RSAConference™2024
San Francisco | May 6 – 9 | Moscone Center

THE ART OF POSSIBLE

SESSION ID: PART4-W02

AI-Equipped Threat Actors vs. AI-Enhanced Cyber Tools: Who Wins?

#RSAC

Shil Sircar
SVP Engineering & Data Science
BlackBerry
X: @sircar

Disclaimer

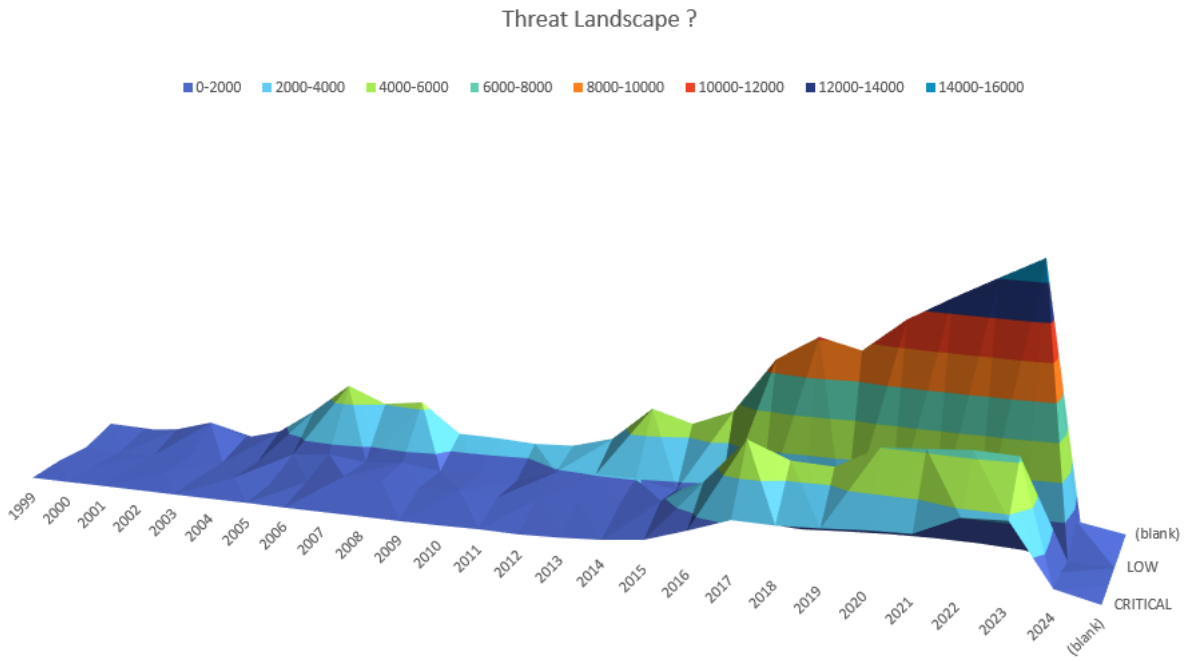
Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference™ or any other co-sponsors. RSA Conference does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

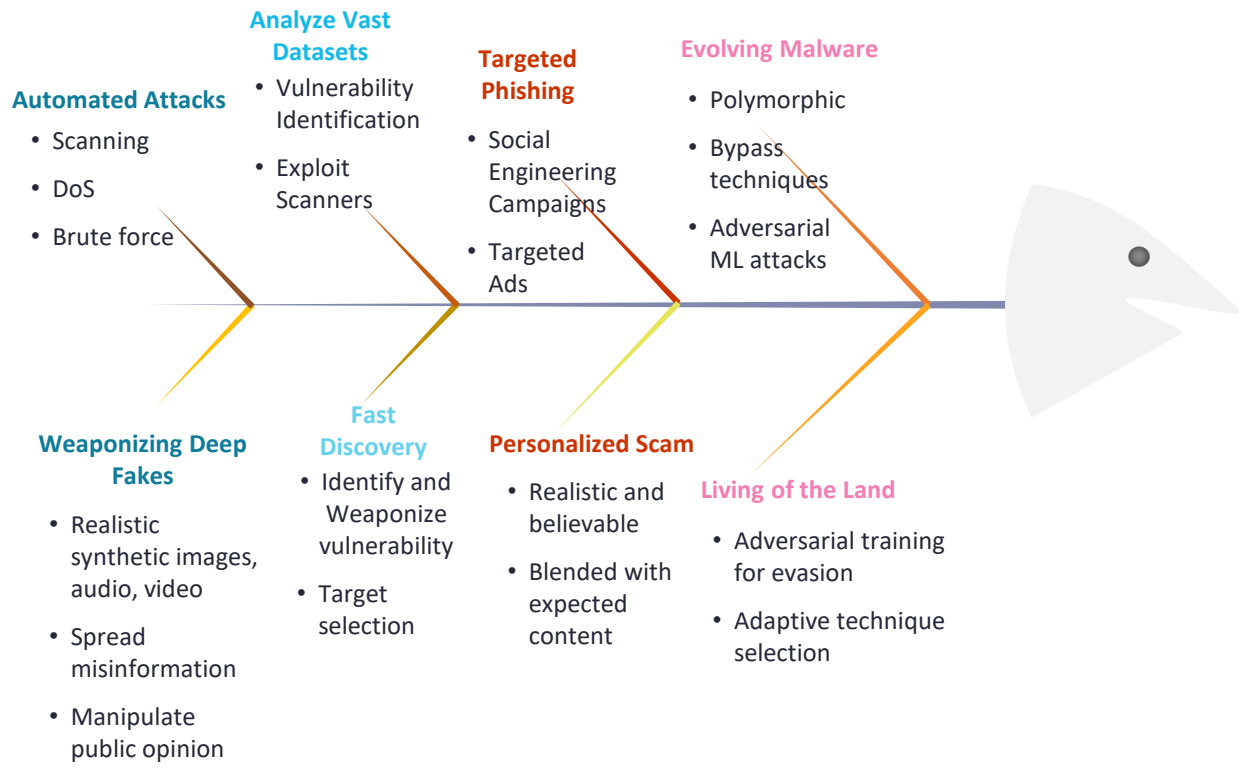
Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

Agenda

- AI-Powered Threat Landscape
- Surge in Novel Cyberattacks
- Predictive Approaches to Defense
- Observed Outcomes
- Key Takeaways

Facing the AI-Powered Threat Landscape

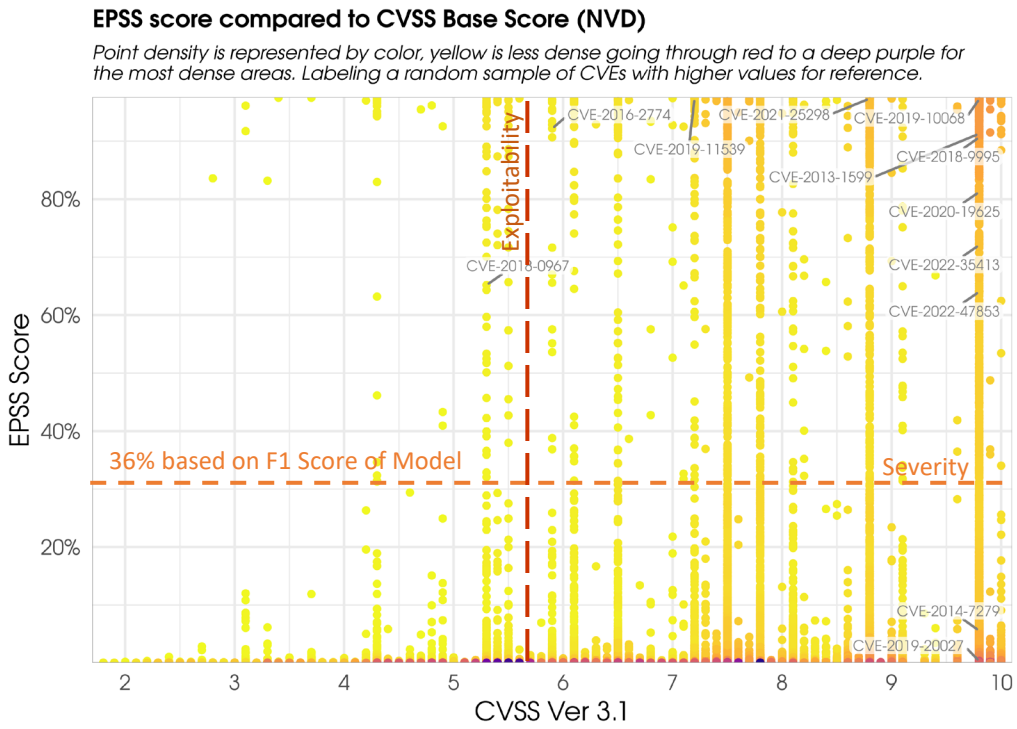




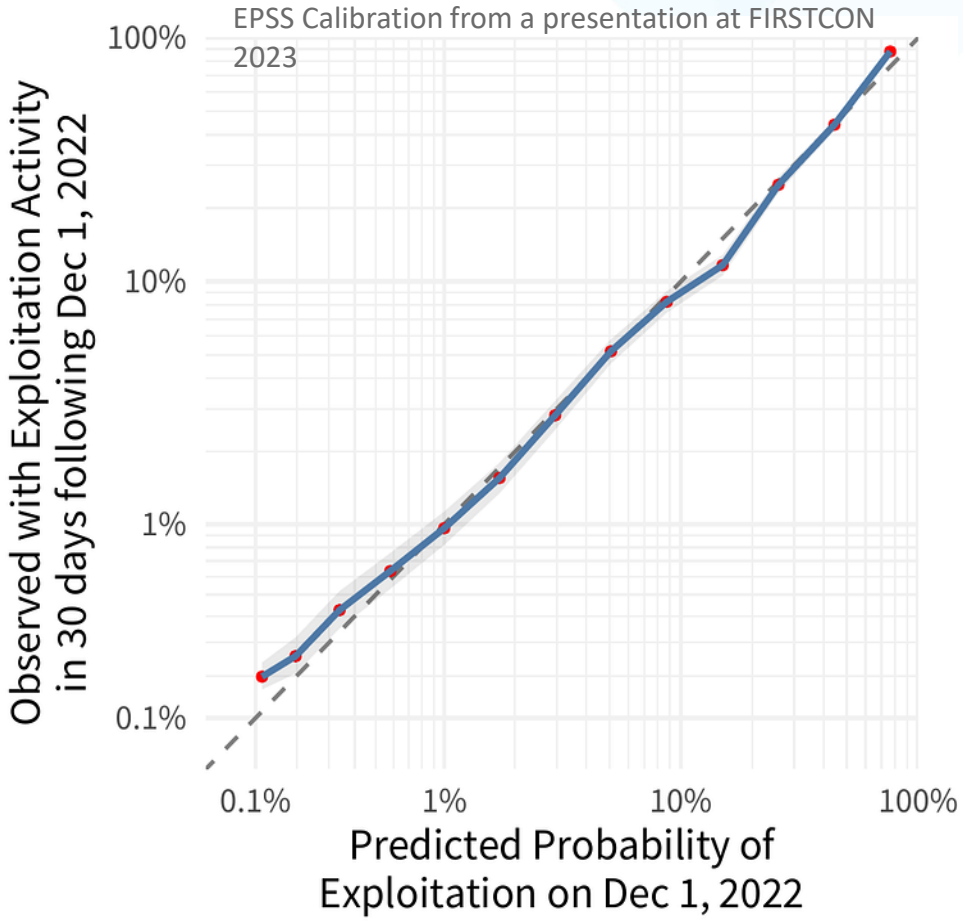
BlackBerry CYLANCE

Discovery - Exploit Prediction

- EPSS predicts probability of a CVE being exploited in 30 days.
- Augment EPSS output with environment specific attributes to prioritize

