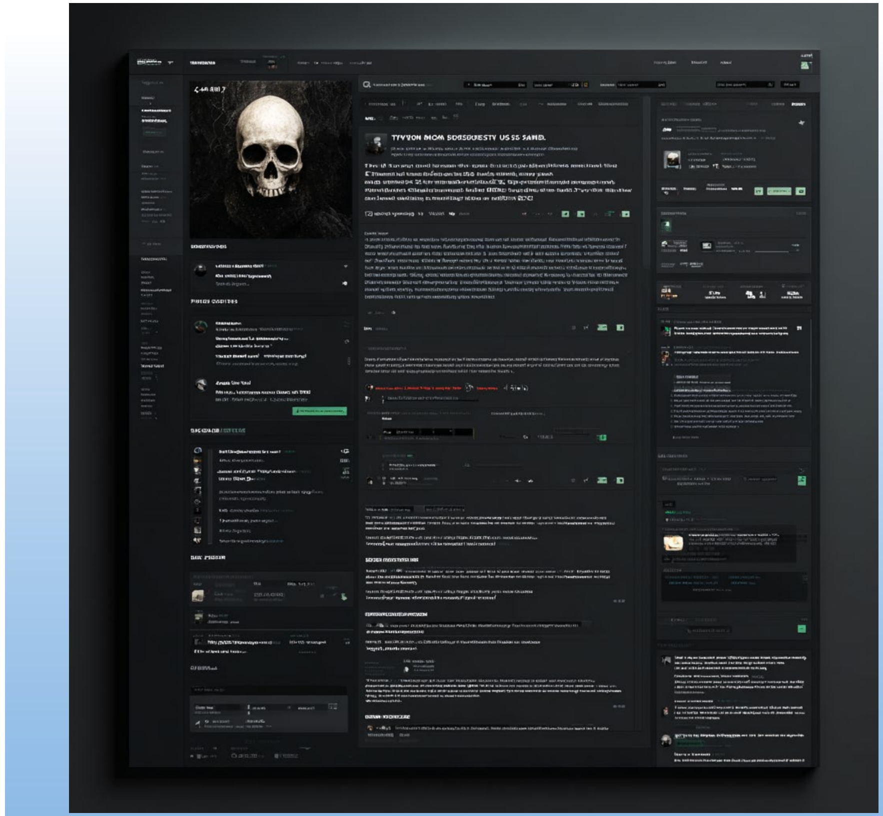
March 4th, 2024: The ARA is Attacked: Inside Job?

Incident Details

Inside Job?

Two months ago, Typhon openly recruited insiders on dark web forums:

- Offering up to $1M USD to install malware on corporate networks
- Targeting employees with privileged access or financial difficulties
- Posted "how-to" guides for bypassing security controls
- ARA was specifically named as a "high-value target" due to sensitive nuclear data



Discussion Prompts:

- **Behavioral Pattern Detection**
  How could AI analyze employee digital behavior patterns to identify someone who might be susceptible to or already compromised by Typhon's $1M recruitment offer?
- **Financial Stress Indicators**
  What AI/ML models could correlate HR data, access patterns, and external risk factors to flag employees experiencing financial pressure who may be targeted by Typhon recruiters?

Tracing the Attack: Digital Forensics

Artifacts Collected during Incident Analysis

Cyber Analysis

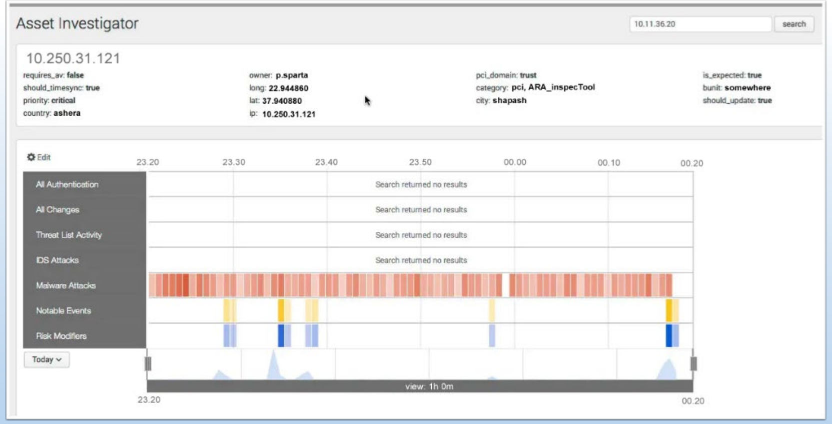
Cyber team identifies the initial infection point on ARA network:

- At 23:20 on March 3rd, laptop (10.250.31.121) initiates unauthorized network scan
- Laptop assigned to ARA inspector begins executing Typhon malware
- Timeline suggests infection occurred before network connection

Attack Propagation from 10.250.31.121

Once connected to ARA network, the infected laptop:

- Typhon immediately began scanning for vulnerable systems
- Malware spread laterally using inspector's privileged credentials
- Established command & control connection to external Typhon server
- Began data exfiltration before encryption phase



Asset Investigator

10.11.36.20 search

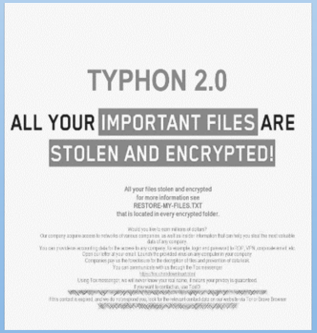
10.250.31.121

requires_av: false
should_timesync: true
priority: critical
country: ashera

owner: p.sparta
long: 22.944860
lat: 37.940880
ip: 10.250.31.121

pci_domain: trust
category: pci, ARA_inspecTool
city: shapash

is_expected: true
bunit: somewhere
should_update: true



TYPHON 2.0

ALL YOUR IMPORTANT FILES ARE STOLEN AND ENCRYPTED!

