

RSACTMConference2024

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.

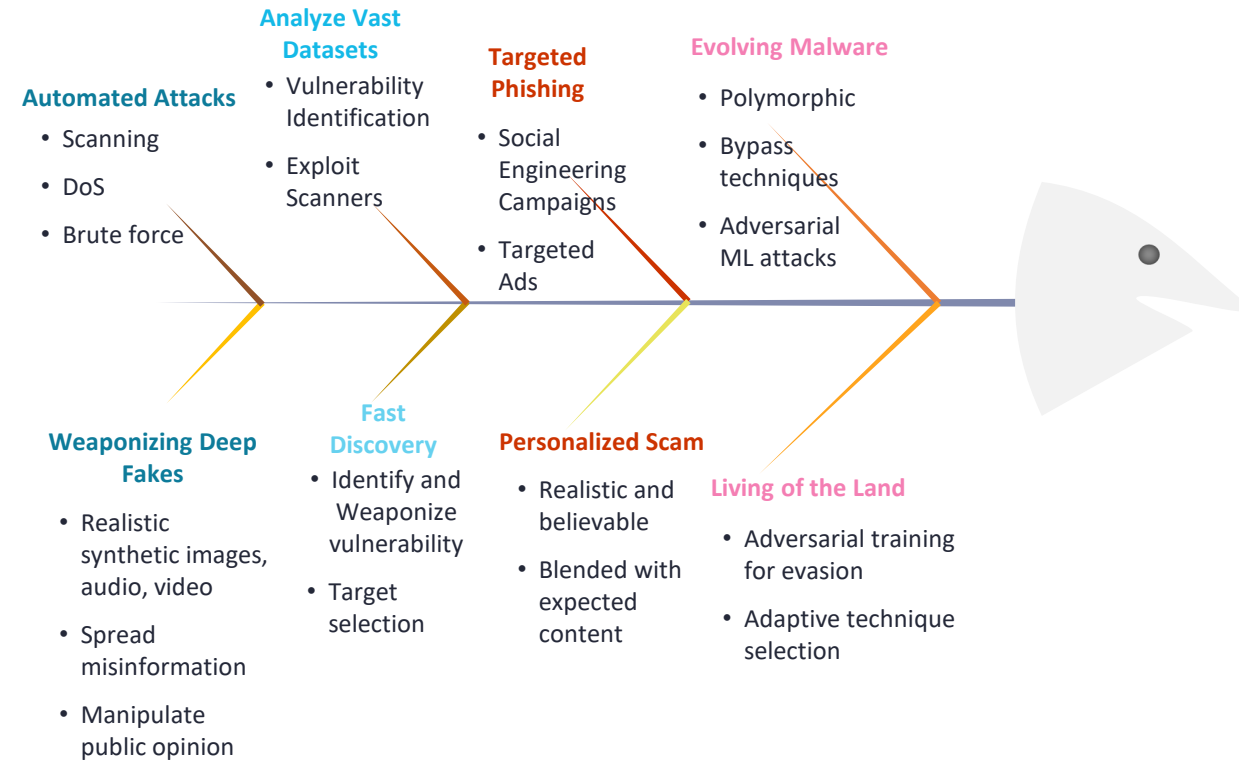
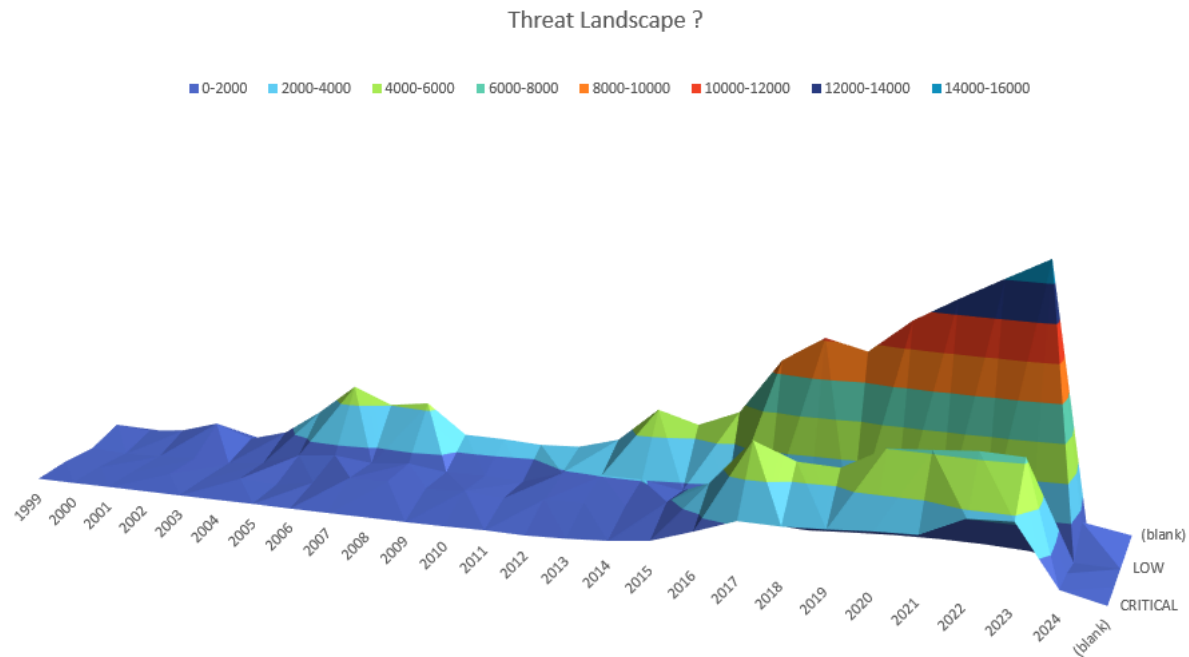
Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2024 RSA Conference LLC or its affiliates. The RSA Conference logo and other trademarks are proprietary. All rights reserved.