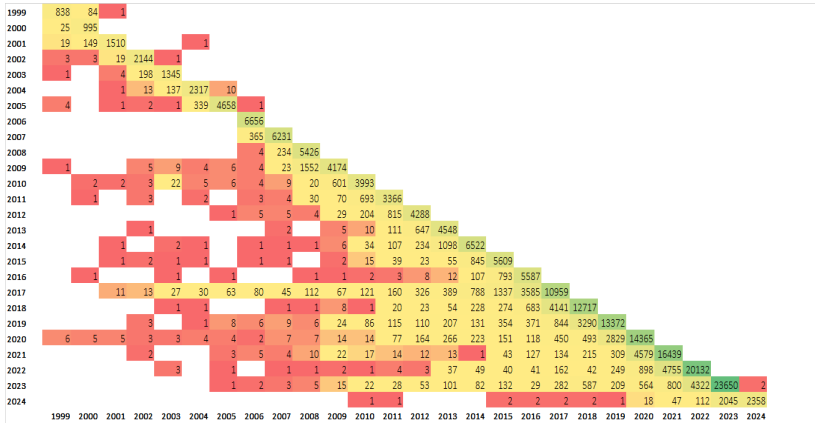


Source: https://first.org/epss/data_stats, 2024-02-28

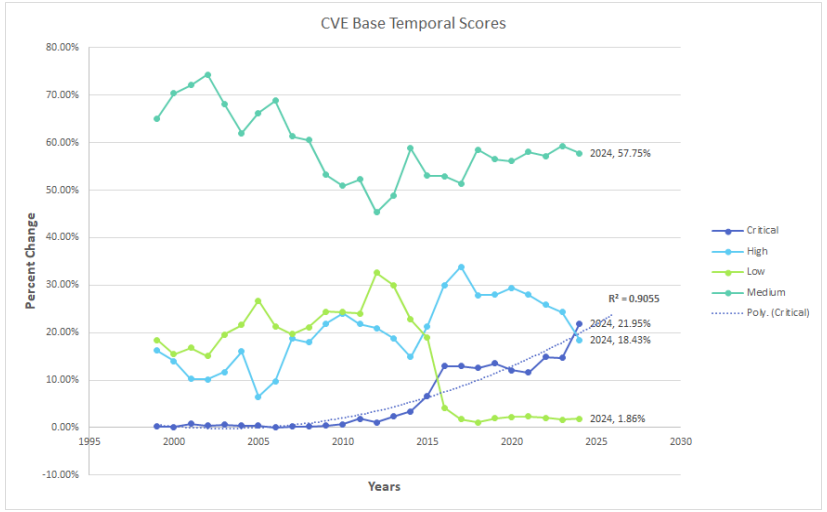


BlackBerry CYLANCE

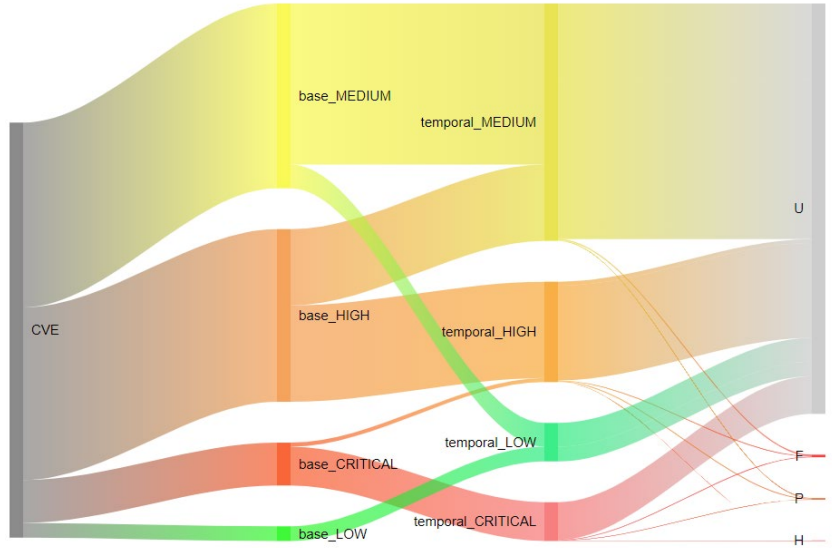
Analysis and Prioritization



Disparity between CVE year and publication year



CVE Base Temporal Score as rate over time



CVE -> CVSS Score -> Exploit Probability Score

Defensive

- Understand vulnerability and exposure to manage
- Threat zero mitigations

- Threat Zero protect first
- Correlate EPSS Predictive score to Inventory

- Cluster cisa_kev, metasploit, epss, nuclei, poc_github and prioritize Functional , PoC, High
- Reduce noise to better focus

Offensive

- Time to identify and exploit vulnerability
- target profile discovery org->product->vulnerability

- Richer threat landscape more targets
- EPSS to discover exploitable vulnerability

- No obvious course

BlackBerry CYLANCE

AI & Generative AI as a Force Multiplier

Leveraging AI and Gen-AI models

Defensive

Data Analysis & Synthesis

- Find & Fix vuln. and mis-configurations
- Correlate Security events
- Reveal probabilities of attack paths

Content Generation

- Generate code
- Make code more secure
- Generate Blue Team Tests

Automation & Augmentation

- Automate Incident response
- Generate and prioritize patch
- Summarize malicious code

Offensive

- Find and exploit vulnerability and configurations
- Locate attack surfaces and exploit
- Generate Novel Malware
- Create better phishing emails
- Modify malware to evade detections
- Automate low effort disruption attacks
- Offer Phishing and Malware-As-Service

Securing AI and Gen-AI Models

Model Development

Secure development environment and OSS/Vendor Supply chain

Model Deployment & Integrity

Secure Runtime environment
Apply Least privilege OWASP top 10

Data Privacy

Differential Privacy and De-identification with Encryption

Agenda

- AI-Powered Threat Landscape
- Surge in Novel Cyberattacks
- Predictive Approaches to Defense
- Observed Outcomes
- Key Takeaways

BLACKBERRY **STOPPED 3.3 MILLION** CYBERATTACKS .

~26 ATTACKS EVERY MIN.

70% INCREASE FROM PREVIOUS REPORTING PERIOD

2.9 NEW UNIQUE MALWARE SAMPLES PER MINUTE.

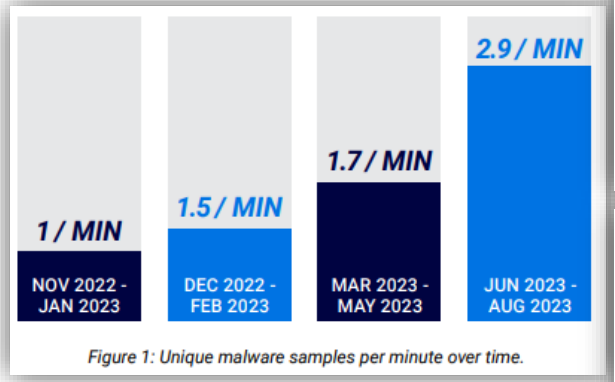


Figure 1: Unique malware samples per minute over time.

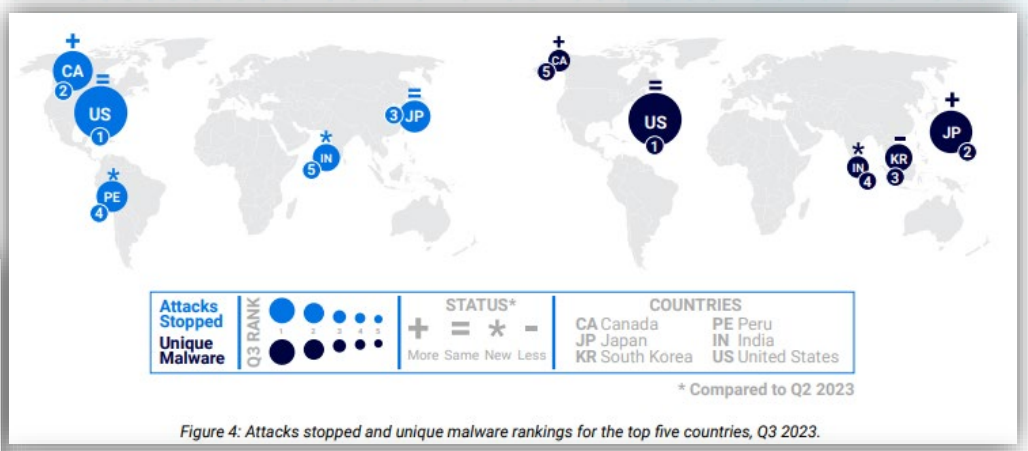


Figure 4: Attacks stopped and unique malware rankings for the top five countries, Q3 2023.