All your files stolen and encrypted
for more information see
RESTORE-MY-FILES.TXT
that is located in every encrypted folder.

Would you like to earn millions of dollars?
Our company acquire access to networks of various companies, as well as insider information that can help you steal the most valuable data of any company.
You can provide us accounting data for the access to any company, for example, login and password to RDP, VPN, corporate email, etc.
Open our letter at your email. Launch the provided virus on any computer in your company.
Companies pay us the foreclosure for the decryption of files and prevention of data leak.
You can communicate with us through the Tox messenger
https://tox.chat/download.html
Using Tox messenger, we will never know your real name, it means your privacy is guaranteed.
If you want to contact us, use ToxID
[redacted]
If this contact is expired, and we do not respond you, look for the relevant contact data on our website via Tor or Brave Browser
[redacted]

The Human Factor: Investigating Penelope Sparta

What do we know so far?

Preliminary Investigation: Inspector Penelope Sparta

- Laptop registered to Inspector Penelope Sparta
- Recently returned from routine NPP facility audit (Feb 28 - Mar 2)
- NPP IT security logs confirm laptop was scanned - no issues detected
- Sparta denies any knowledge of Typhon or suspicious activity

Discussion Prompts:

- **Behavioral Baseline Analysis**
  What AI models could establish normal behavior patterns for inspectors like Penelope and detect deviations that might indicate compromise or coercion?
- **Multi-Source Correlation**
  How could AI integrate data from badge access, network logs, email patterns, and financial systems to build a comprehensive risk profile for insider threat detection?
- **Predictive Risk Scoring**
  What ML approaches could assign dynamic risk scores to employees based on their access levels, recent life events, and proximity to sensitive systems?

Suspended pending investigation

Cyber Attacks Look Easy But They Are Not

Attack Planning and Attack Trees

The Attacker’s Dilemma

For every successful attack, multiple conditions must align:

- Attackers need detailed planning and reconnaissance
- Each step in the attack chain must succeed
- One detection or failure can expose the entire operation

The Typhon Attack Chain

1. Recruit insider with the right access ($1M offer)
2. Insider must bypass security controls undetected
3. Malware must evade endpoint protection
4. Lateral movement requires valid credentials
5. Data exfiltration needs unmonitored channels
6. Encryption must complete before detection

YOUR FILES ARE ENCRYPTED

Your photos, documents and other important files have been encrypted with unique key, generated for this computer.

The Defender’s Advantage: AI Opportunities

Defense Planning and Attack Trees

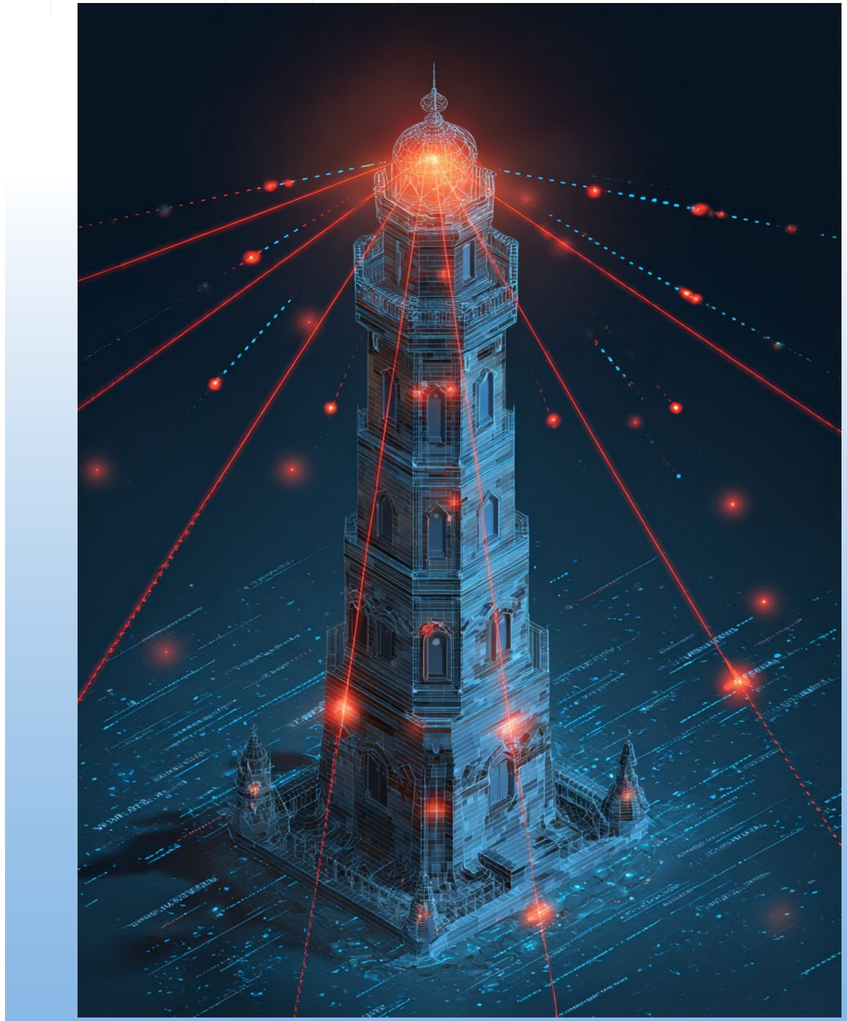
The Defender’s Advantage

- AI can monitor every link in the attack chain simultaneously
- Behavioral anomalies at ANY stage can trigger alerts
- Machine learning improves detection with each attempt
- Focus on what attackers MUST do, not what they might do