Agenda

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

Facing the AI-Powered Threat Landscape

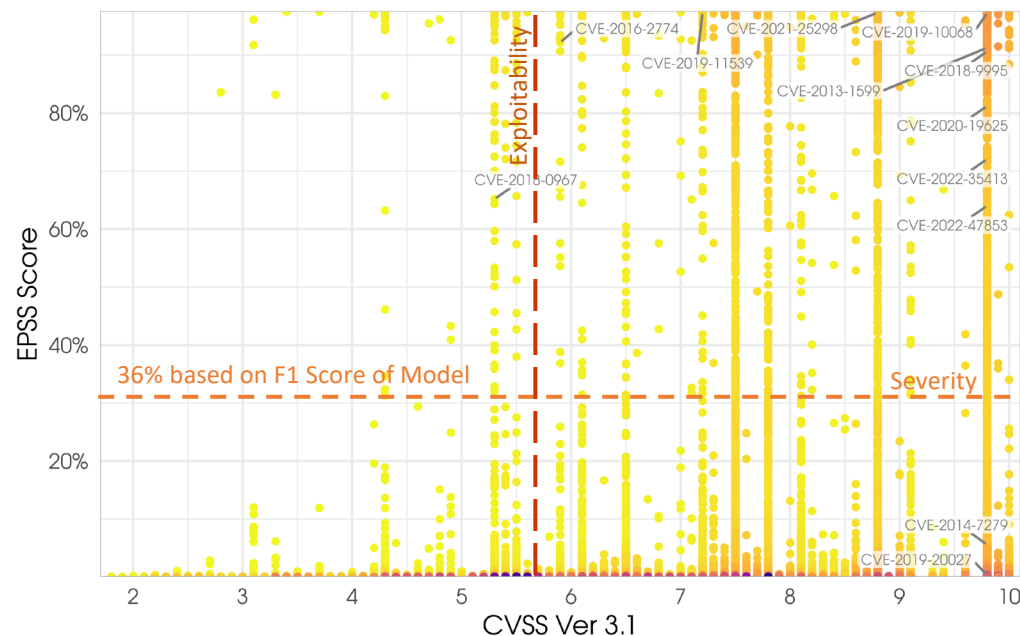


Discovery - Exploit Prediction

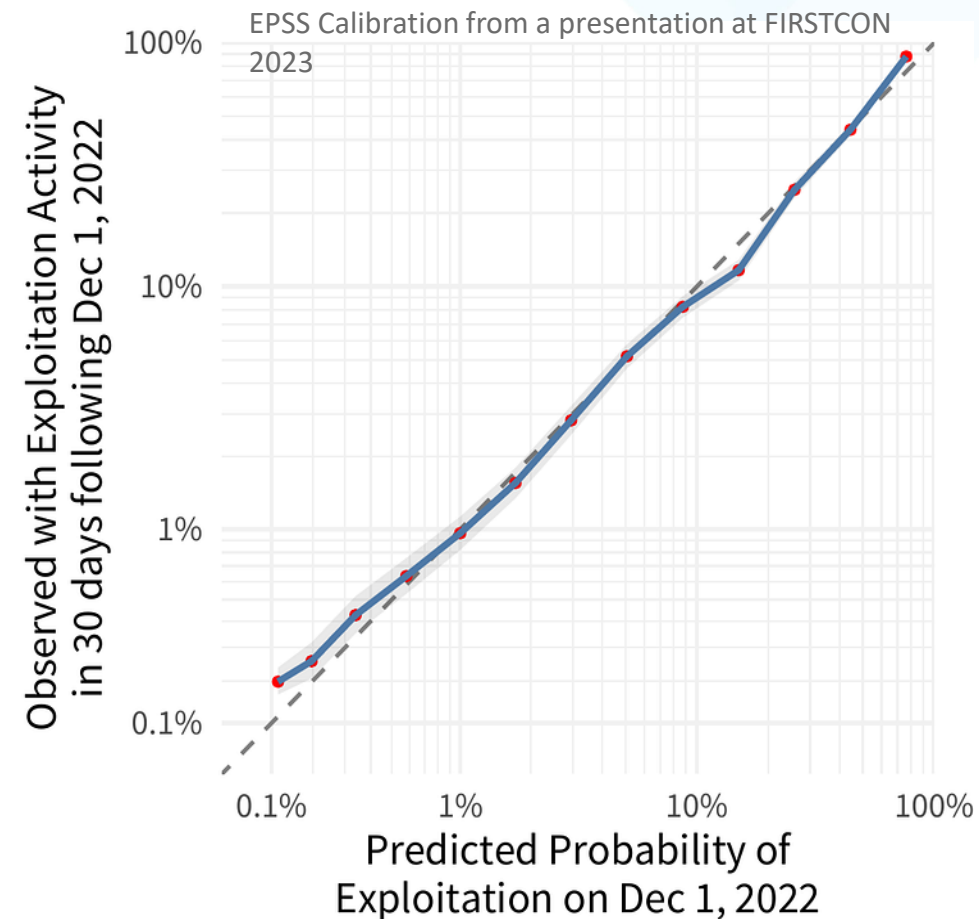
- EPSS predicts **probability** of a CVE being **exploited** in 30 days.
- Augment EPSS output with environment specific attributes to prioritize

EPSS score compared to CVSS Base Score (NVD)

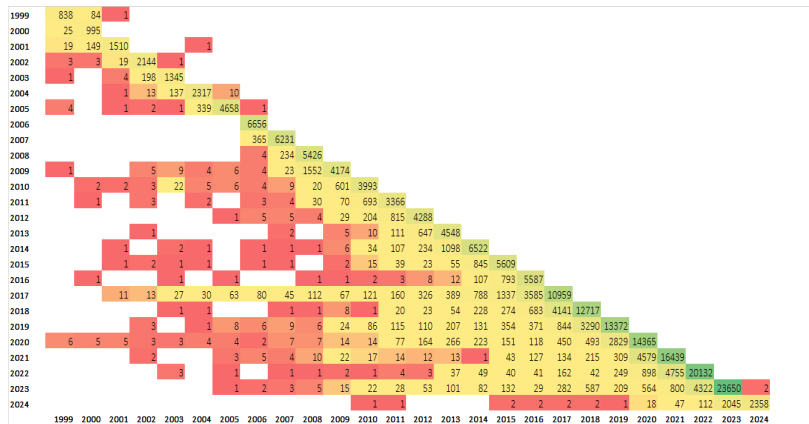
Point density is represented by color, yellow is less dense going through red to a deep purple for the most dense areas. Labelling a random sample of CVEs with higher values for reference.



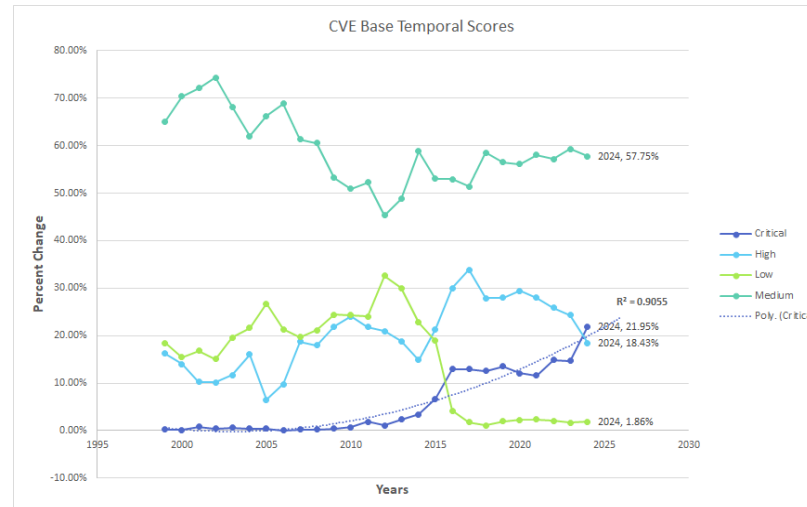
Source: https://first.org/epss/data_stats, 2024-02-28



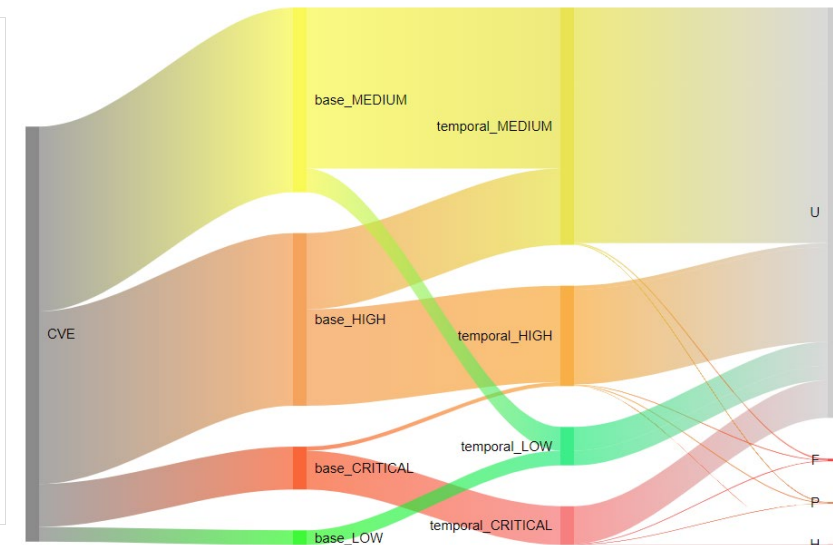
Analysis and Prioritization



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

Offensive

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

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

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

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

- No obvious course

AI & Generative AI as a Force Multiplier

Leveraging AI and Gen-AI models

Defensive	Data Analysis & Synthesis			Content Generation			Automation & Augmentation	
	Find & Fix vuln. and mis-configurations	Correlate Security events	Reveal probabilities of attack paths	Generate code	Make code more secure	Generate Blue Team Tests	Automate Incident response	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	Model Deployment & Integrity	Data Privacy
Secure development environment and OSS/Vendor Supply chain	Secure Runtime environment Apply Least privilege OWASP top 10	Differential Privacy and De-identification with Encryption