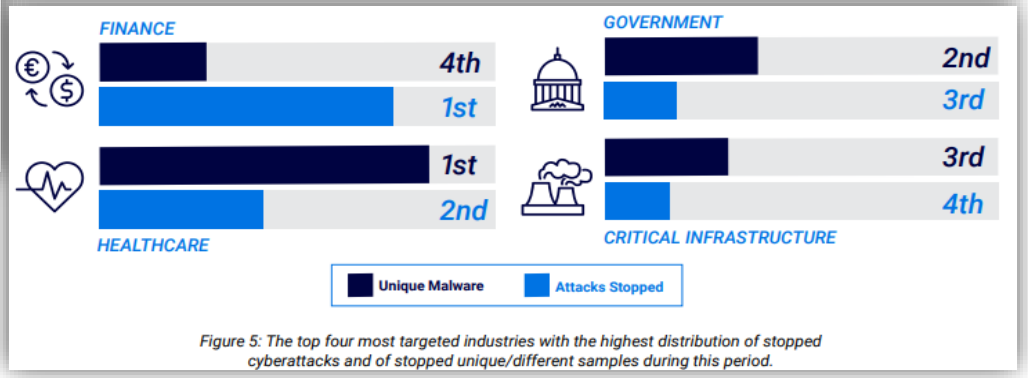
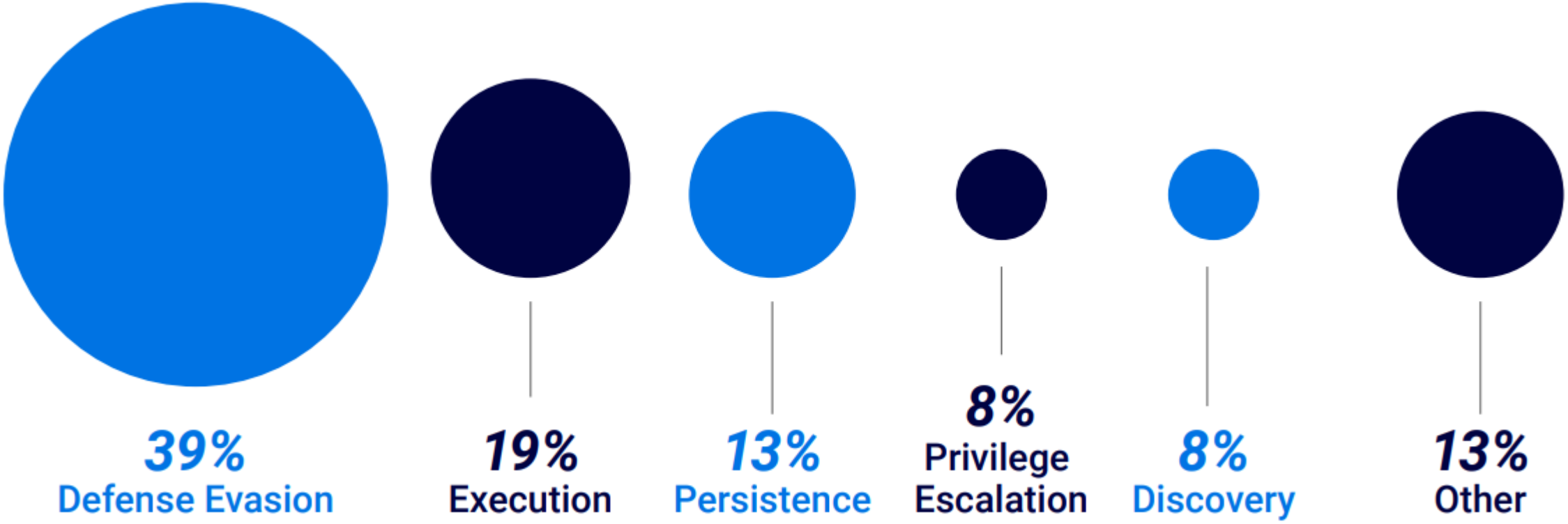


Figure 5: The top four most targeted industries with the highest distribution of stopped cyberattacks and of stopped unique/different samples during this period.

BlackBerry CTI Report – Nov 2023

From Traditional Threats to AI-Driven Attacks

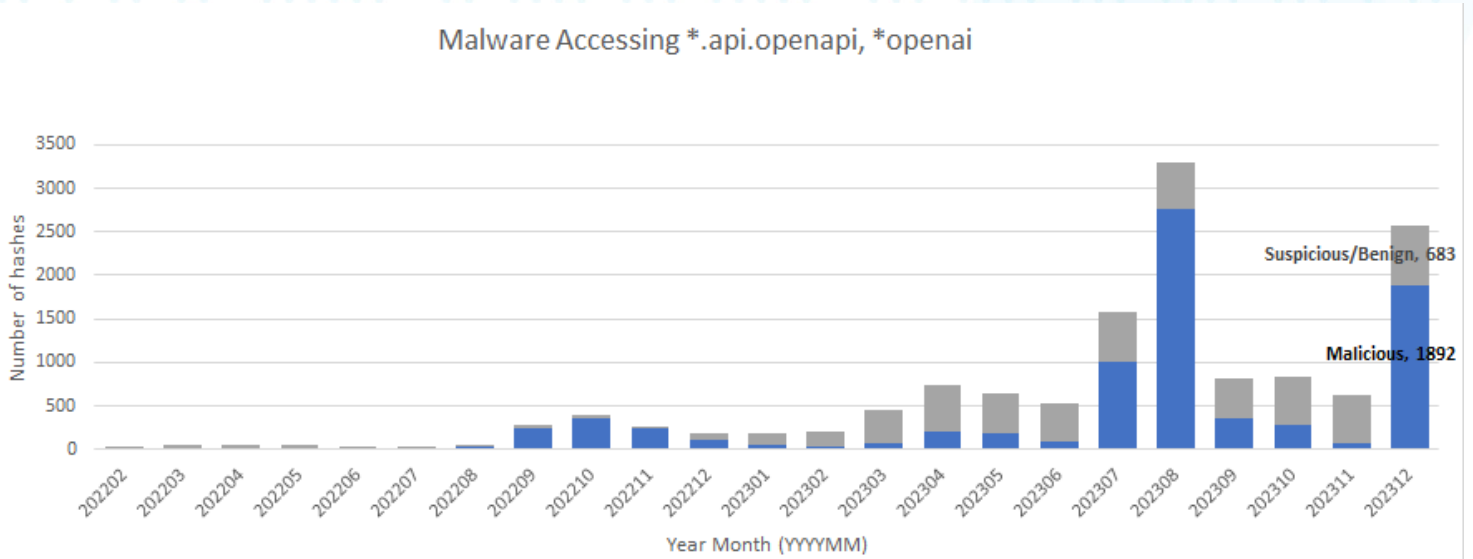
Shift from Traditional cyber threats to AI-assisted ? attacks, highlighting the most prevalent portion to be defense evasion.

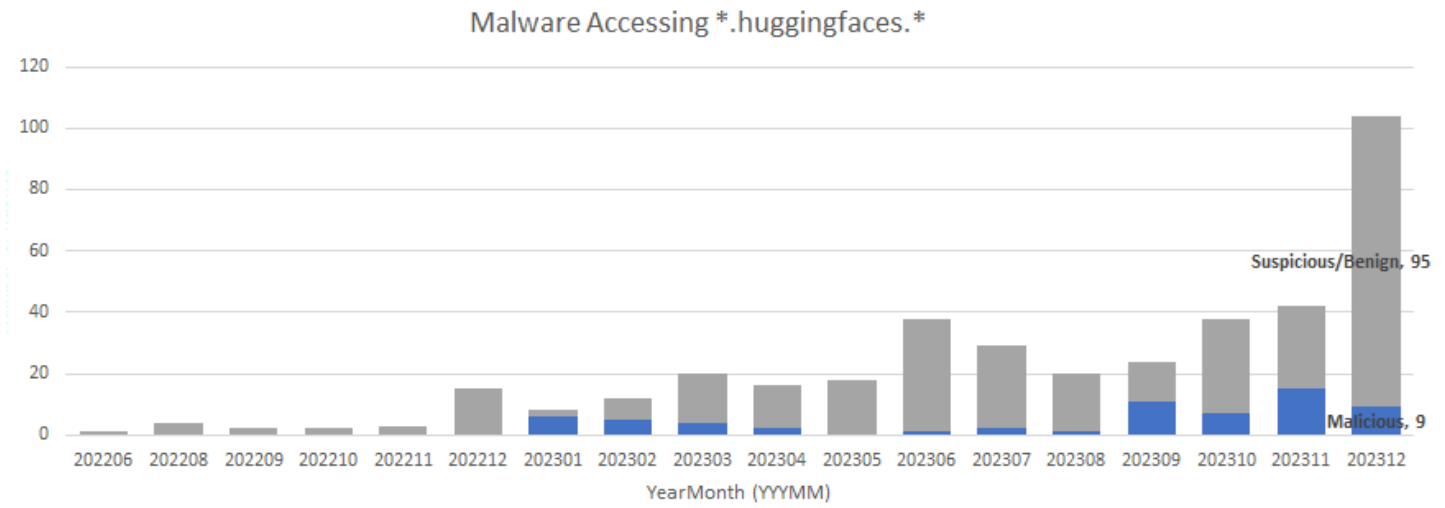


Tactics From BlackBerry CTI Report – Nov 2023

Observations from Prevalence Data

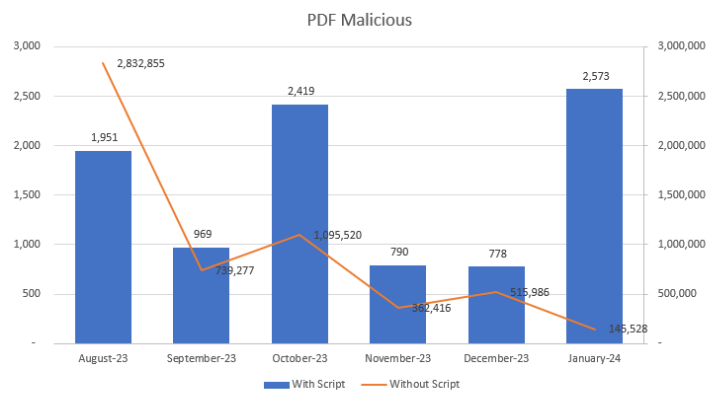
- Malware ML model convicts that touch OpenAI API
- Malware our models convict that are using hugging face site