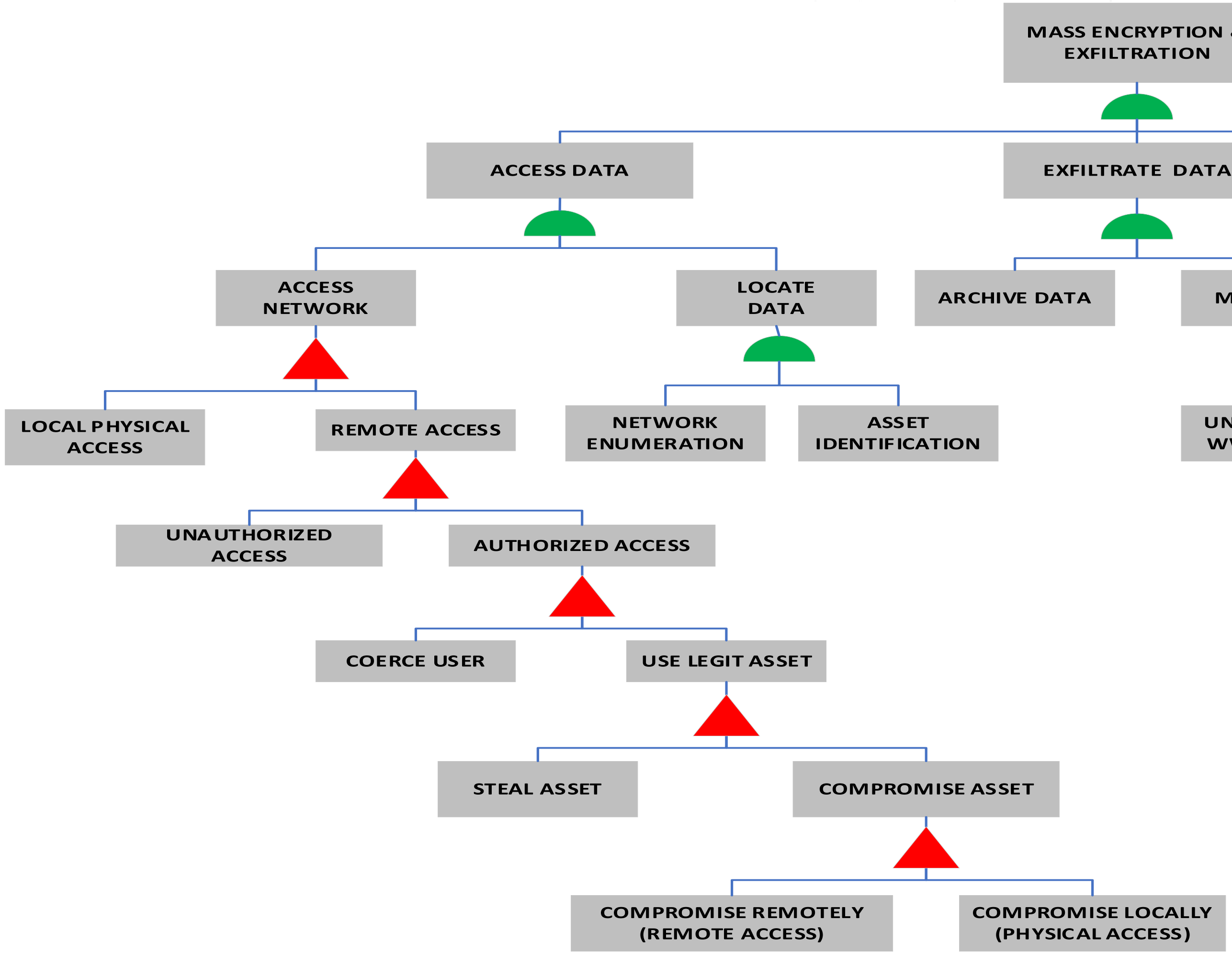
Key Insight: Penelope's laptop had to perform specific, detectable actions. What AI capabilities could have identified these necessary attack behaviors?

Discussion Prompts:

- **Attack Chain Analytics**
  How could AI models recognize the sequential pattern of reconnaissance → privilege escalation → lateral movement → data staging that Typhon must follow?
- **Impossible Travel Detection**
  What ML algorithms could flag suspicious impossibilities, like Penelope's credentials being used simultaneously at the NPP site and ARA headquarters?