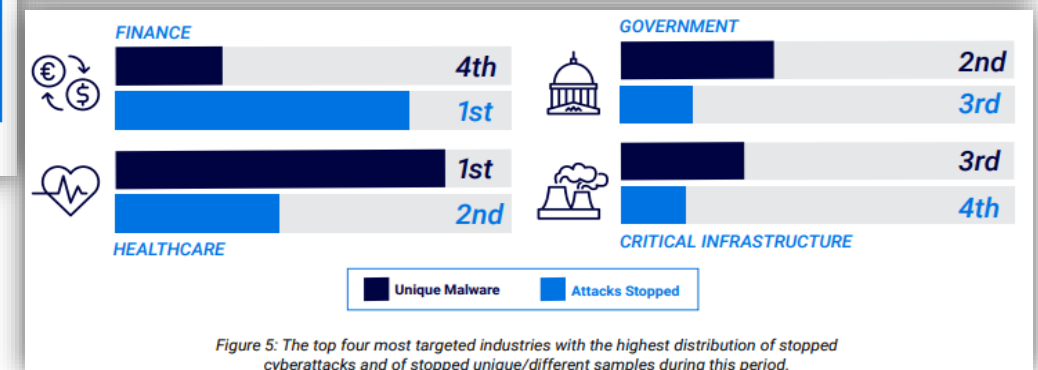
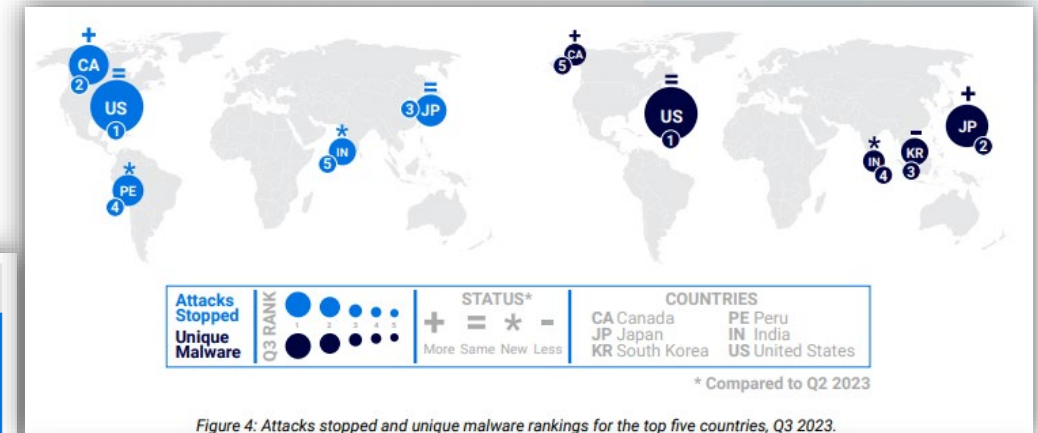
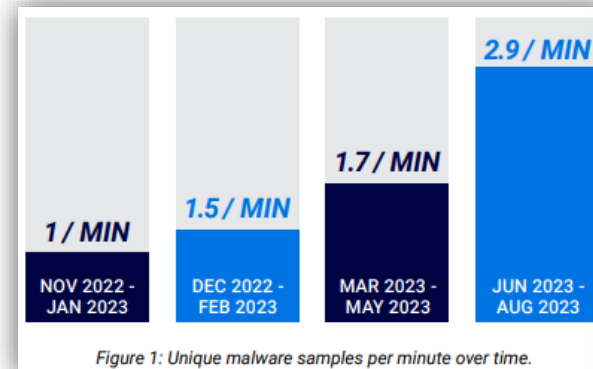
Agenda

- 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

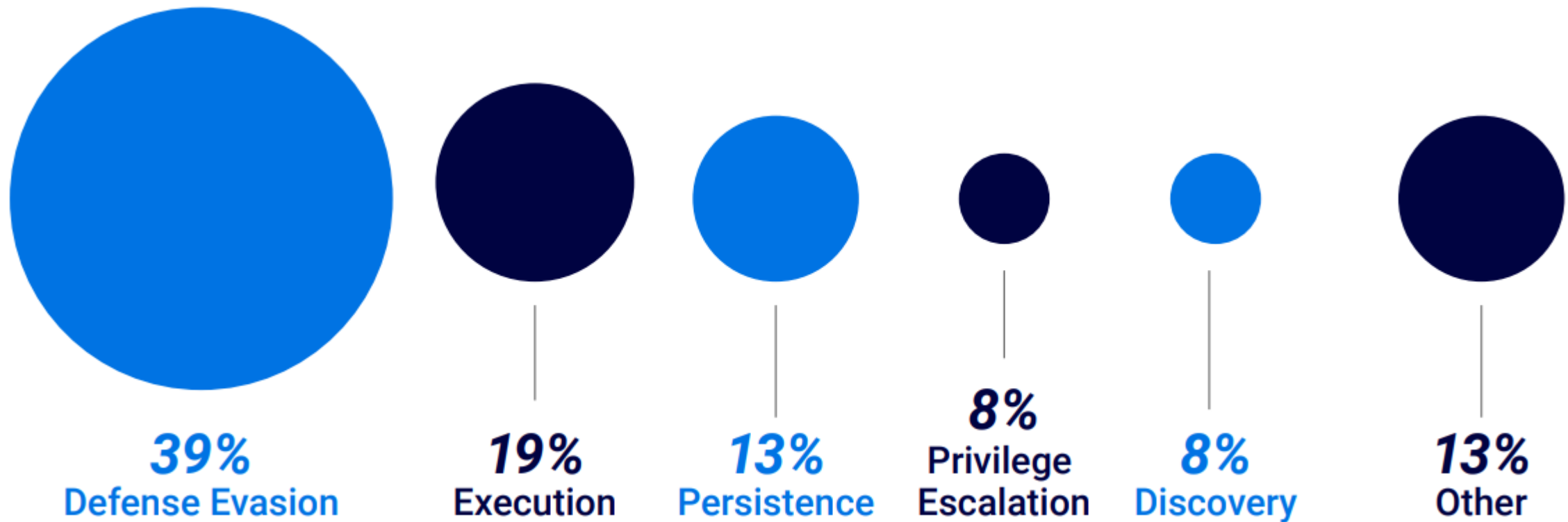
2.9 NEW UNIQUE MALWARE SAMPLES PER MINUTE.



BlackBerry CTI Report – Nov 2023

From Traditional Threats to AI-Driven Attacks

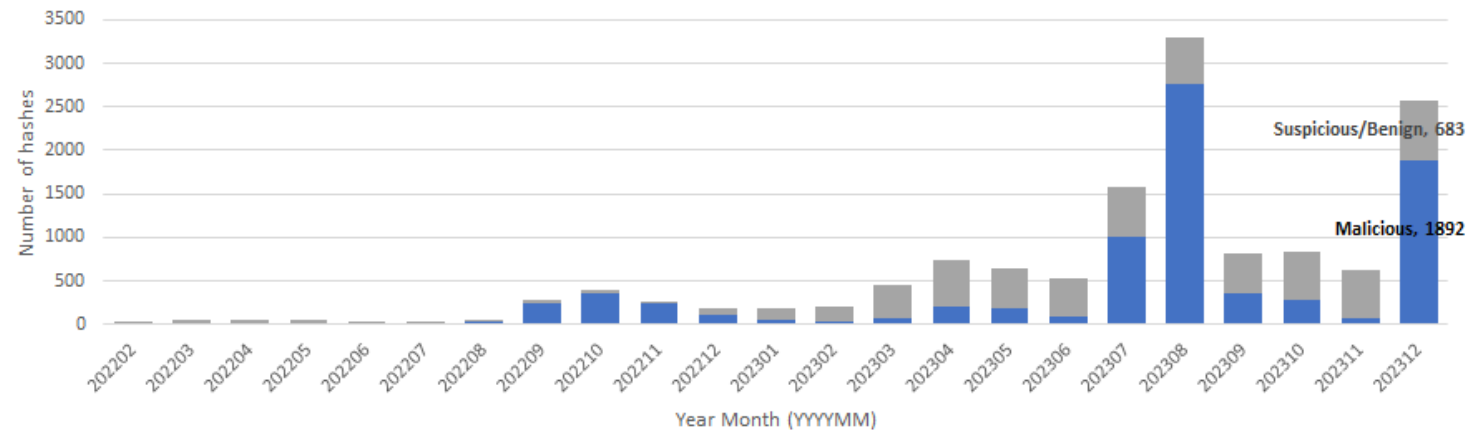
Shift from **Traditional** cyber threats to **AI-assisted** attacks, highlighting the most prevalent portion to be defense evasion.