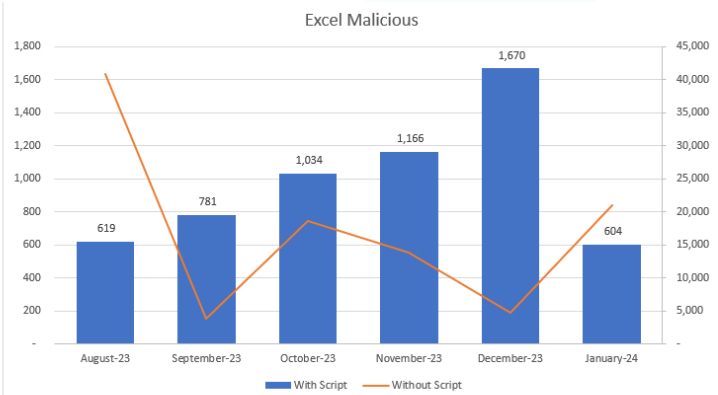


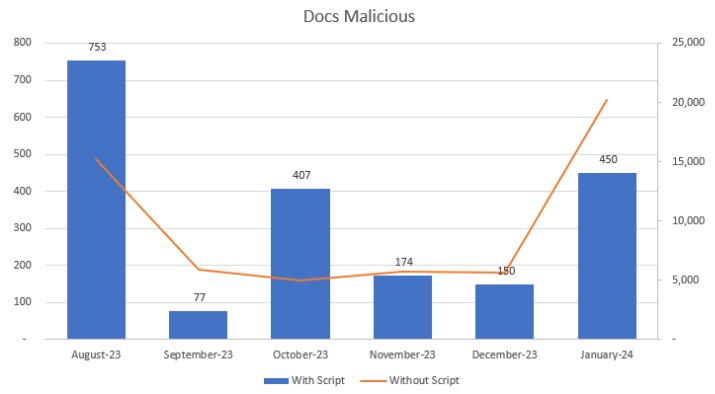


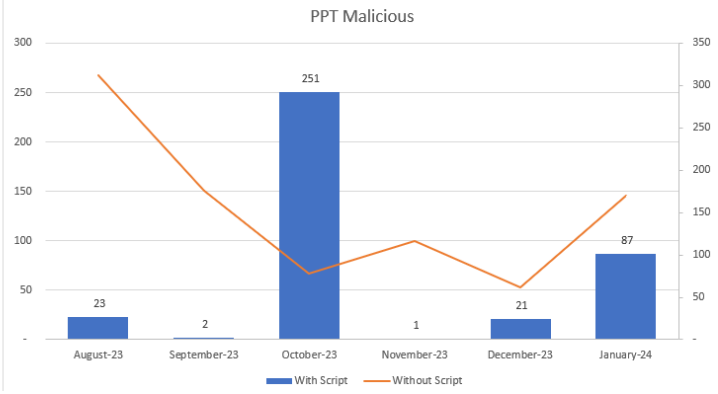
Observations from Prevalence Data

- Malicious document files that ML Model convicts
- Doc files exhibit maliciousness without embedded scripts