Typhon Master Attack Plan



MASS ENCRYPTION & EXFILTRATION

ACCESS DATA

EXFILTRATE DATA

ENCRYPT DATA

ACCESS NETWORK

LOCATE DATA

ARCHIVE DATA

MOVE DATA

LOAD ENCRYPTOR

RUN ENCRYPTOR

LOCAL PHYSICAL ACCESS

REMOTE ACCESS

NETWORK ENUMERATION

ASSET IDENTIFICATION

UNRESTRICTED WWW ACCESS

UNAUTHORIZED ACCESS

AUTHORIZED ACCESS

COERCE USER

USE LEGIT ASSET

STEAL ASSET

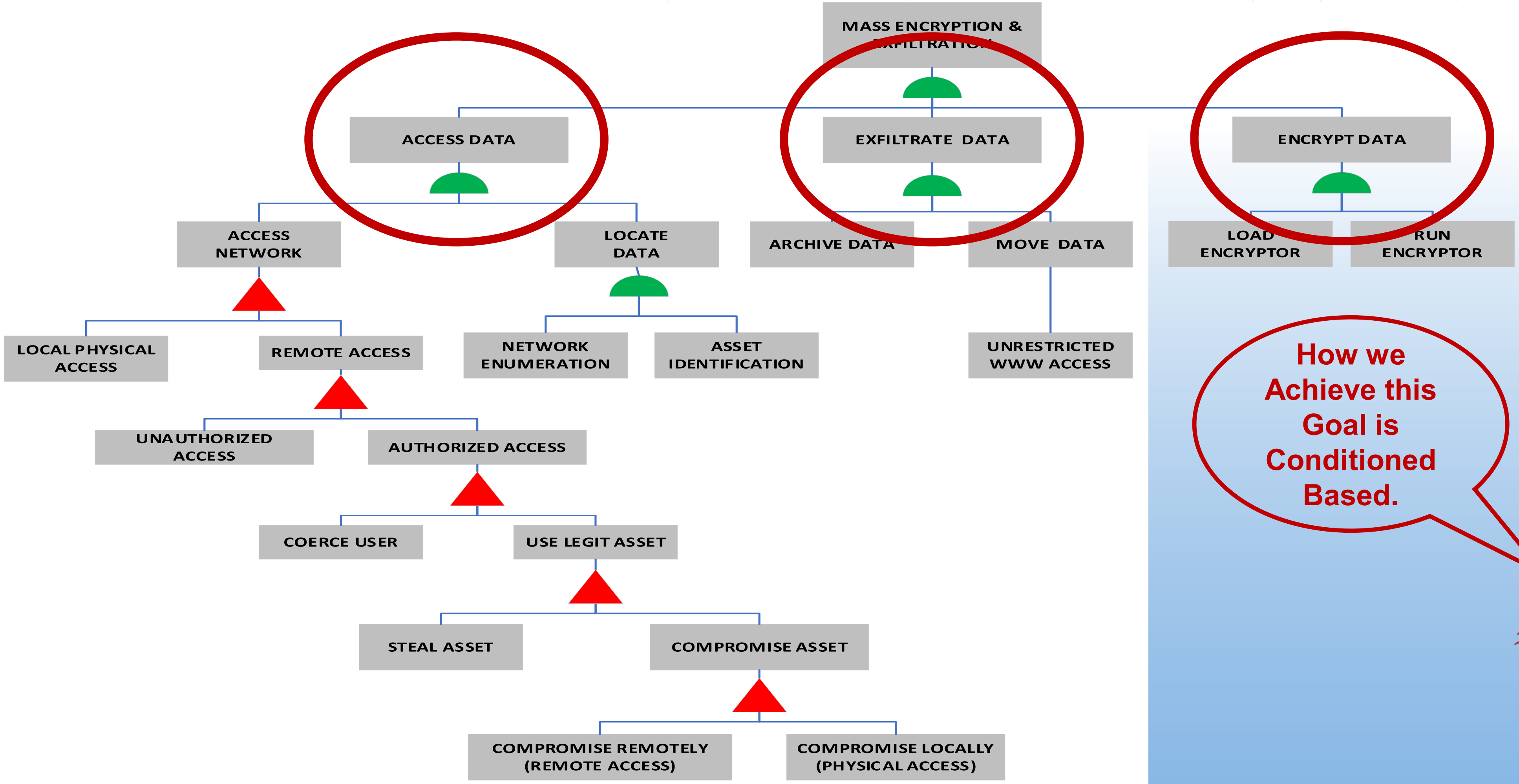
COMPROMISE ASSET

COMPROMISE REMOTELY (REMOTE ACCESS)

COMPROMISE LOCALLY (PHYSICAL ACCESS)

Our goals are consistent with all victims!