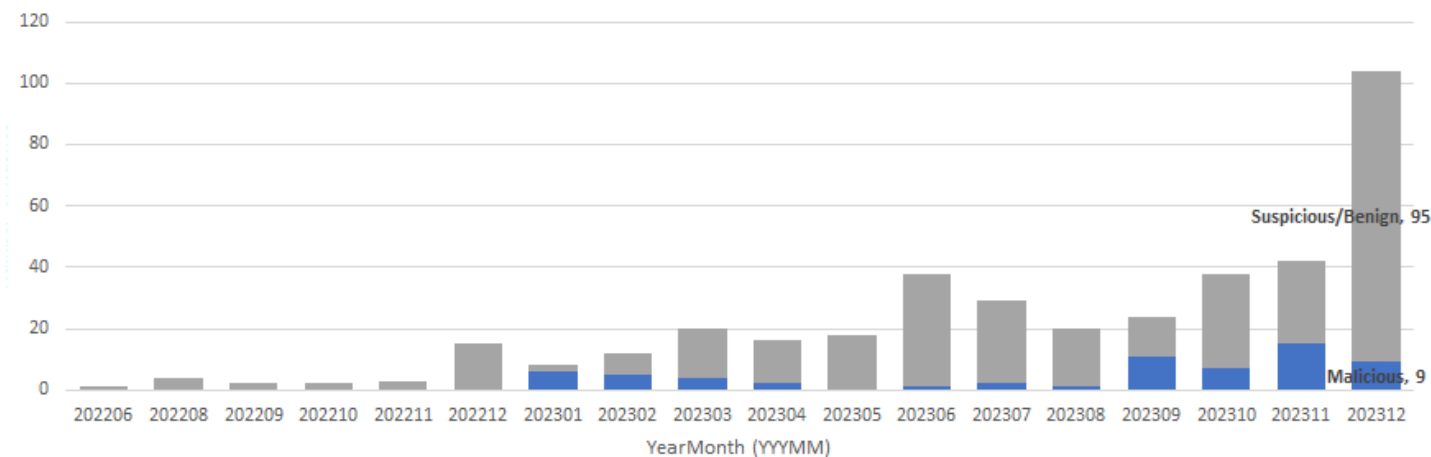
Observations from Prevalence Data

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

Malware Accessing *.api.openapi, *openai

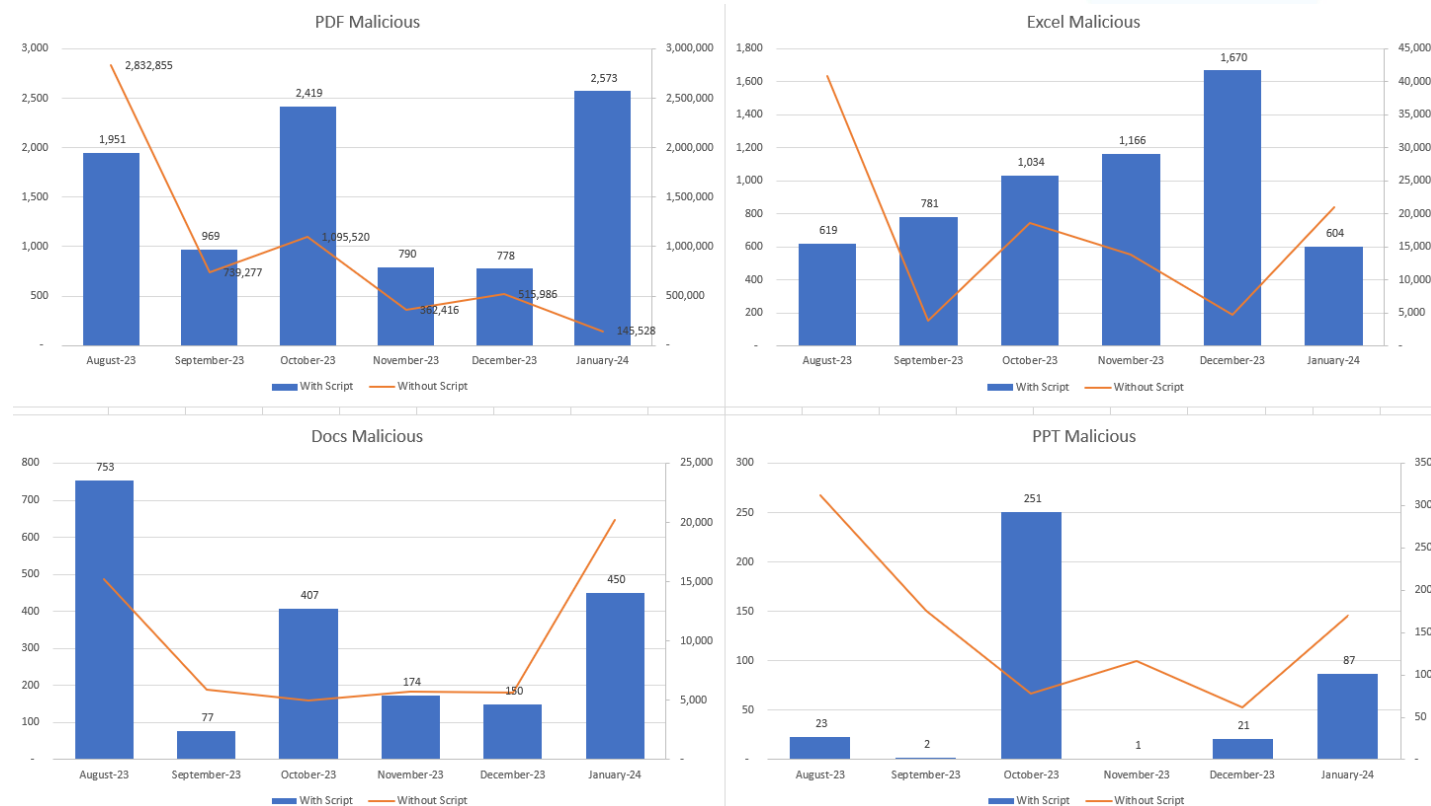


Malware Accessing *.huggingfaces.*



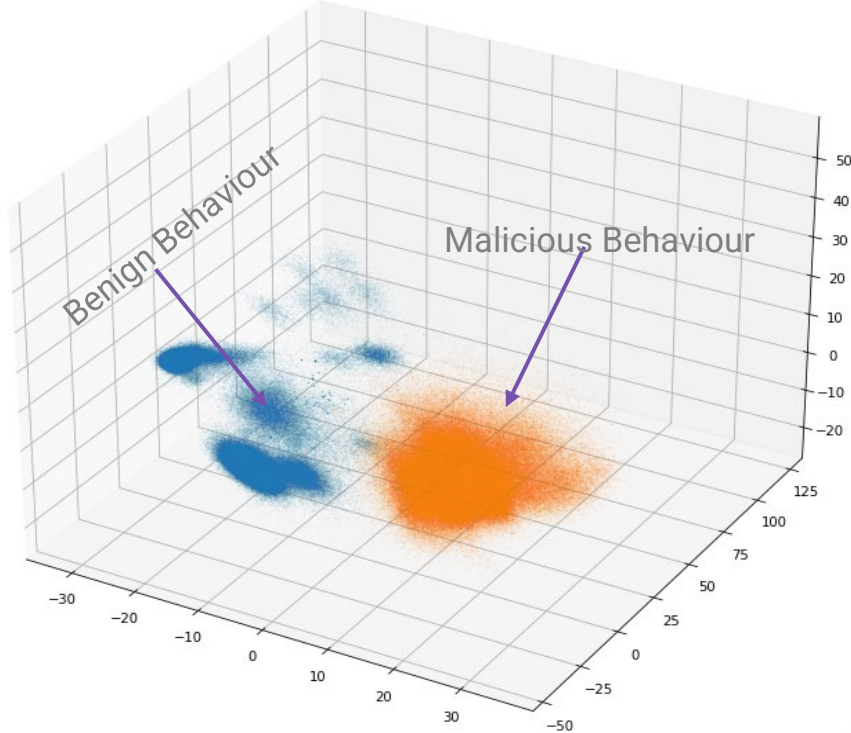
Observations from Prevalence Data

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

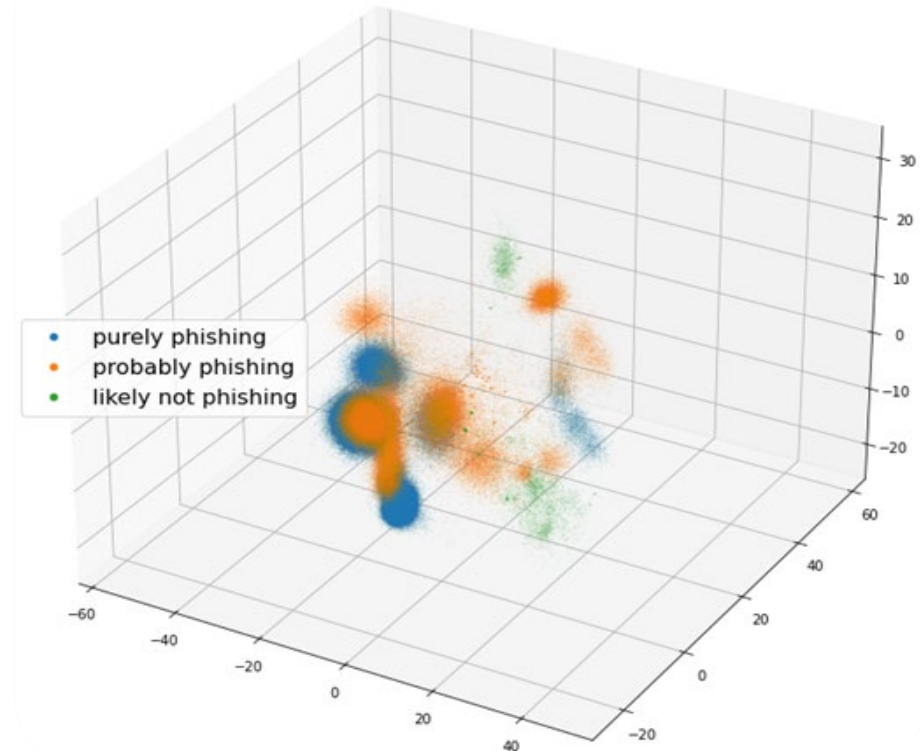


Probing-In Documents - With and Without Scripts

PDF Document Model