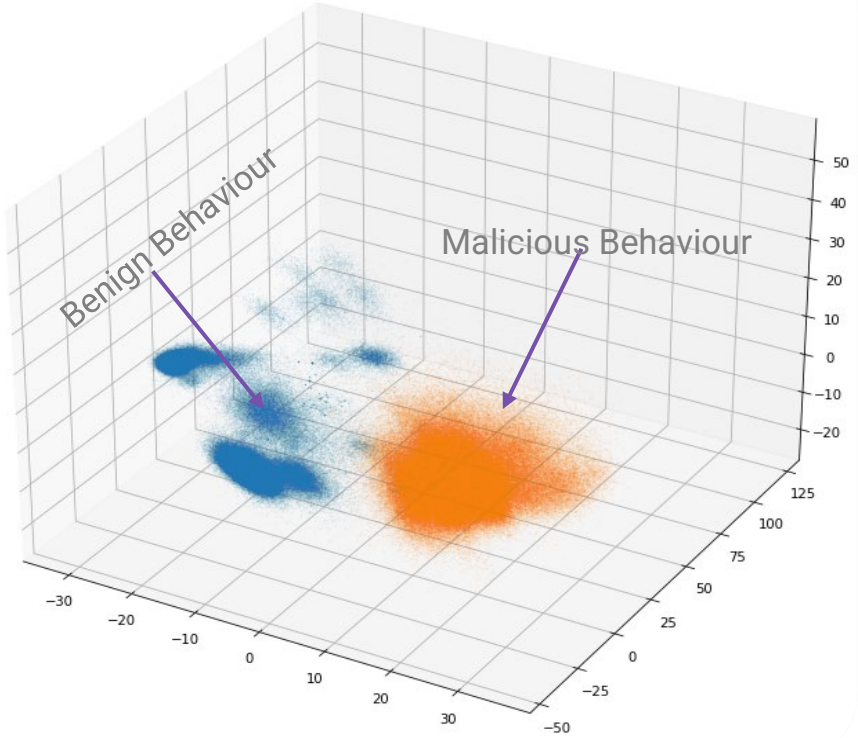




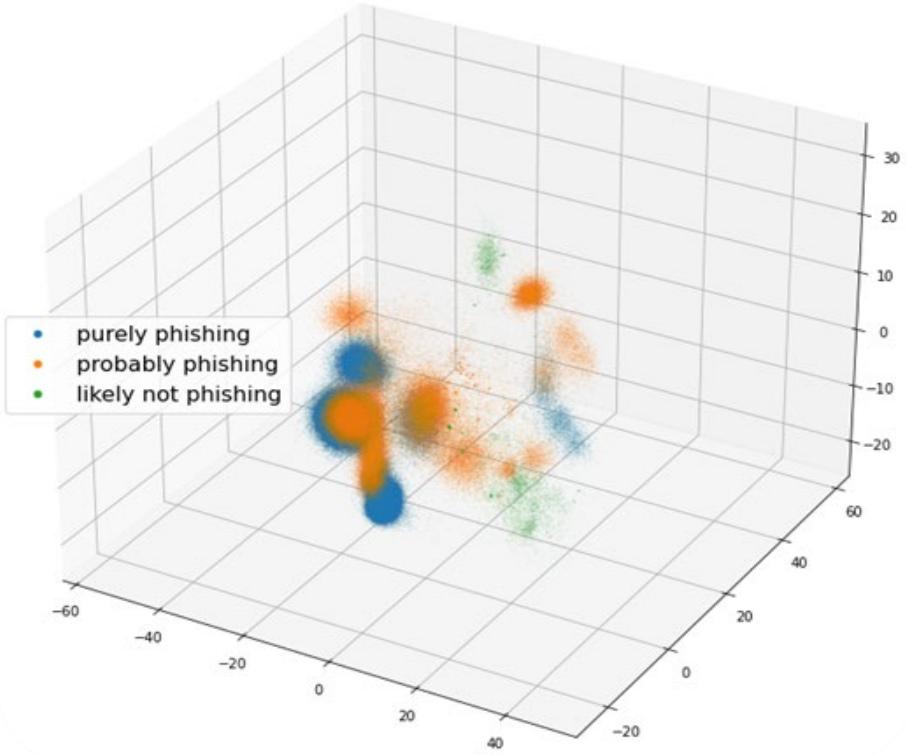


Probing-In Documents - With and Without Scripts

PDF Document Model



Analysing for Content

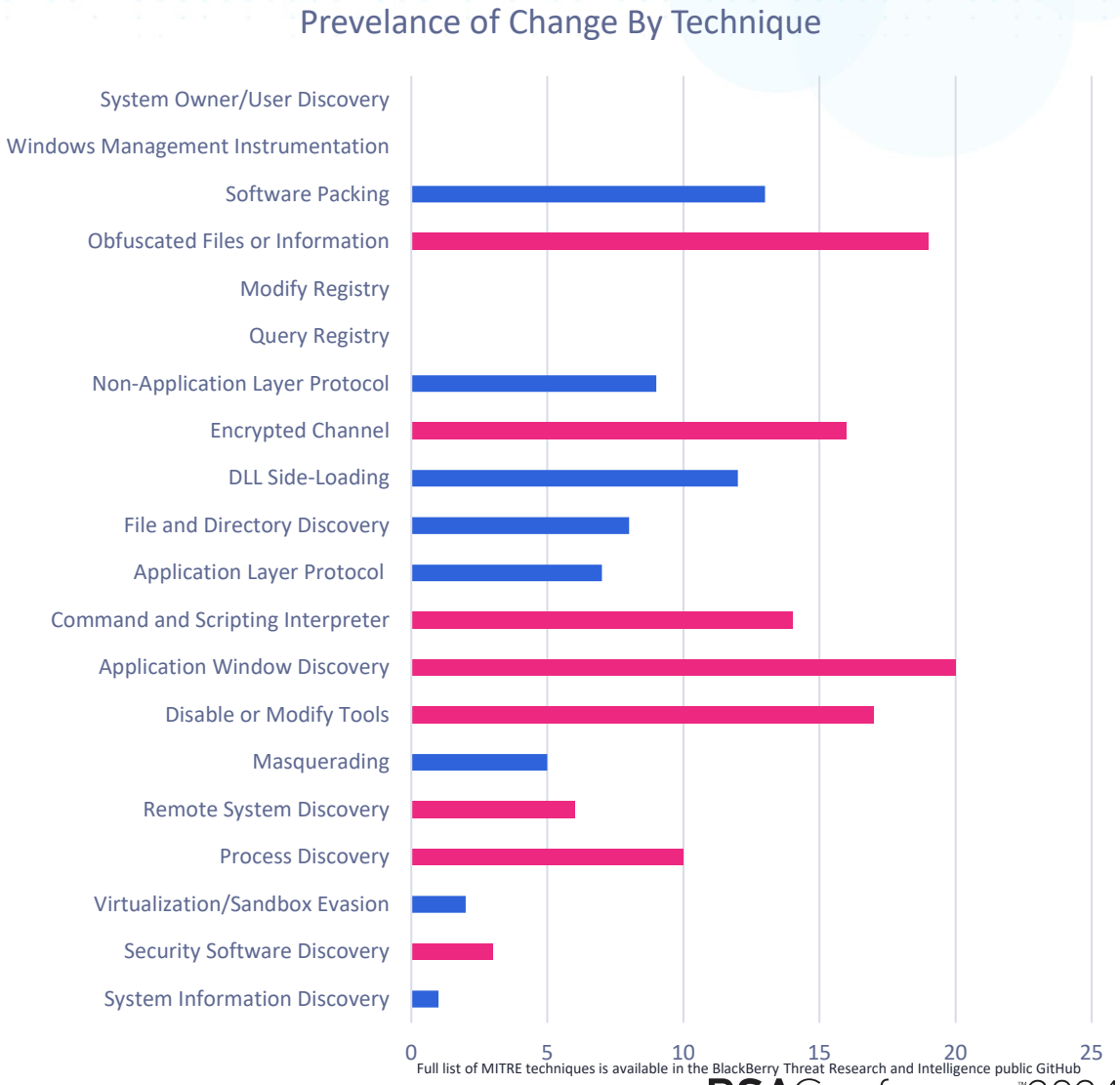


Agenda

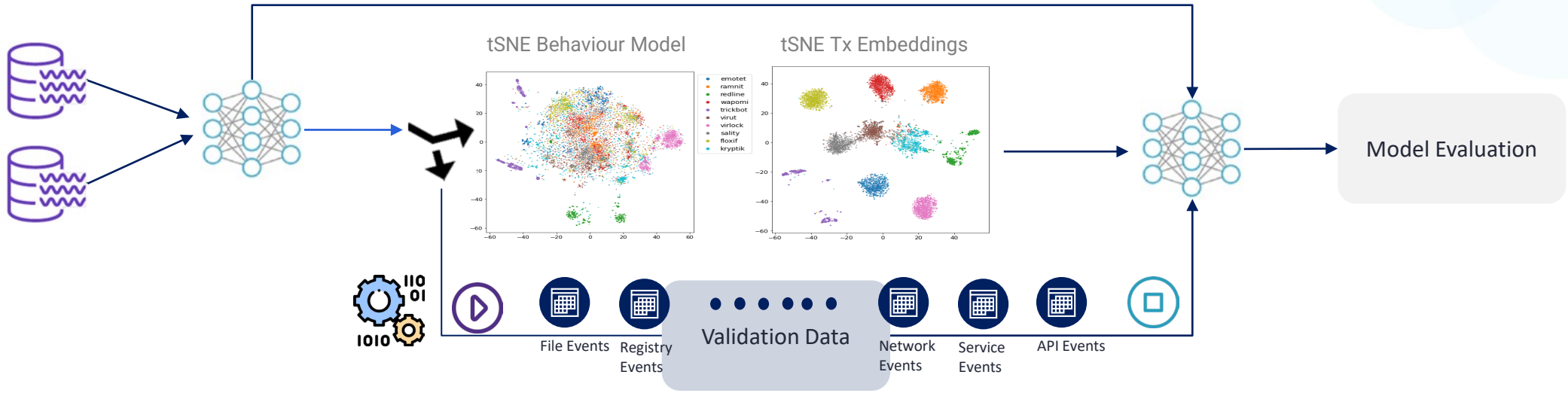
- Surge in Novel Cyberattacks
- AI-Powered Threat Landscape
- Predictive Approaches to Defense
- Observed Outcomes
- Key Takeaways

Intent and Behavior cannot be hidden

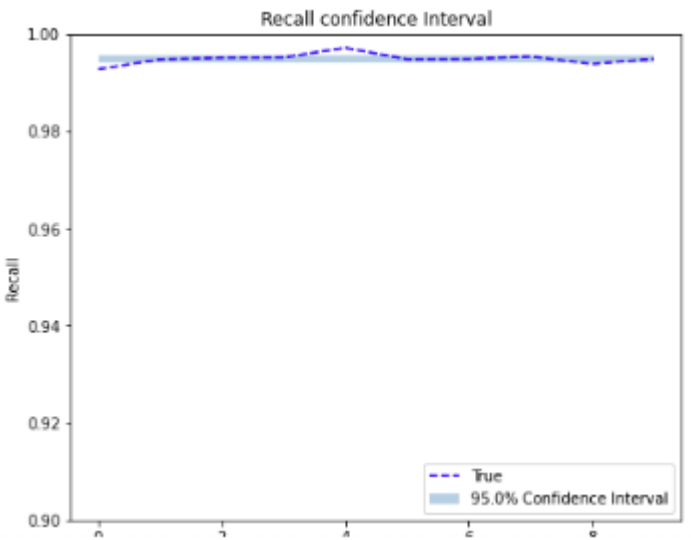
- Observed behaviour taken by threat groups – Nov 2023 Report
- Prioritization based on these top 20 techniques
- Discovery tactic is the most prevalent, associated with four of the five techniques in the top five.

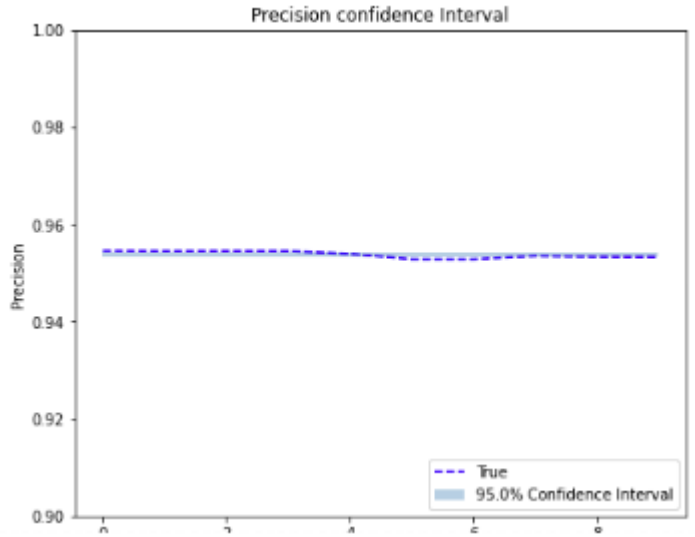


Process Behavior prediction



How confident are we with Models Recall And Precision ?



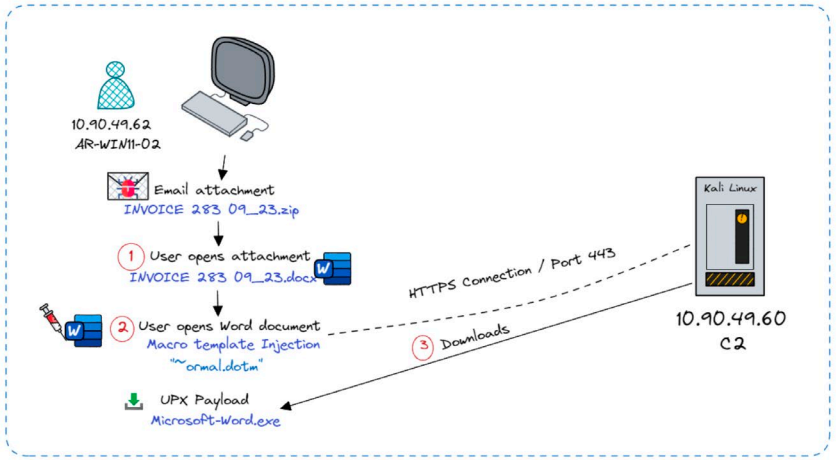


BlackBerry CYLANCE

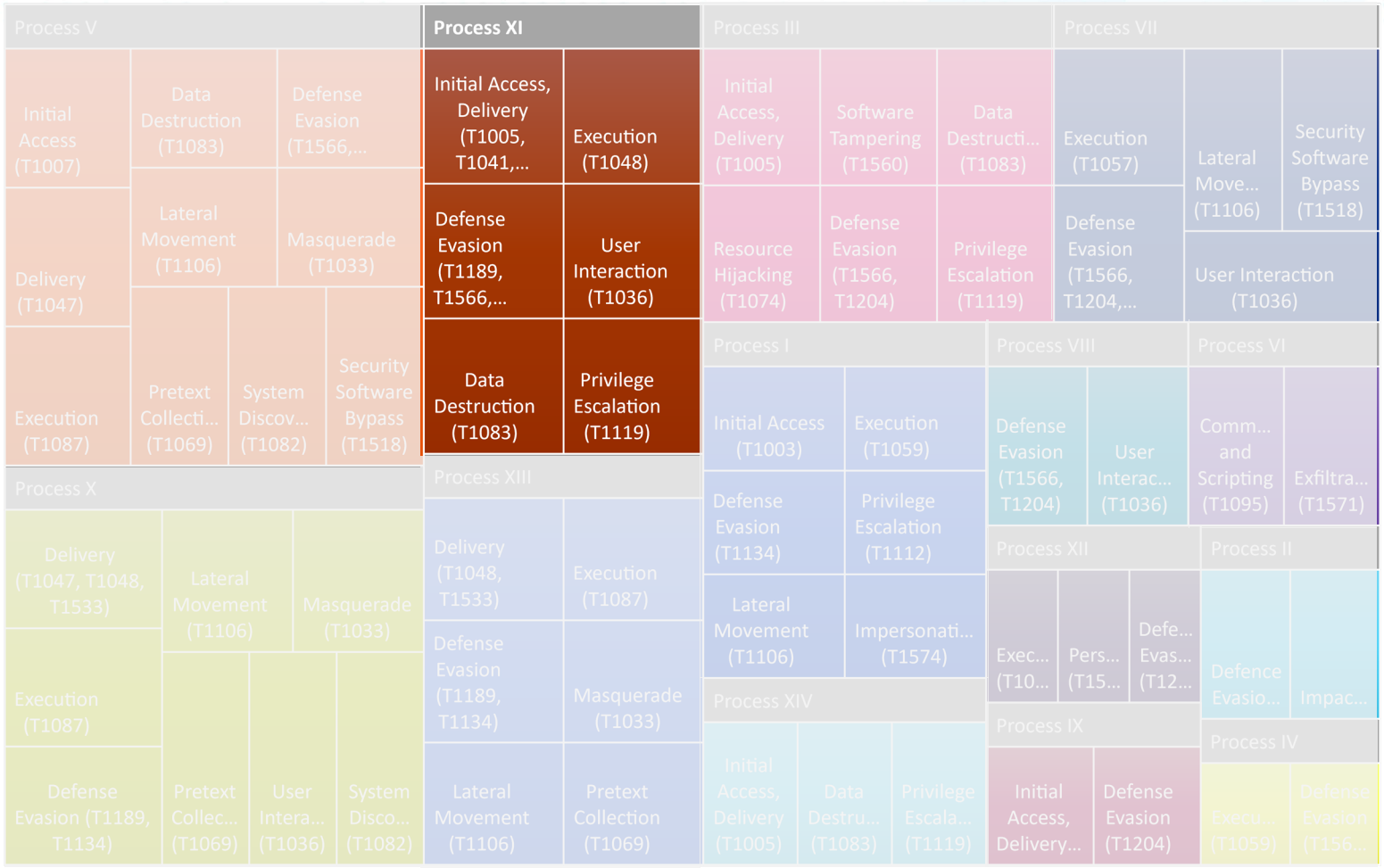
Agenda

- Surge in Novel Cyberattacks
- AI-Powered Threat Landscape
- Predictive Approaches to Defense
- Observed Outcomes
- Key Takeaways