Typhon Conditions for Success



MASS ENCRYPTION & EXFILTRATION

ACCESS DATA

EXFILTRATE DATA

ENCRYPT DATA

ACCESS NETWORK

LOCATE DATA

ARCHIVE DATA

MOVE DATA

LOAD ENCRYPTOR

RUN ENCRYPTOR

LOCAL PHYSICAL ACCESS

REMOTE ACCESS

NETWORK ENUMERATION

ASSET IDENTIFICATION

UNRESTRICTED WWW ACCESS

UNAUTHORIZED ACCESS

AUTHORIZED ACCESS

COERCE USER

USE LEGIT ASSET

STEAL ASSET

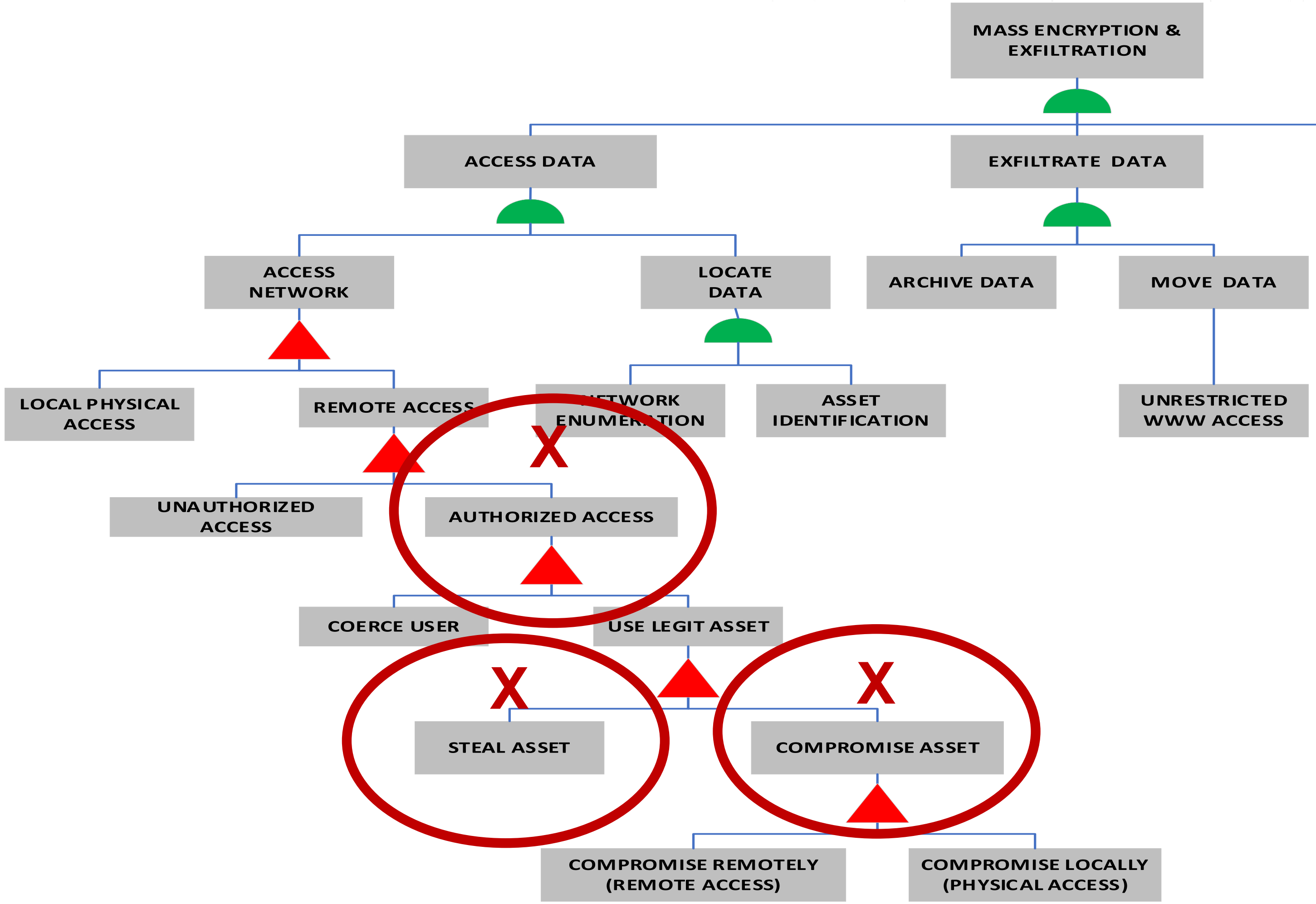
COMPROMISE ASSET

COMPROMISE REMOTELY (REMOTE ACCESS)

COMPROMISE LOCALLY (PHYSICAL ACCESS)

How we Achieve this Goal is Conditioned Based.

Low Probability of Meeting all Conditions



Attack Progression Frustrated Without Insider

The Aftermath: Typhon’s Impact

Measuring the Breach Consequences

100GB of Data Released

- Typhon publishes 100GB of stolen ARA data on dark web
- Leaked data includes:
  - *Security violation reports from nuclear facilities*
  - *Reactor design documentation*
  - *Confidential inspection reports*
  - *Personal information of all inspectors and licensees*

Analyzing the Impacts

- Public outrage about nuclear security breach
- Licensees lose trust in ARA's ability to protect sensitive data
- Criminal groups now targeting specific facilities using leaked data
- ARA operations paralyzed - all inspections suspended
- International regulatory credibility severely damaged

Forensic Timeline Reveals

March 1st @ 14:32

Thumb drive insertion detected during NPP inspection

March 1st @ 14:33

NPP's automated media scan showed "clean"

March 3rd @ 23:20

Typhon dormant until network connection

Attack triggered by Penelope’s valid login credentials

The Investigation Deepens: Was it Really Penelope?

Uncovering the Truth Behind the Attack

Penelope’s Account

- Insists all security protocols were followed during inspection
- Thumb drive was scanned per NPP's portable media policy
- No unusual files accessed, or suspicious behavior noted
- Passed polygraph; bank records show no large deposits

Critical Forensic Findings

- Typhon variant specifically designed to evade NPP's scanner
- Malware remained dormant until detecting ARA network
- Required valid inspector credentials to activate
- Attack sophistication suggests insider knowledge of both NPP and ARA systems

Artificial Intelligence Response Strategies: Post-Breach Analysis

Learning from the Incident

Discussion Prompts:

- **Behavioral Biometrics**
  How could AI analyze keystroke dynamics, mouse movements, and system interaction patterns to verify if it was Penelope using her credentials?
- **Anomaly Detection in Dormant Threats**
  What ML techniques could identify dormant malware by detecting subtle system changes even when the malware isn't actively executing?
- **Cross-Organization Threat Intelligence**
  How could federated learning allow NPPs and ARA to share AI threat detection models without exposing sensitive data?
- **Impact Scenario Modeling**
  How could Large Language Models (LLMs) analyze the leaked data to rapidly generate potential attack scenarios, helping ARA predict and prevent secondary attacks on specific facilities now exposed in the breach?