Analysing for Content



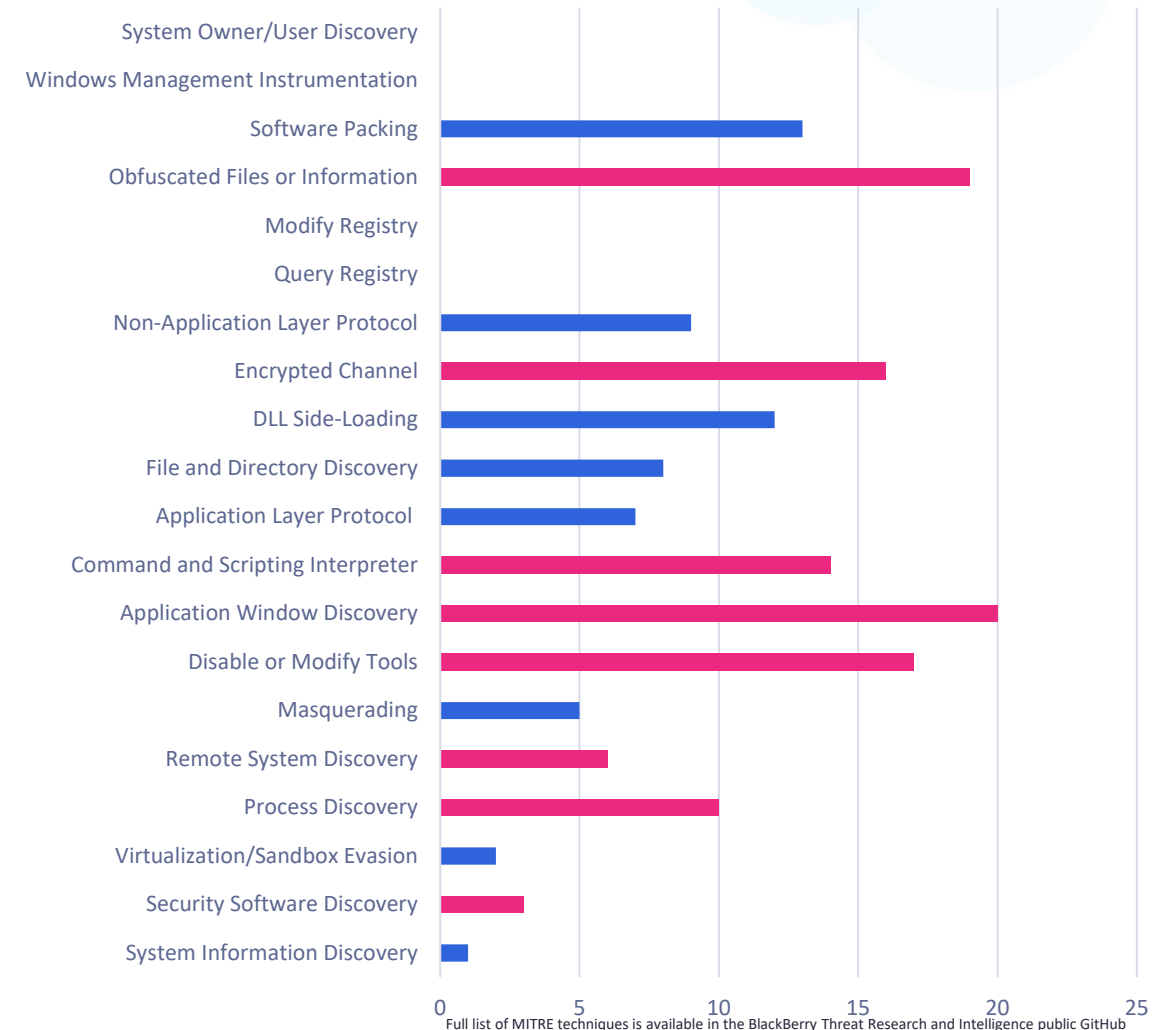
Agenda

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

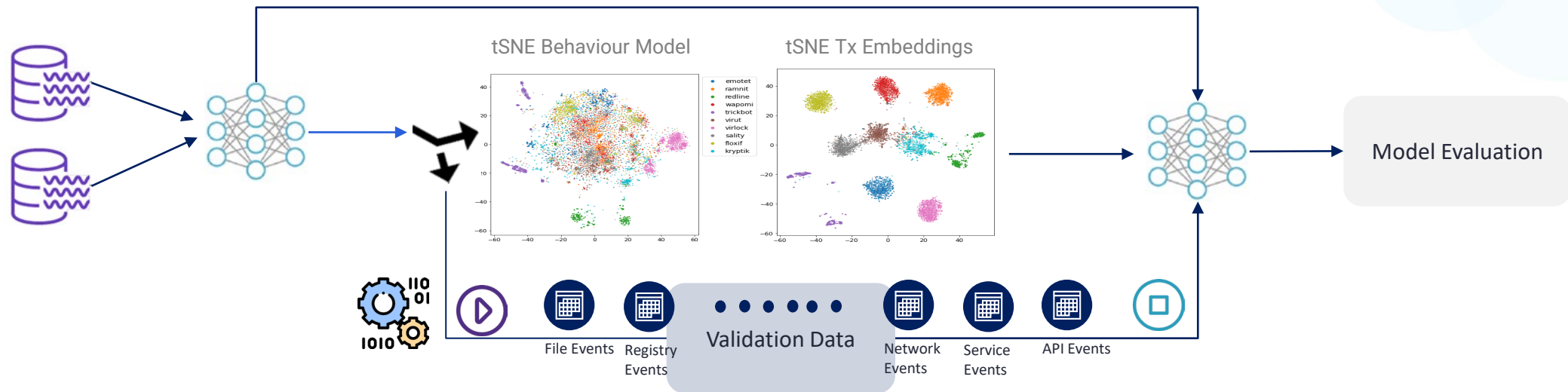
Intent and Behavior cannot be hidden

- 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.

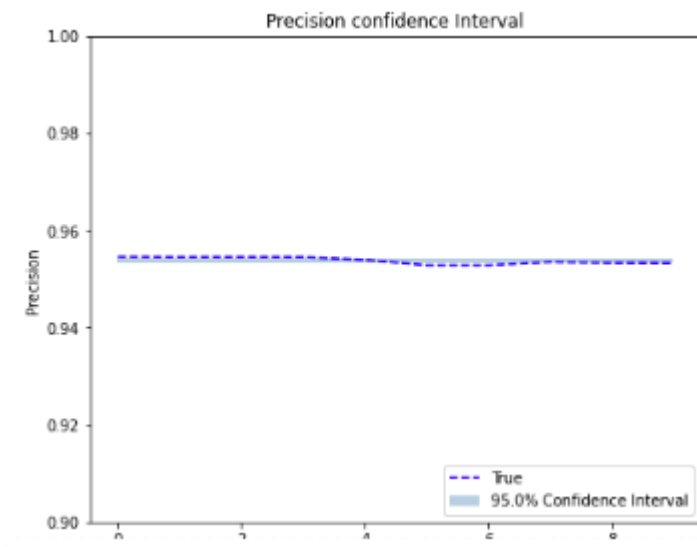
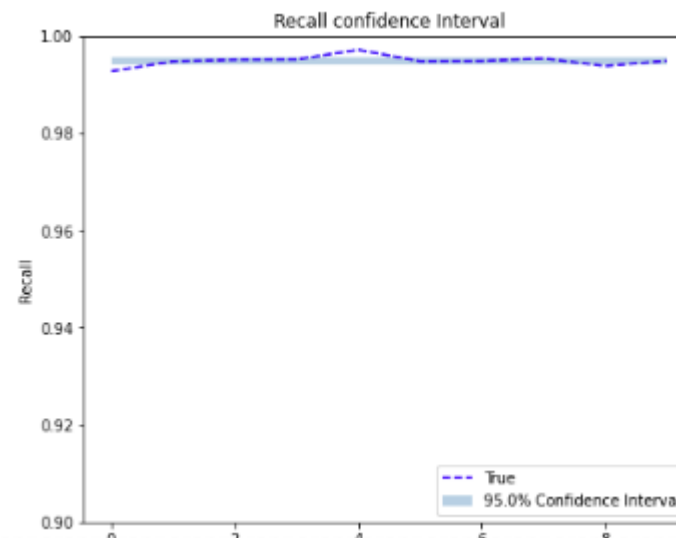
Prevalence of Change By Technique



Process Behavior prediction



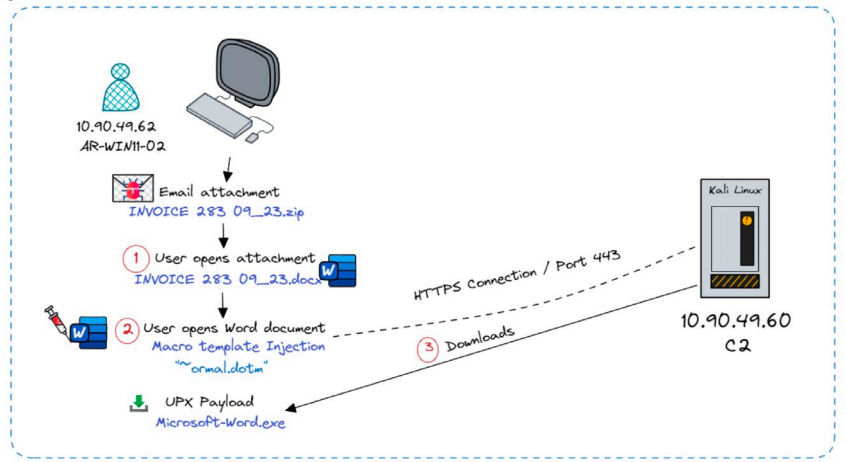
How **confident** are we with **Models** Recall And Precision ?