Step 1

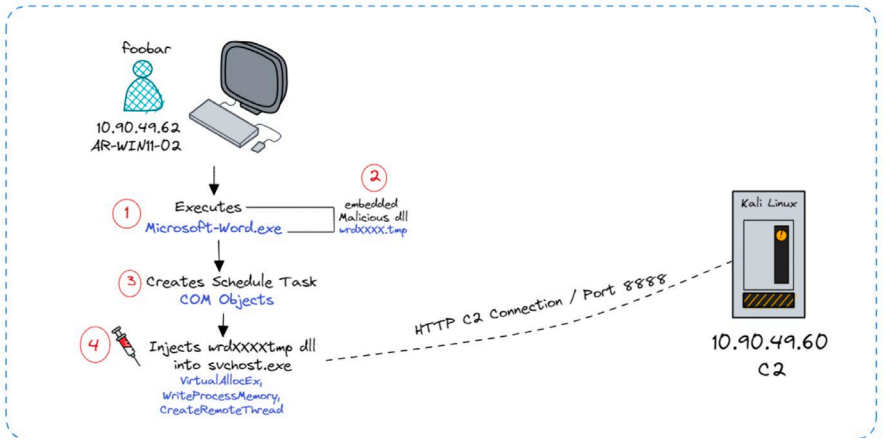


Malicious Behaviour Detections based on Mitre TTP

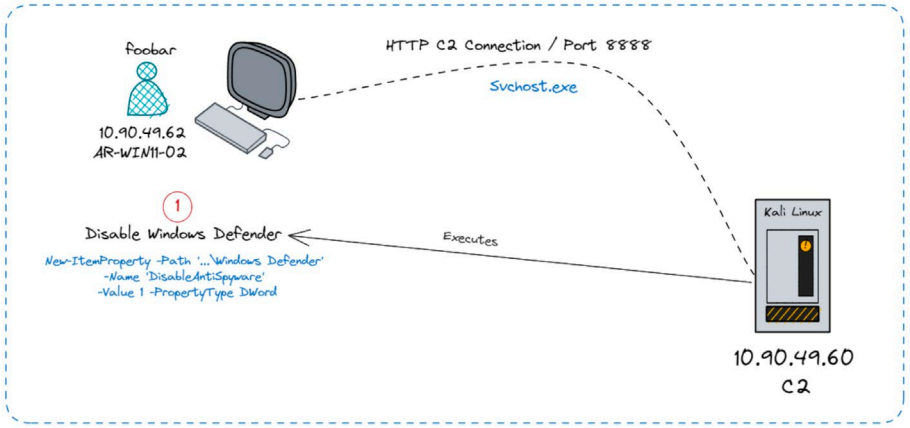


Model Score (maliciousness): 0.9883317

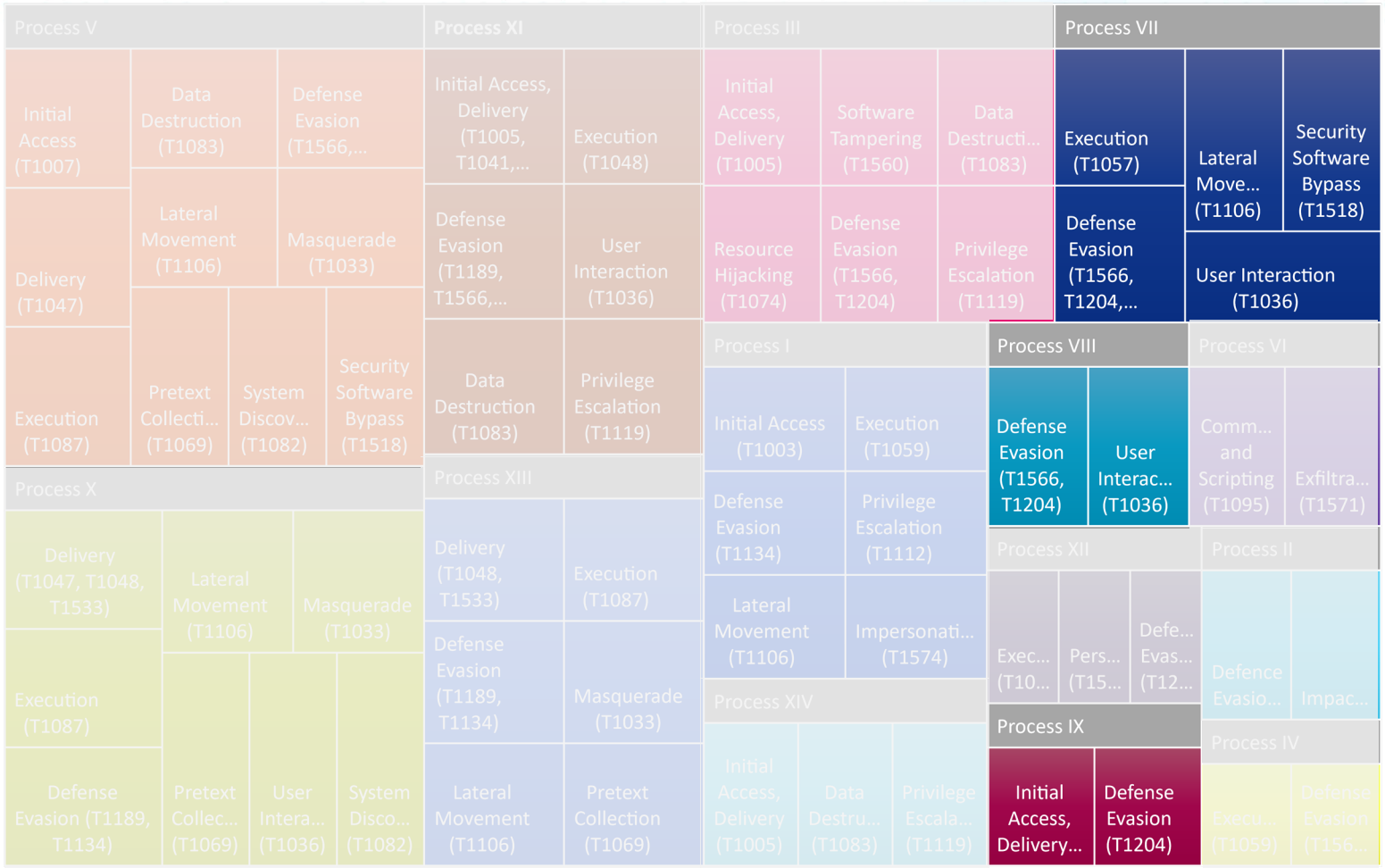
Step 2



Step 3

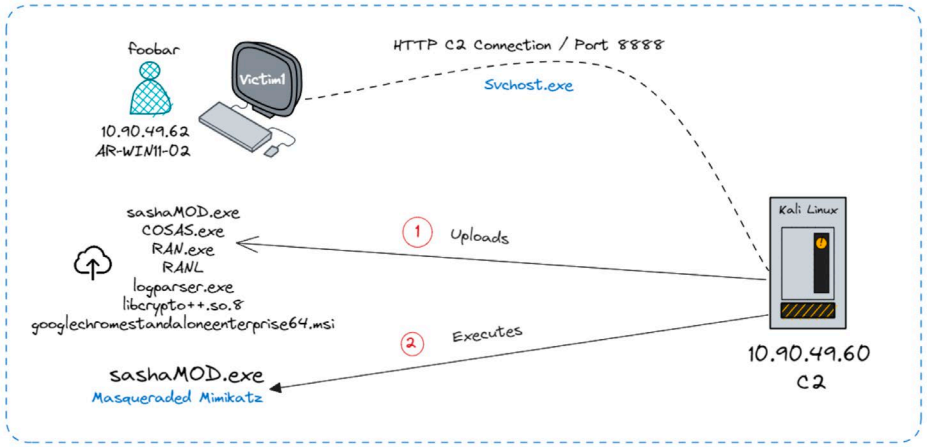


Malicious Behaviour Detections based on Mitre TTP

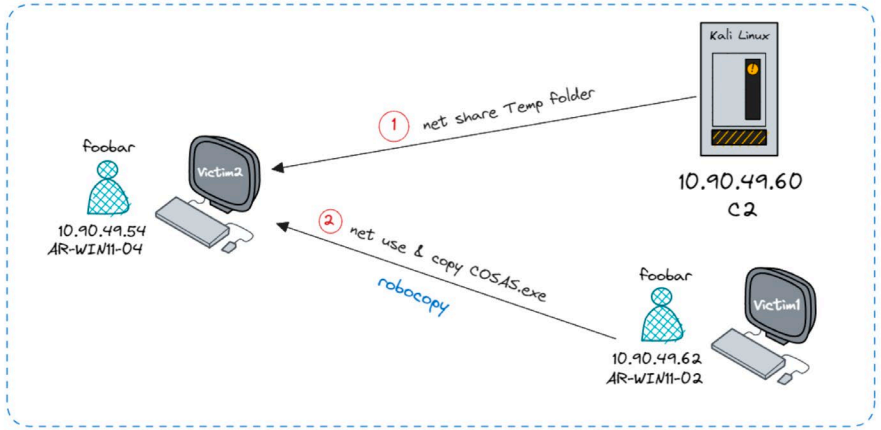


Model Score (maliciousness): 0.97976273, 0.96840143, 0.82308644

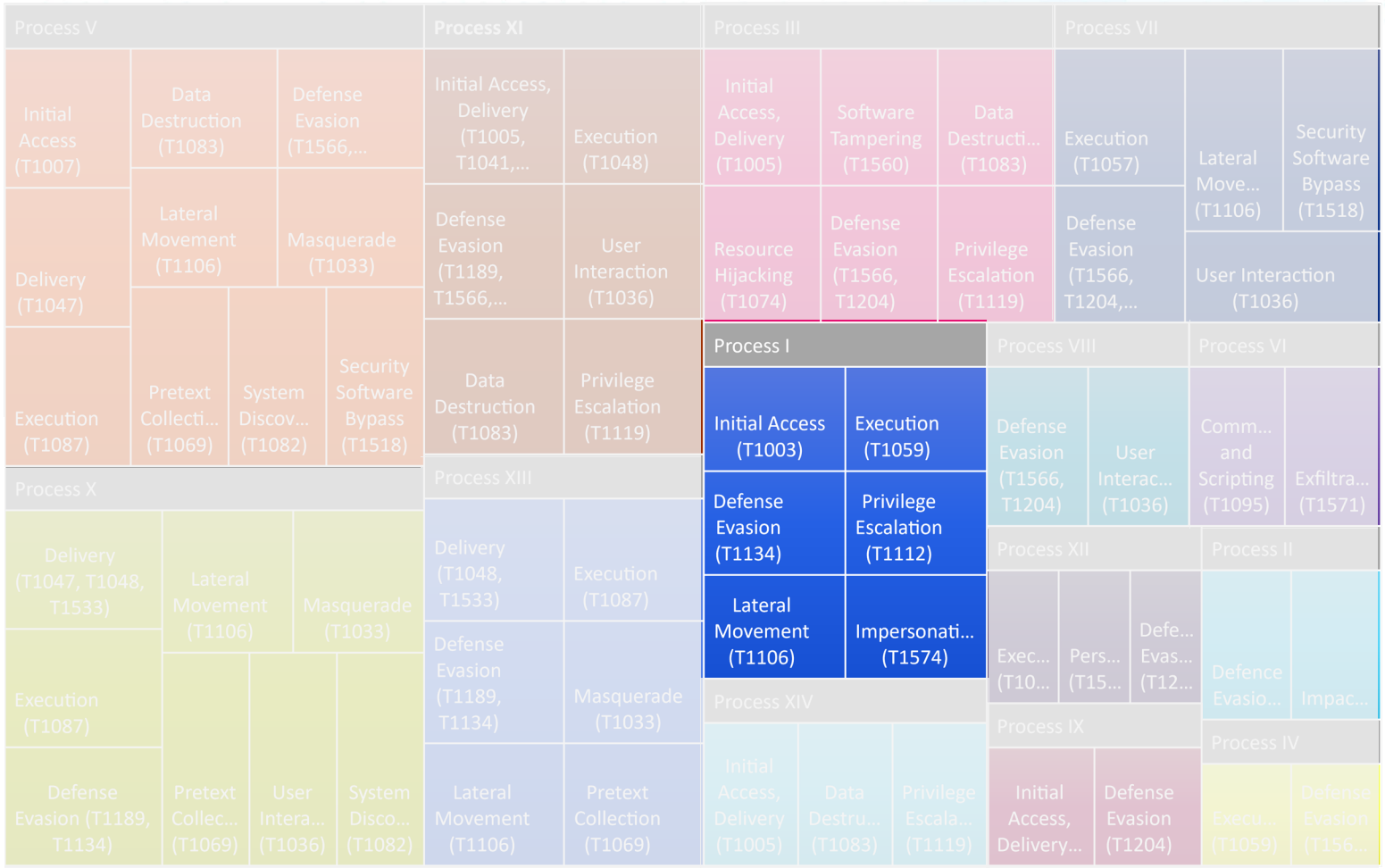
Step 4



Step 5

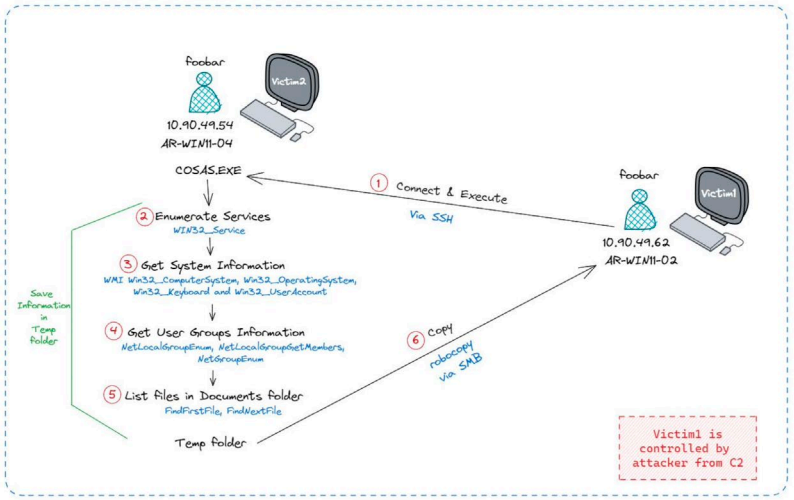


Malicious Behaviour Detections based on Mitre TTP

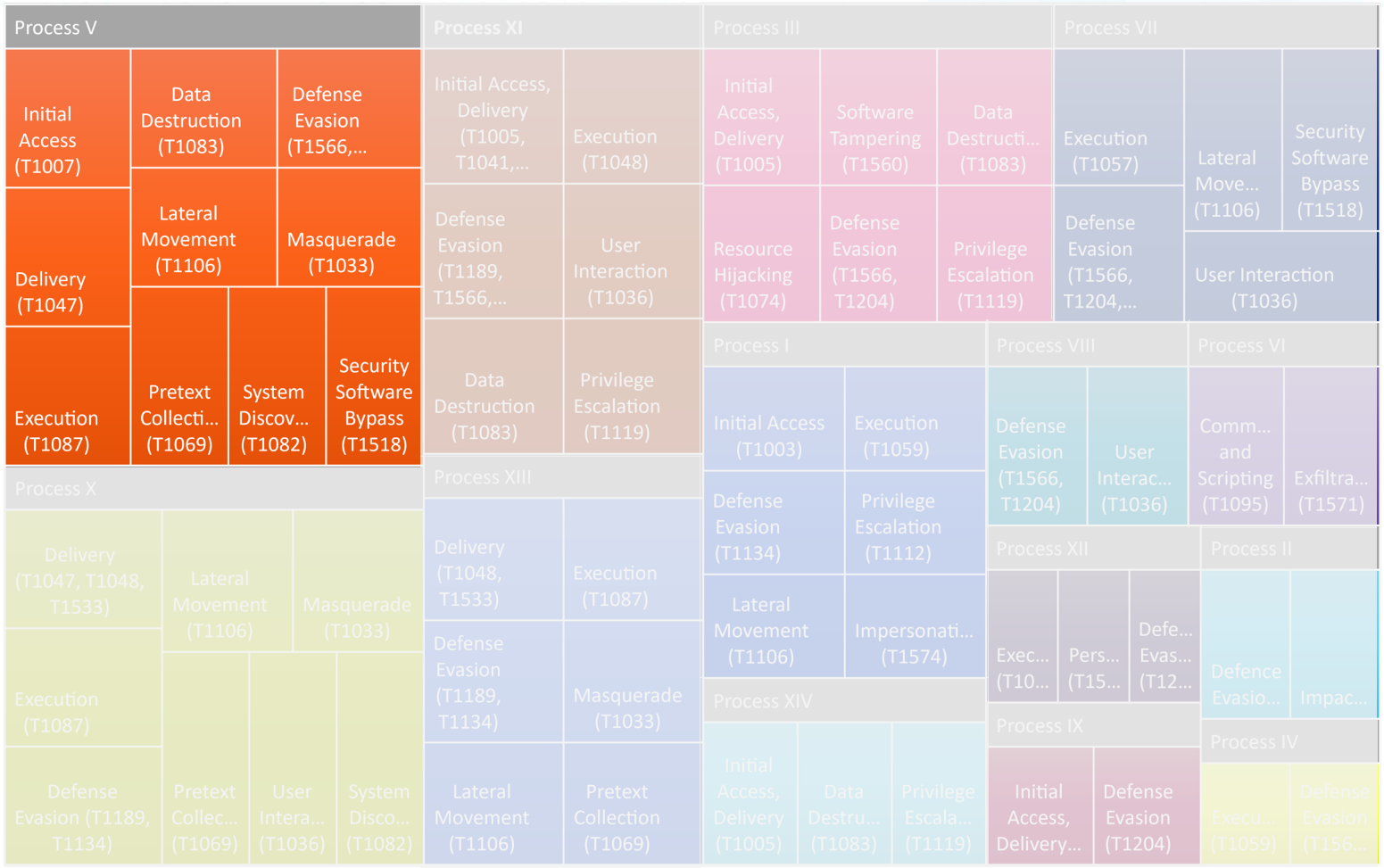


Model Score (maliciousness): 0.79279643

Step 6

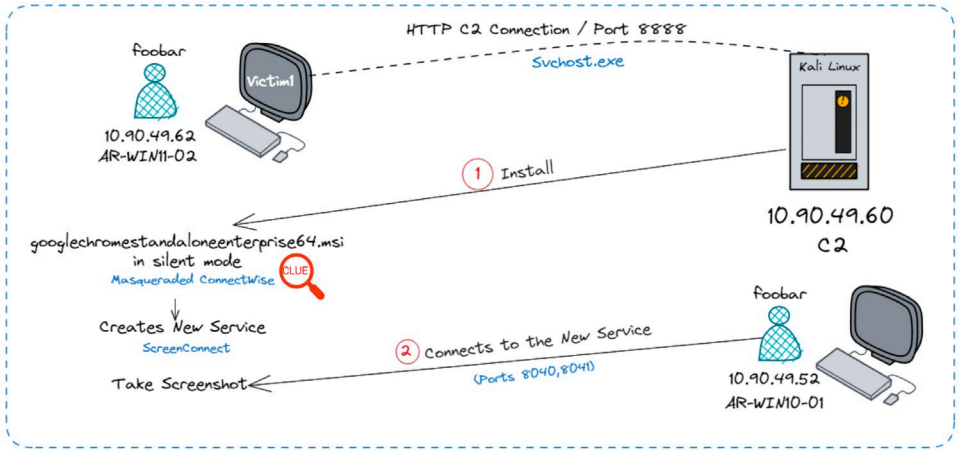


Malicious Behaviour Detections based on Mitre TTP

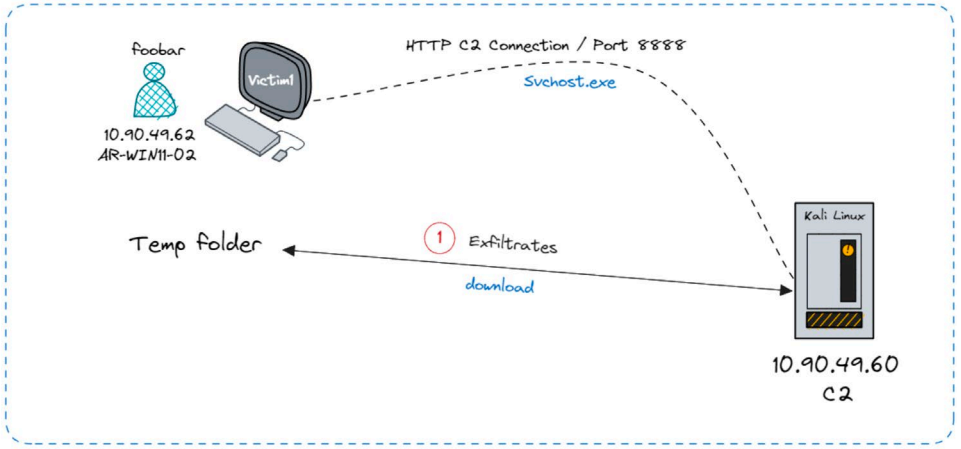