Another Insider: IT Tech Philippe

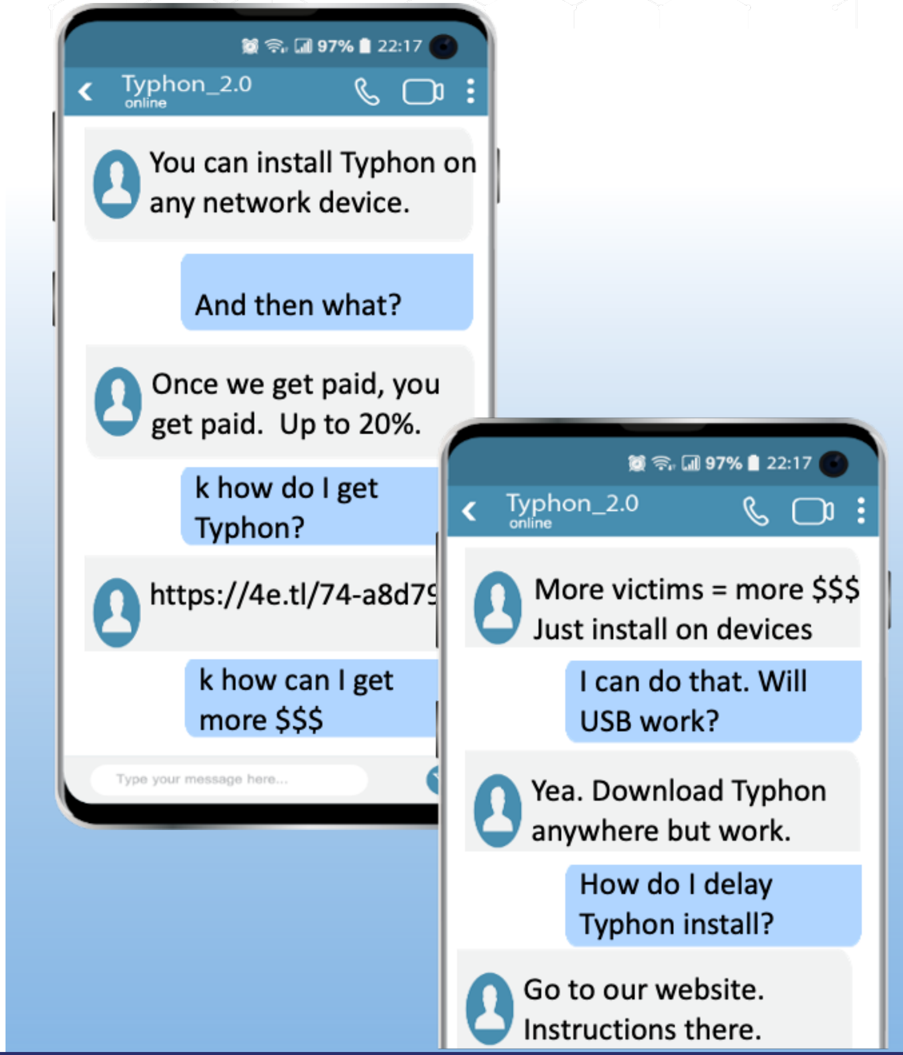
Discovery at the NPP

Parallel Attack at the NPP

- While ARA dealt with Typhon, the same NPP Penelope audited detected a similar attack
- NPP's cyber team caught Typhon before network installation
- Success due to ARA's rapid threat intelligence sharing
- NPP security began investigating all personnel with system access

Uncovering Philippe

- Philippe: NPP IT technician with privileged access to security systems
- Responsible for scanning visiting inspector equipment (including Penelope's laptop)
- Used his scanner access to plant Typhon on devices while showing "clean" scan results
- Confiscated phone revealed direct communication with Typhon group
- Admitted to installing malware on multiple inspector laptops for $500K



You can install Typhon on any network device.

And then what?

Once we get paid, you get paid. Up to 20%.

k how do I get Typhon?

https://4e.tl/74-a8d7

k how can I get more $$$

More victims = more $$$ Just install on devices

I can do that. Will USB work?

Yea. Download Typhon anywhere but work.

How do I delay Typhon install?

Go to our website. Instructions there.

Insider Threat Indicators: The Philippe Profile

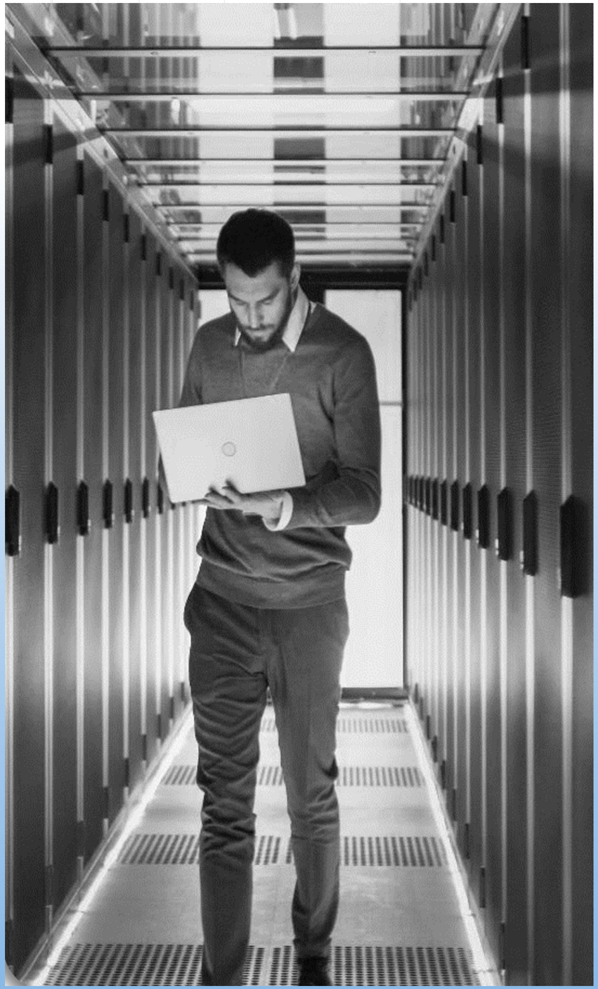
Recognizing the Warning Signs

Behavioral Red Flags

- Manager noted long-standing performance issues:
  - *Brilliant technically but felt routine tasks were "beneath him”*
  - *Hostile responses to criticism or error correction*
  - *Recent written reprimand for verbal abuse of coworkers*
  - *Increasing resentment toward management*
- Classic insider threat profile: disgruntled, technically capable, privileged access