Agenda

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

Step 1

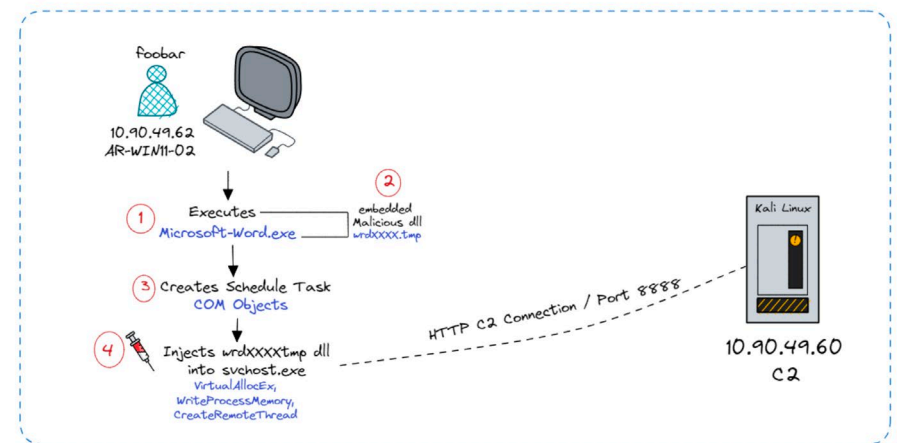


Malicious Behaviour Detections based on Mitre TTP

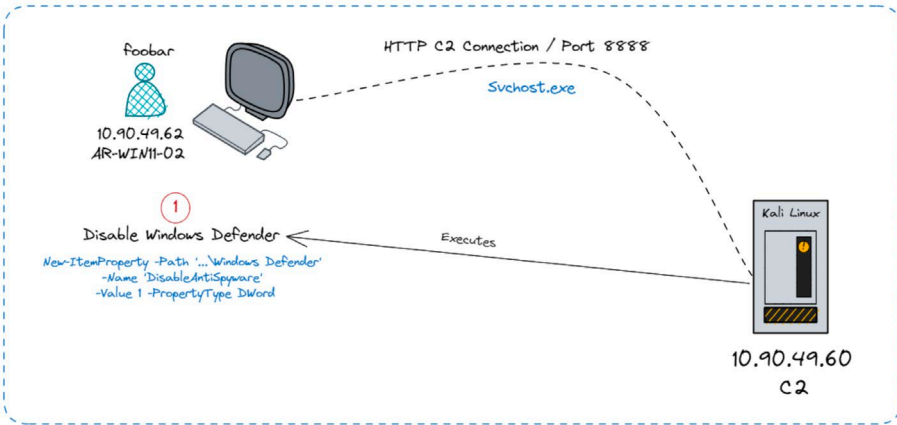
Process V				Process XI		Process III			Process VII					
Initial Access (T1007)	Data Destruction (T1083)	Defense Evasion (T1566,...)		Initial Access, Delivery (T1005, T1041,...)	Execution (T1048)	Initial Access, Delivery (T1005)	Software Tampering (T1560)	Data Destructi... (T1083)	Execution (T1057)	Lateral Move... (T1106)	Security Software Bypass (T1518)			
	Lateral Movement (T1106)	Masquerade (T1033)												
Delivery (T1047)	Lateral Movement (T1106)	Masquerade (T1033)		Defense Evasion (T1189, T1566,...)	User Interaction (T1036)	Resource Hijacking (T1074)	Defense Evasion (T1566, T1204)	Privilege Escalation (T1119)	Defense Evasion (T1566, T1204,...)	User Interaction (T1036)				
Execution (T1087)	Pretext Collecti... (T1069)	System Discov... (T1082)	Security Software Bypass (T1518)	Data Destruction (T1083)	Privilege Escalation (T1119)	Process I		Process VIII		Process VI				
						Initial Access (T1003)	Execution (T1059)	Defense Evasion (T1566, T1204)	User Interac... (T1036)	Comm... and Scripting (T1095)	Exfiltra... (T1571)			
Process X				Process XIII		Defense Evasion (T1134)	Privilege Escalation (T1112)	Process XII		Process II				
Delivery (T1047, T1048, T1533)		Lateral Movement (T1106)		Masquerade (T1033)		Delivery (T1048, T1533)	Execution (T1087)	Process XIV		Exec... (T10...)	Pers... (T15...)	Defe... (T12...)	Defence Evasio...	Impac...
Execution (T1087)						Defense Evasion (T1189, T1134)	Masquerade (T1033)	Process XIV		Process IX		Process IV		
								Initial Access, Delivery (T1005)	Data Destru... (T1083)	Privilege Escala... (T1119)	Initial Access, Delivery...	Defense Evasion (T1204)	Execu... (T1059)	Defense Evasion (T156...