Model Score (maliciousness): 0.9131431

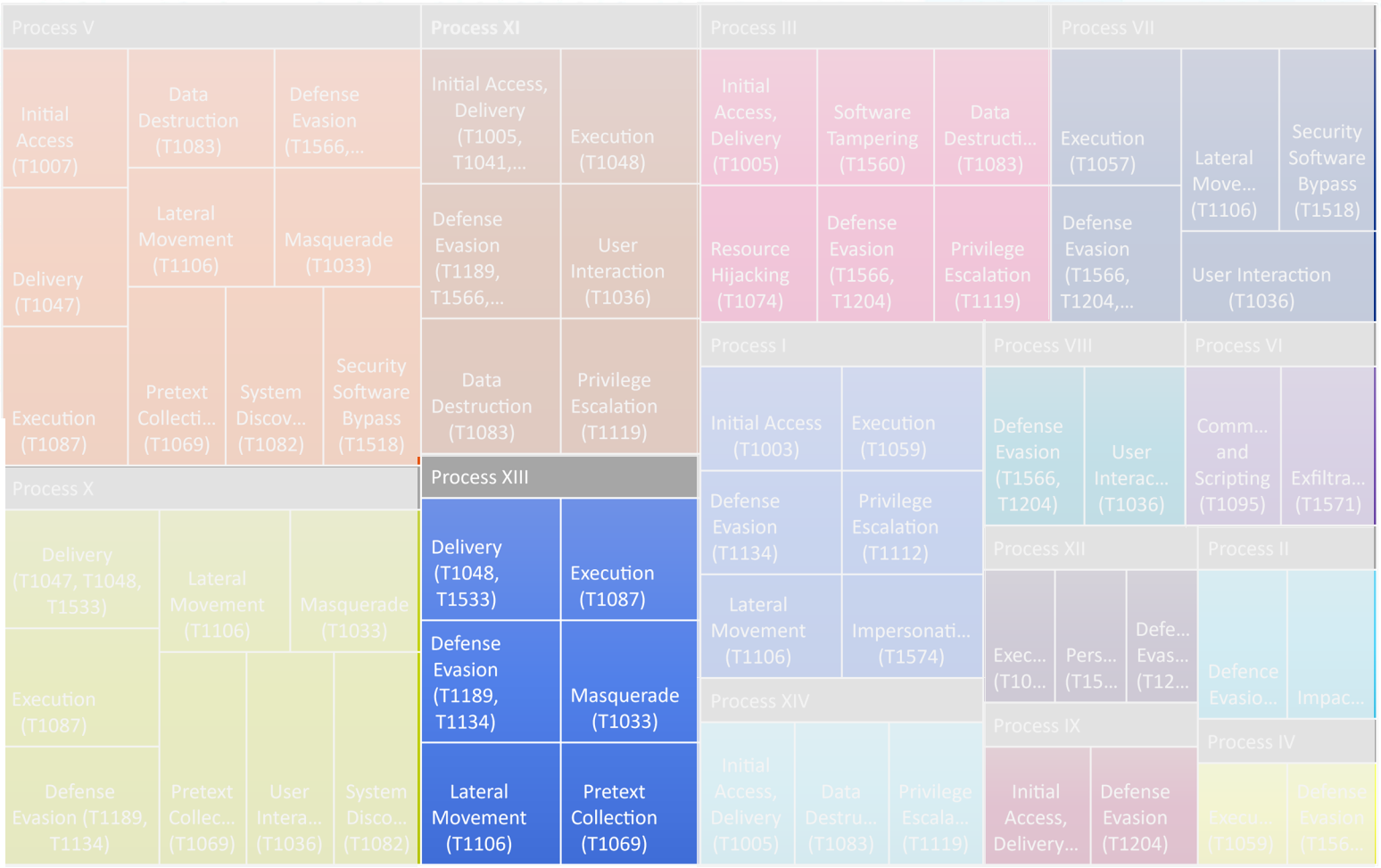
Step 7



Step 8

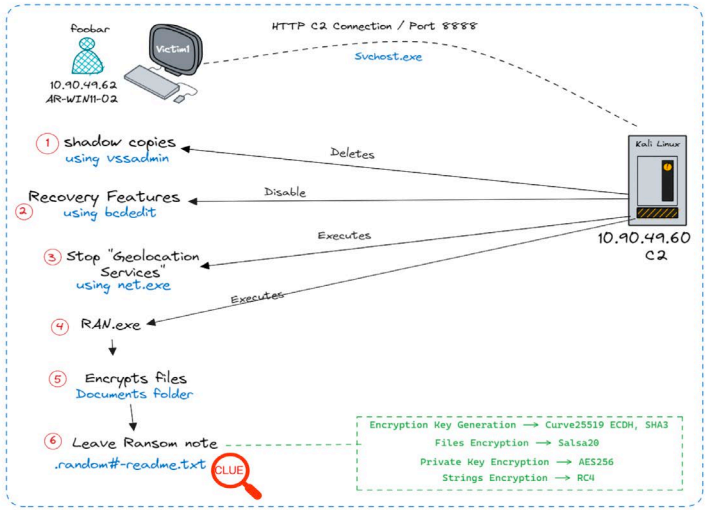


Malicious Behaviour Detections based on Mitre TTP

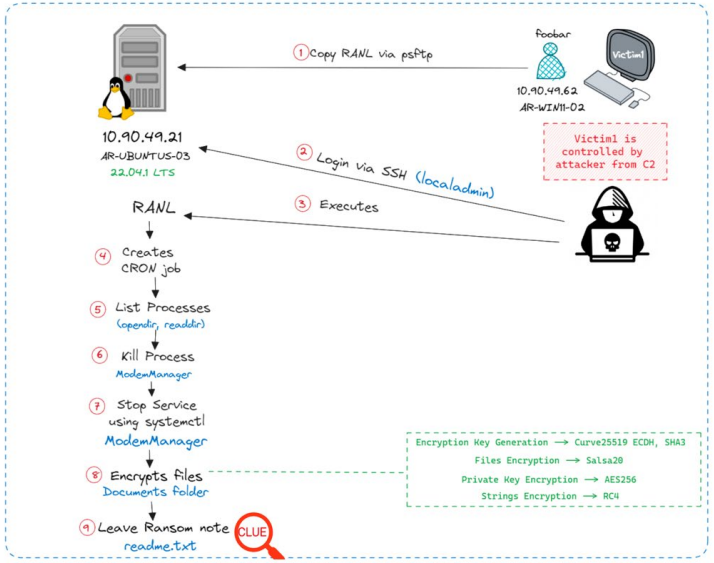


Model Score (maliciousness): 0.9131431

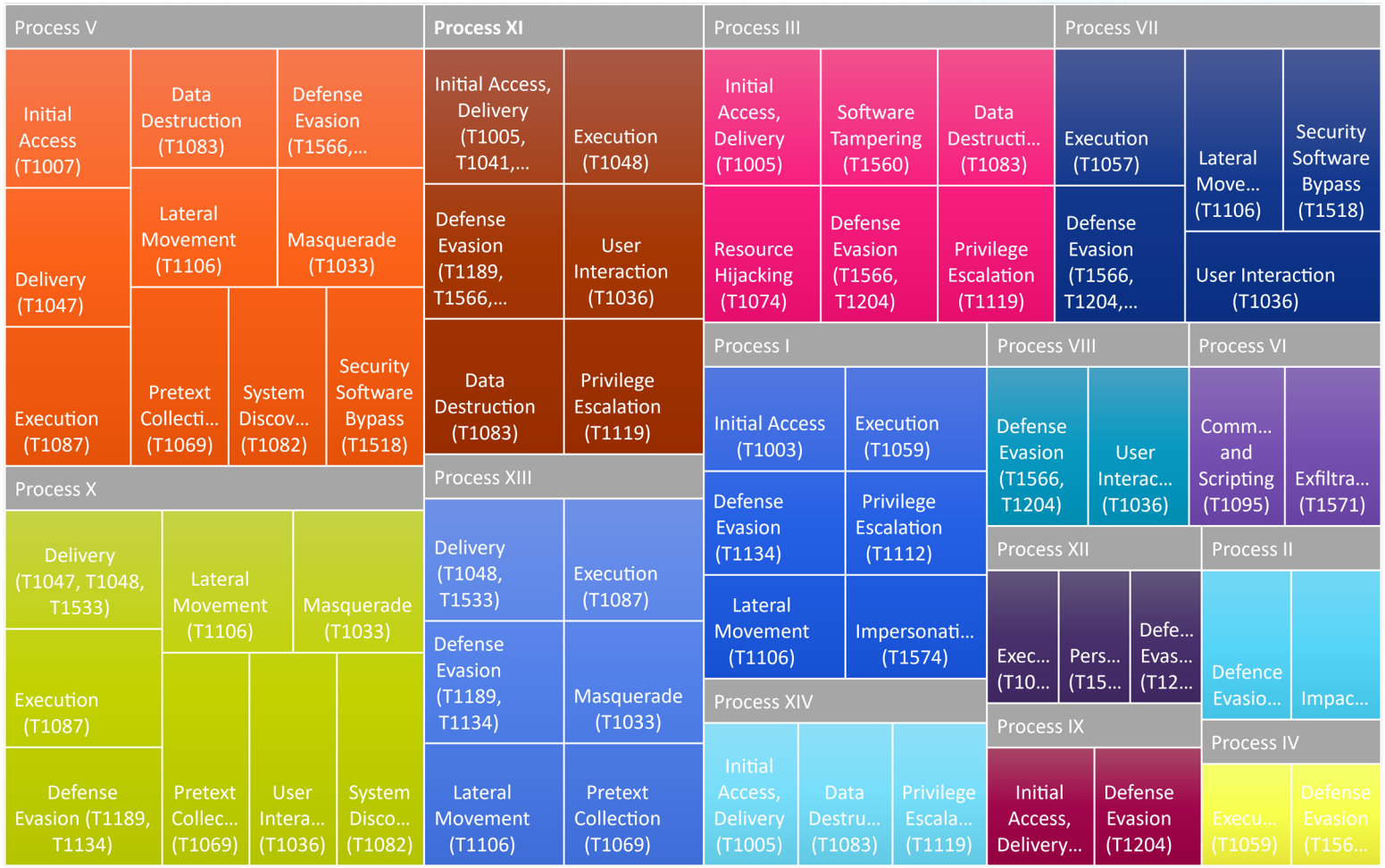
Step 9



Step 10



Malicious Behaviour Detections based on Mitre TTP



Model Score (maliciousness):0.8190064, 0.8338462

Agenda

- Surge in Novel Cyberattacks
- AI-Powered Threat Landscape
- Predictive Approaches to Defense
- Observed Outcomes
- Key Takeaways

Apply What You Have Learned Today

- Short Term you/your team should: Explore
  - Evaluate if Exploit predictability scoring Model is applicable
  - Review analysis and prioritization of exposure
  - Align your search with documented techniques ATLAS Matrix | MITRE ATLAS™
- Near-Term you/ your team should: Enable
  - Consider AI Center of Excellence in Cyber Security
  - Invest in Protect first Methodology
  - Explore open-source tools/projects https://github.com/cylance/IntroductionToMachineLearningForSecurityPros
- Mid-Term you / your team should: Evolve
  - Leverage community intelligence in threat simulation with Generative models:
  - Community links here : Educational Resources · OWASP/www-project-top-10-for-large-language-model-applications Wiki · GitHub
  - Validate detection and protection efficacy and align with emerging legislation
  - Legislation/ Responsible AI links here: The Act Texts | EU Artificial Intelligence Act & Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems (canada.ca)
  - Responsible Artificial Intelligence (Responsible AI) | Groups | LinkedIn