The Pathway to Compromise

- Financial pressure + workplace grievances = vulnerability
- Technical skills + privileged access = capability
- External recruitment + perceived injustice = motivation
- Perfect storm for insider threat activation



AI-Powered Insider Detection: Behavioral Analytics

Early Warning Systems

Discussion Prompts

- **Sentiment Analysis & Communication Patterns**
  How could Natural Language Processing analyze emails, tickets, and chat logs to detect escalating negativity, hostility, or disengagement before it becomes a security risk?
- **Holistic Risk Scoring**
  What AI model could combine HR data (reprimands, reviews), access logs (after-hours, unusual systems), and behavioral indicators to create dynamic insider risk scores?
- **Peer Comparison Analytics**
  How could unsupervised learning identify employees whose behavior patterns deviate significantly from their peer group across multiple dimensions simultaneously?

Anatomy of the Attack: Three Types of Insiders

Understanding Insider Threat Profiles

Creating accurate threat profiles is foundational to preventing and detecting insider behaviors. Each type requires a different AI detection strategy.

Philippe - The Malicious Insider

- *IT technician with privileged scanner access. Motivated by financial gain and workplace grievances. Deliberately planted Typhon for $500K payment*
- ***AI Focus**: Behavioral changes, sentiment analysis, anomalous access patterns*

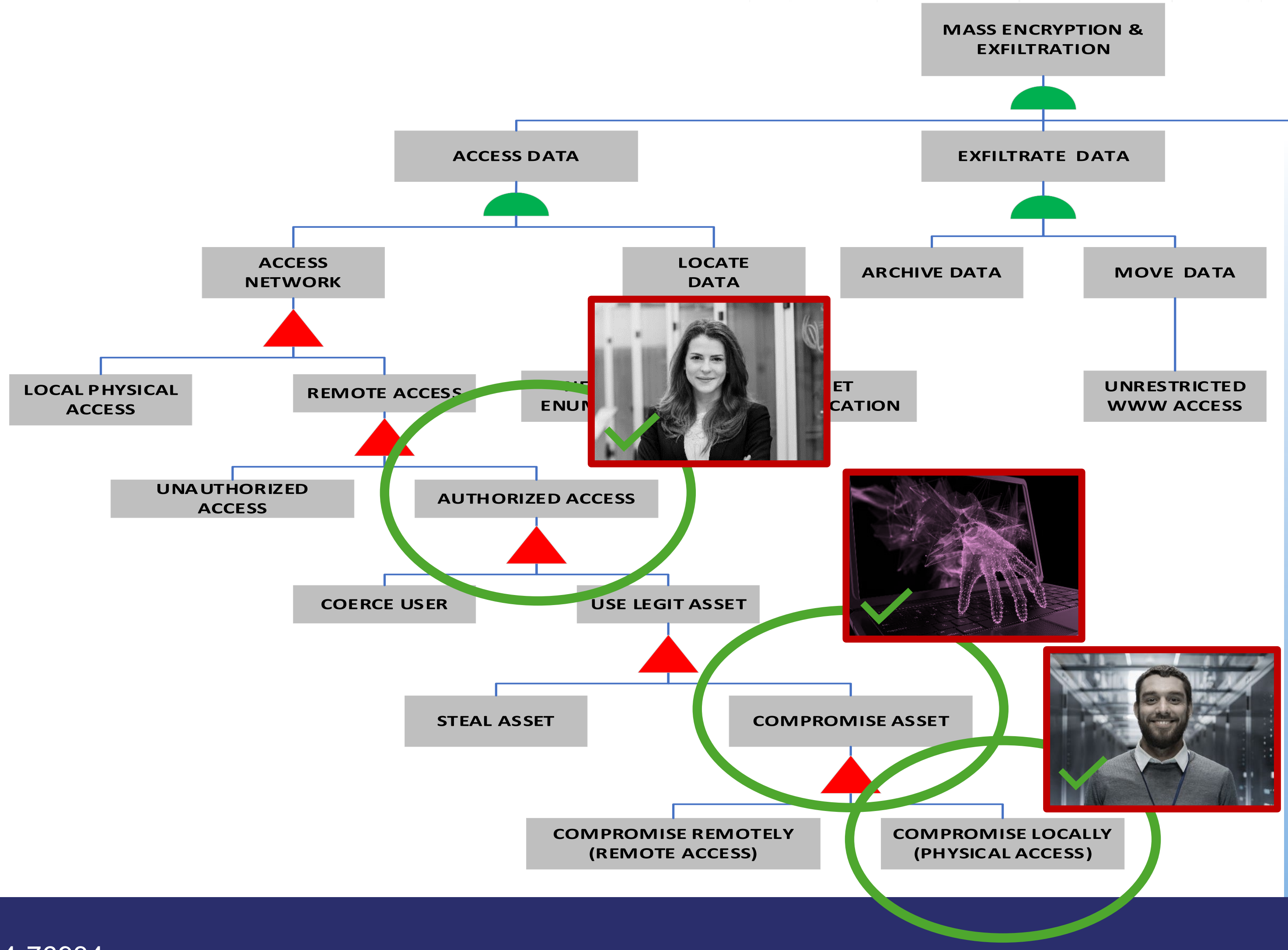
Penelope - The Unwitting Insider

- *Authorized inspector following all protocols. Laptop compromised without her knowledge. Became attack vector despite compliance*
- ***AI Focus**: Device integrity, unusual system behavior, credential misuse*