Model Score (maliciousness): 0.9883317

Step 2



Step 3

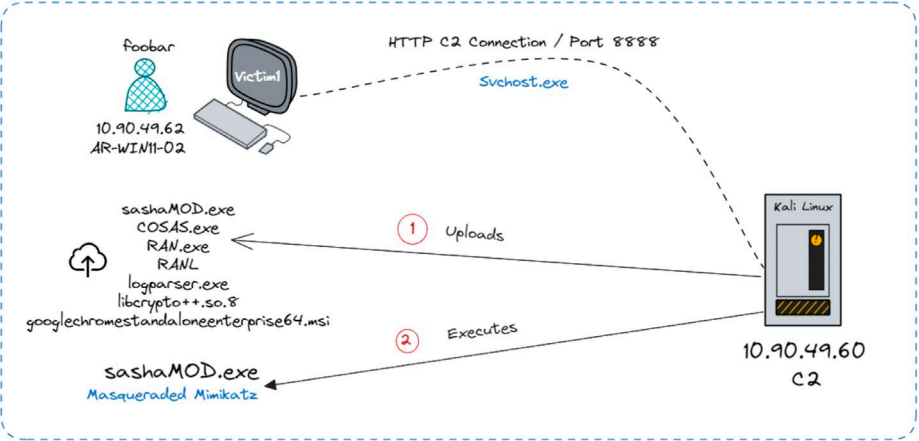


Malicious Behaviour Detections based on Mitre TTP

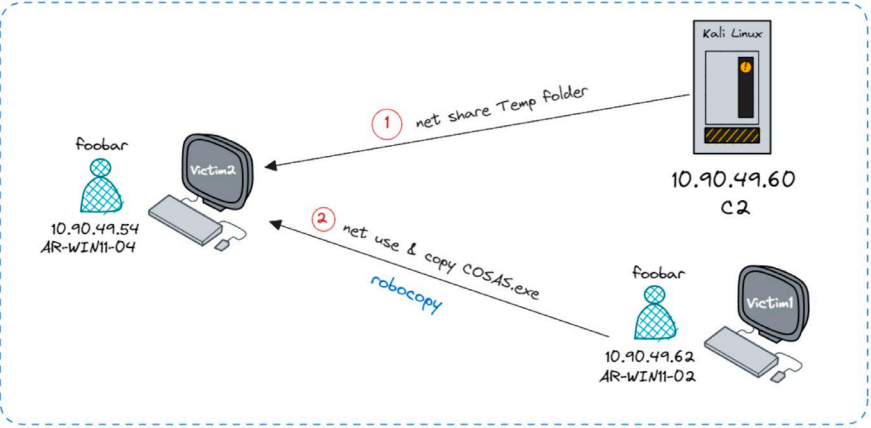
Process V				Process XI		Process III			Process VII			
Initial Access (T1007)	Data Destruction (T1083)	Defense Evasion (T1566,...)		Initial Access, Delivery (T1005, T1041,...)	Execution (T1048)	Initial Access, Delivery (T1005)	Software Tampering (T1560)	Data Destructi... (T1083)	Execution (T1057)	Lateral Move... (T1106)	Security Software Bypass (T1518)	
	Delivery (T1047)	Lateral Movement (T1106)	Masquerade (T1033)		Defense Evasion (T1189, T1566,...)	User Interaction (T1036)	Resource Hijacking (T1074)	Defense Evasion (T1566, T1204)				Privilege Escalation (T1119)
Execution (T1087)		Pretext Collecti... (T1069)	System Discov... (T1082)	Security Software Bypass (T1518)	Data Destruction (T1083)	Privilege Escalation (T1119)	Process I		Process VIII		Process VI	
	Initial Access (T1003)						Execution (T1059)	Defense Evasion (T1566, T1204)	User Interac... (T1036)	Comm... and Scripting (T1095)	Exfiltra... (T1571)	
Process X				Process XIII		Defense Evasion (T1134)	Privilege Escalation (T1112)	Process XII	Process II			
Delivery (T1047, T1048, T1533)	Lateral Movement (T1106)		Masquerade (T1033)	Delivery (T1048, T1533)	Execution (T1087)							
Execution (T1087)				Defense Evasion (T1189, T1134)	Masquerade (T1033)							
	Process XIV		Process IX			Process IV						
Defense Evasion (T1189, T1134)	Pretext Collec... (T1069)	User Intera... (T1036)	System Disco... (T1082)	Lateral Movement (T1106)	Pretext Collection (T1069)	Initial Access, Delivery (T1005)	Data Destru... (T1083)	Privilege Escala... (T1119)	Initial Access, Delivery...	Defense Evasion (T1204)	Execu... (T1059)	Defense Evasion (T156...

Model Score (maliciousness): 0.97976273, 0.96840143, 0.82308644

Step 4



Step 5

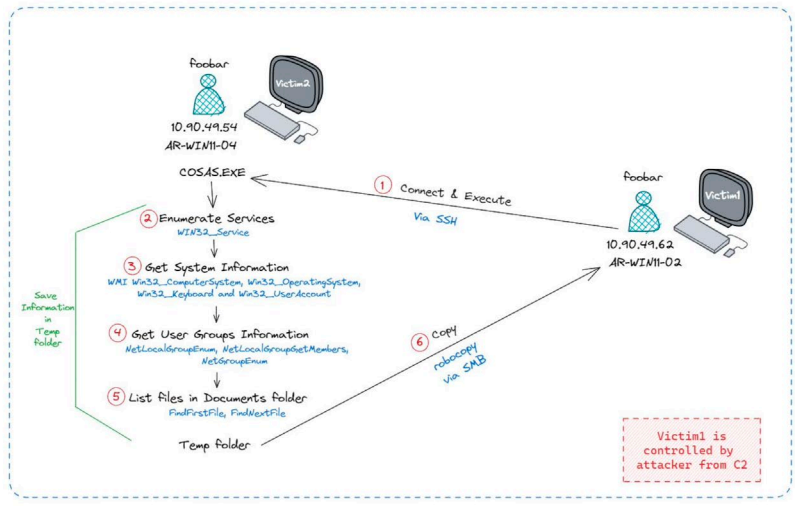


Malicious Behaviour Detections based on Mitre TTP

Process V				Process XI		Process III			Process VII		
Initial Access (T1007)	Data Destruction (T1083)	Defense Evasion (T1566,...)		Initial Access, Delivery (T1005, T1041,...)	Execution (T1048)	Initial Access, Delivery (T1005)	Software Tampering (T1560)	Data Destructi... (T1083)	Execution (T1057)	Lateral Move... (T1106)	Security Software Bypass (T1518)
Delivery (T1047)	Lateral Movement (T1106)		Masquerade (T1033)	Defense Evasion (T1189, T1566,...)	User Interaction (T1036)	Resource Hijacking (T1074)	Defense Evasion (T1566, T1204)	Privilege Escalation (T1119)	Defense Evasion (T1566, T1204,...)	User Interaction (T1036)	
Execution (T1087)	Pretext Collecti... (T1069)	System Discov... (T1082)	Security Software Bypass (T1518)	Data Destruction (T1083)	Privilege Escalation (T1119)	Process I		Process VIII		Process VI	
Process X				Process XIII		Initial Access (T1003)	Execution (T1059)	Defense Evasion (T1566, T1204)	User Interac... (T1036)	Comm... and Scripting (T1095)	Exfiltra... (T1571)
Delivery (T1047, T1048, T1533)	Lateral Movement (T1106)		Masquerade (T1033)	Delivery (T1048, T1533)	Execution (T1087)	Defense Evasion (T1134)	Privilege Escalation (T1112)	Process XII		Process II	
Execution (T1087)				Defense Evasion (T1189, T1134)	Masquerade (T1033)	Lateral Movement (T1106)	Impersonati... (T1574)	Exec... (T10...)	Pers... (T15...)	Defe... (T12...)	Defence Evasio... Impac...
Defense Evasion (T1189, T1134)	Pretext Collec... (T1069)	User Intera... (T1036)	System Disco... (T1082)	Lateral Movement (T1106)	Pretext Collection (T1069)	Process XIV		Process IX		Process IV	
						Initial Access, Delivery (T1005)	Data Destru... (T1083)	Privilege Escala... (T1119)	Initial Access, Delivery... (T1005)	Defense Evasion (T1204)	Execu... (T1059) Defense Evasion (T156...)

Model Score (maliciousness): 0.79279643

Step 6

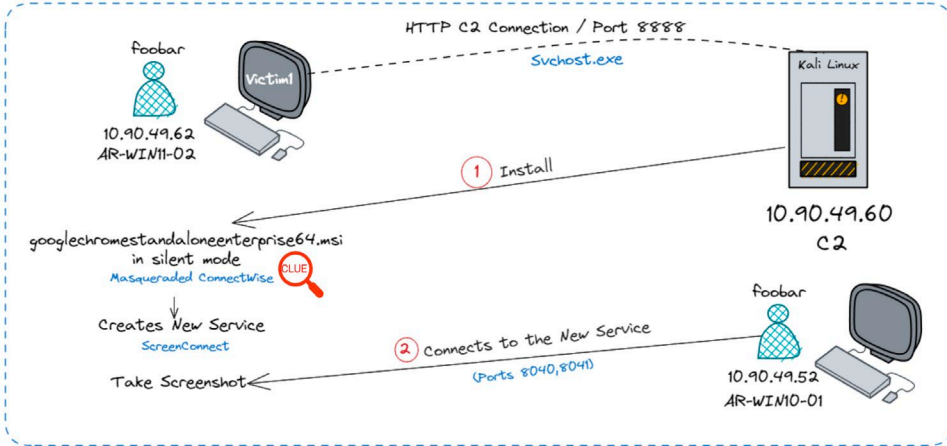


Malicious Behaviour Detections based on Mitre TTP

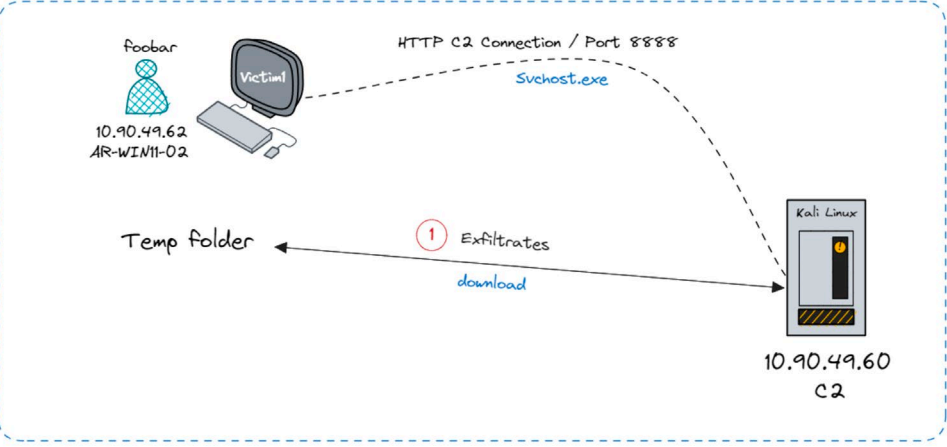
Process V				Process XI		Process III			Process VII				
Initial Access (T1007)	Data Destruction (T1083)	Defense Evasion (T1566,...)		Initial Access, Delivery (T1005, T1041,...)	Execution (T1048)	Initial Access, Delivery (T1005)	Software Tampering (T1560)	Data Destructi... (T1083)	Execution (T1057)	Lateral Move... (T1106)	Security Software Bypass (T1518)		
	Delivery (T1047)	Lateral Movement (T1106)	Masquerade (T1033)	Defense Evasion (T1189, T1566,...)	User Interaction (T1036)	Resource Hijacking (T1074)	Defense Evasion (T1566, T1204)	Privilege Escalation (T1119)				Defense Evasion (T1566, T1204,...)	User Interaction (T1036)
Execution (T1087)		Pretext Collecti... (T1069)	System Discov... (T1082)	Security Software Bypass (T1518)	Data Destruction (T1083)	Privilege Escalation (T1119)	Process I		Process VIII		Process VI		
	Initial Access (T1003)						Execution (T1059)	Defense Evasion (T1566, T1204)	User Interac... (T1036)	Comm... and Scripting (T1095)	Exfiltra... (T1571)		
Process X				Process XIII		Defense Evasion (T1134)	Privilege Escalation (T1112)	Process XII		Process II			
Delivery (T1047, T1048, T1533)		Lateral Movement (T1106)		Masquerade (T1033)		Delivery (T1048, T1533)	Execution (T1087)	Process XIV		Defence Evasio... (T1059)	Impac...		
Execution (T1087)				Defense Evasion (T1189, T1134)		Masquerade (T1033)		Process IX				Process IV	
Defense Evasion (T1189, T1134)		Pretext Collec... (T1069)	User Intera... (T1036)	System Disco... (T1082)	Lateral Movement (T1106)	Pretext Collection (T1069)	Initial Access, Delivery (T1005)	Data Destru... (T1083)	Privilege Escala... (T1119)	Initial Access, Delivery...	Defense Evasion (T1204)	Execu... (T1059)	Defense Evasion (T156...