Technical Systems - The Automated Insider

- *Legitimate tools weaponized (scanners, update servers). Trusted processes exploited for malware delivery. No human intention but enables attack*
- ***AI Focus**: Process anomalies, unexpected connections, file behavior changes*

Attack Requirements: What Typhon Must Accomplish



MASS ENCRYPTION & EXFILTRATION

ACCESS DATA

EXFILTRATE DATA

ENCRYPT DATA

ACCESS NETWORK

LOCATE DATA

ARCHIVE DATA

MOVE DATA

LOAD ENCRYPTOR

RUN ENCRYPTOR

LOCAL PHYSICAL ACCESS

REMOTE ACCESS

ENU

ET CATION

UNRESTRICTED WWW ACCESS

UNAUTHORIZED ACCESS

AUTHORIZED ACCESS

COERCE USER

USE LEGIT ASSET

STEAL ASSET

COMPROMISE ASSET

COMPROMISE REMOTELY (REMOTE ACCESS)

COMPROMISE LOCALLY (PHYSICAL ACCESS)

Philippe - Compromise Locally (Physical Access)

- Plant malware using privileged scanner access during inspections
- Modify scan results to show "clean" while installing Typhon

Penelope - Authorized Access (Network Entry)

- Provide valid credentials for ARA network access
- Enable dormant malware to activate and spread using her privileges

Technical Systems - Compromise Asset (Attack Infrastructure)

- Allow Typhon to evade detection by security tools
- Provide unmonitored pathways for data exfiltration and encryption

Bottom Line: Remove ANY insider from this chain = Attack fails

Mitigating Malicious Insiders: The Philippe Prevention

Cross-Functional AI Approach

Behavioral Detection

- AI sentiment analysis on communications to flag escalating hostility
- ML models to identify employees matching disgruntlement patterns
- Automated alerts when behavioral risk scores exceed thresholds

Administrative Integration

- HR actions (reprimands) automatically trigger enhanced cyber monitoring
- AI correlates personnel issues with access logs and system usage
- Risk dashboard combines HR, security, and IT data streams

Technical Controls

- AI monitors privileged account usage for anomalous patterns
- ML algorithms detect unusual file modifications or scan result tampering
- Automated reporting on high-risk employee technical activities

Defense in Depth

Limit privileged access scope using AI-recommended least privilege models

Deploy behavioral biometrics to verify identity during critical operations

Physical access correlation with digital activities to detect policy violations

Protecting Unwitting Insiders: The Penelope Safeguards

Cross Functional AI Approach

Physical Security & Device Integrity

- AI-powered device tracking to alert when laptops leave inspector control
- Blockchain verification of device state before/after external scans
- Automated alerts for unauthorized hardware access attempts

Behavioral Monitoring

- ML baseline of normal inspector system usage patterns
- AI detection of commands/processes inspectors never typically run
- Real-time alerts when devices exhibit non-human behavior patterns

Technical Safeguards

- AI-monitored application whitelisting with anomaly detection
- ML analysis of network traffic for unusual data flows
- Automated USB/port lockdown with biometric override only

Defense in Depth

Network segmentation with AI monitoring cross-segment traffic

Role-based access with ML-powered privilege escalation detection

Continuous device health monitoring using behavioral analytics

Securing Technical Access: System-Level Defenses

Cross Functional AI Approach

Infrastructure Protection

- AI-powered anomaly detection for all system-to-system communications
- ML models trained on normal device behavior to flag deviations
- Real-time backup integrity monitoring with tamper detection

Process Security

- AI baseline of normal inspector laptop activities and connections
- ML detection of unusual process spawning or privilege escalation
- Behavioral analysis of all automated scanning and update processes

Access Control Evolution

- Dynamic least-privilege assignment based on AI risk scoring
- ML-powered separation of duties with automated compliance checking
- Zero-trust architecture with continuous authentication

Defense in Depth

Air-gapped AI system for backup verification and threat analysis

Immutable audit logs with ML anomaly detection

Automated rollback capabilities triggered by AI threat detection

Strategic Focus: Disrupting the Attack Chain

Targeting What Attackers Must Do

The Adversary Must:

- Recruit or compromise an insider (human or technical)
- Evade detection during initial access and lateral movement
- Maintain persistence while exfiltrating data
- Execute encryption without triggering response

Discussion Prompts

- **Condition-Based Detection Models**
  How can AI models be trained specifically on the non-negotiable steps attackers must take, rather than trying to predict all possible attack variations?
- **Automated Response Orchestration**
  What ML systems could instantly recognize when critical attack conditions are met and automatically trigger isolation, rollback, or defensive actions?
- **Continuous Learning from Near-Misses**
  How can AI systems learn from attacks like Typhon to continuously refine detection of 'must-do' adversary behaviors across the entire ARA infrastructure?

Fundamental Principle

Don't chase infinite possibilities.
Target necessary adversary actions.

AI should monitor the few things adversaries cannot avoid, not the many things they might try.

Contact

Chris Spirito, Idaho National Laboratory

Email: christopher.spirito@inl.gov