Model Score (maliciousness): 0.9131431

Step 7



Step 8

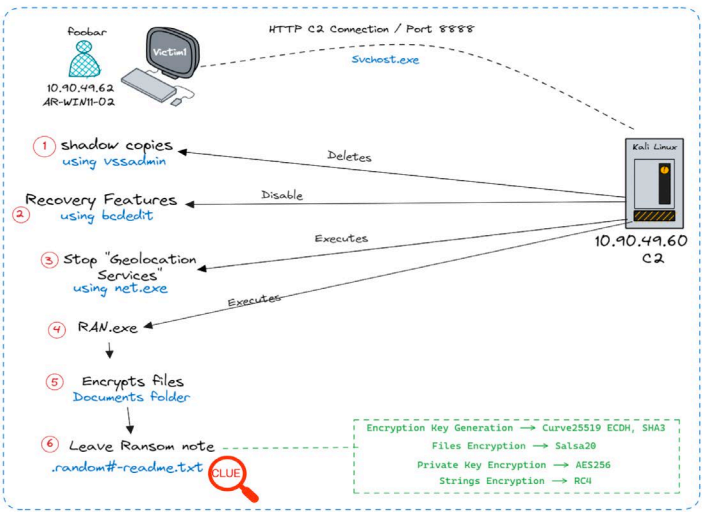


Malicious Behaviour Detections based on Mitre TTP

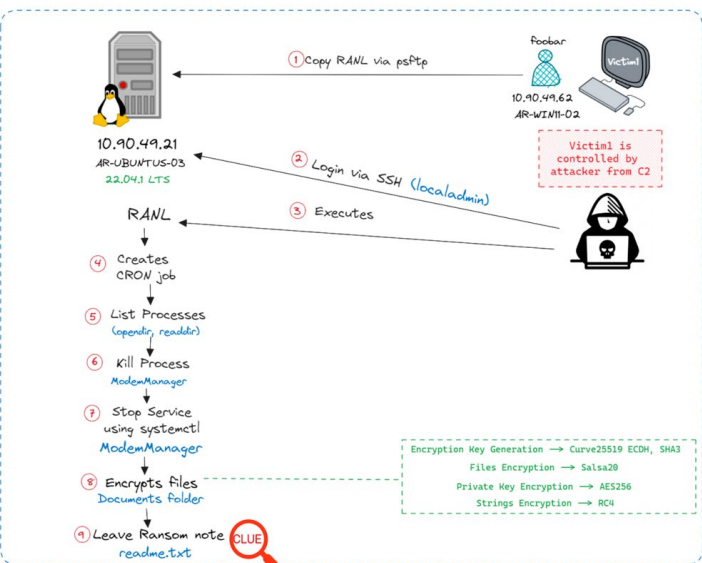
Process V				Process XI		Process III			Process VII				
Initial Access (T1007)	Data Destruction (T1083)	Defense Evasion (T1566,...)		Initial Access, Delivery (T1005, T1041,...)	Execution (T1048)	Initial Access, Delivery (T1005)	Software Tampering (T1560)	Data Destructi... (T1083)	Execution (T1057)	Lateral Move... (T1106)	Security Software Bypass (T1518)		
	Delivery (T1047)	Lateral Movement (T1106)	Masquerade (T1033)		Defense Evasion (T1189, T1566,...)	User Interaction (T1036)	Resource Hijacking (T1074)	Defense Evasion (T1566, T1204)				Privilege Escalation (T1119)	Defense Evasion (T1566, T1204,...)
Execution (T1087)		Pretext Collecti... (T1069)	System Discov... (T1082)	Security Software Bypass (T1518)	Data Destruction (T1083)	Privilege Escalation (T1119)	Process I		Process VIII		Process VI		
	Initial Access (T1003)						Execution (T1059)	Defense Evasion (T1566, T1204)	User Interac... (T1036)	Comm... and Scripting (T1095)	Exfiltra... (T1571)		
Process X				Process XIII		Defense Evasion (T1134)	Privilege Escalation (T1112)	Process XII		Process II			
Delivery (T1047, T1048, T1533)	Lateral Movement (T1106)		Masquerade (T1033)		Delivery (T1048, T1533)	Execution (T1087)	Lateral Movement (T1106)	Impersonati... (T1574)	Exec... (T10...	Pers... (T15...	Defe... (T12...	Defence Evasio...	Impac...
Execution (T1087)					Defense Evasion (T1189, T1134)	Masquerade (T1033)	Process XIV		Process IX		Process IV		
Defense Evasion (T1189, T1134)	Pretext Collec... (T1069)	User Intera... (T1036)	System Disco... (T1082)	Lateral Movement (T1106)	Pretext Collection (T1069)	Initial Access, Delivery (T1005)	Data Destru... (T1083)	Privilege Escala... (T1119)	Initial Access, Delivery...	Defense Evasion (T1204)	Execu... (T1059)	Defense Evasion (T156...	

Model Score (maliciousness): 0.9131431

Step 9



Step 10



Malicious Behaviour Detections based on Mitre TTP

Process V				Process XI		Process III			Process VII				
Initial Access (T1007)	Data Destruction (T1083)		Defense Evasion (T1566,...)	Initial Access, Delivery (T1005, T1041,...)	Execution (T1048)	Initial Access, Delivery (T1005)	Software Tampering (T1560)	Data Destructi... (T1083)	Execution (T1057)		Lateral Move... (T1106)	Security Software Bypass (T1518)	
	Lateral Movement (T1106)		Masquerade (T1033)	Defense Evasion (T1189, T1566,...)	User Interaction (T1036)	Resource Hijacking (T1074)	Defense Evasion (T1566, T1204)	Privilege Escalation (T1119)	Defense Evasion (T1566, T1204,...)		User Interaction (T1036)		
Execution (T1087)				Pretext Collecti... (T1069)	System Discov... (T1082)	Security Software Bypass (T1518)	Data Destruction (T1083)	Privilege Escalation (T1119)	Process I		Process VIII		Process VI
Process X				Process XIII		Initial Access (T1003)	Execution (T1059)	Defense Evasion (T1566, T1204)	User Interac... (T1036)	Comm... and Scripting (T1095)	Exfiltra... (T1571)		
Delivery (T1047, T1048, T1533)		Lateral Movement (T1106)		Masquerade (T1033)	Delivery (T1048, T1533)	Execution (T1087)	Defense Evasion (T1134)	Privilege Escalation (T1112)	Process XII		Process II		
Execution (T1087)					Defense Evasion (T1189, T1134)	Masquerade (T1033)	Lateral Movement (T1106)	Impersonati... (T1574)	Exec... (T10...	Pers... (T15...	Defe... (T12...	Defence Evasio...	Impac...
					Process XIV			Process IX		Process IV			
Defense Evasion (T1189, T1134)		Pretext Collec... (T1069)	User Intera... (T1036)	System Disco... (T1082)	Lateral Movement (T1106)	Pretext Collection (T1069)	Initial Access, Delivery (T1005)	Data Destru... (T1083)	Privilege Escala... (T1119)	Initial Access, Delivery...	Defense Evasion (T1204)	Execu... (T1059)	Defense Evasion (T156...

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](#)