Can We Tell the Threat Actor from Their ATT&CK TIDs?

RYUSUKE MASUOKA,
TOSHITAKA SATOMI, AND
KOJI YAMADA,
FUJITSU SYSTEM INTEGRATION LABORATORIES, JAPAN

Presenter Introduction

Ryusuke Masuoka, Ph.D, CISSP
Technical Advisor | Global Fujitsu Distinguished Engineer
FIRST Fujitsu FJC-CERT Team Representative
Founding Member, MITRE Engenutiy
Center for Threat Informed Defense (CTID)
OASIS CTI TC Voting Member
Fujitsu System Integration Laboratories
Chief Cybersecurity Advisor, *Japan Ministry of Defense*
Member, *Japan Cybercrime Control Center (JC3)*

Toshitaka Satomi
Researcher
Creator of ATT&CK Powered Suit (APS)
Holder of ALL MITRE ATT&CK Defender (MAD) certificates and badges
Fujitsu System Integration Laboratories
Member, *Japan Cybercrime Control Center (JC3)*

Koji Yamada, CISSP
Research Manager
FIRST FJC-CERT Team Member
(Network Monitoring, CTI)
Fujitsu System Integration Laboratories

Research Question

Can We Tell the Threat Actor from Their ATT&CK TIDs?

TID: Technique IDentifier

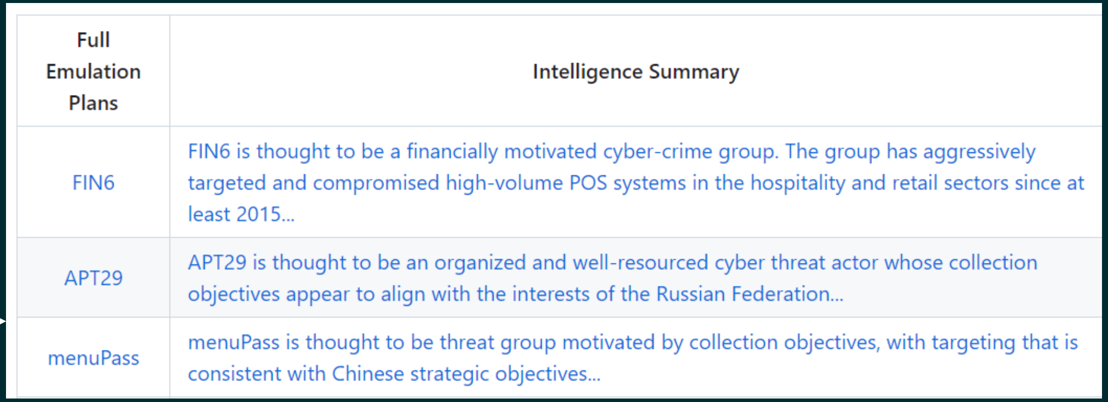
Why This Question Matters

Attribution improves your cyber defense

1. Incident Response
   - Enables effective countermeasures
2. Adversary Emulation
   - Helps determine scenarios for red teaming/BAS tools
3. SOC Assessment
   - Prioritizes controls to check based on threat actors

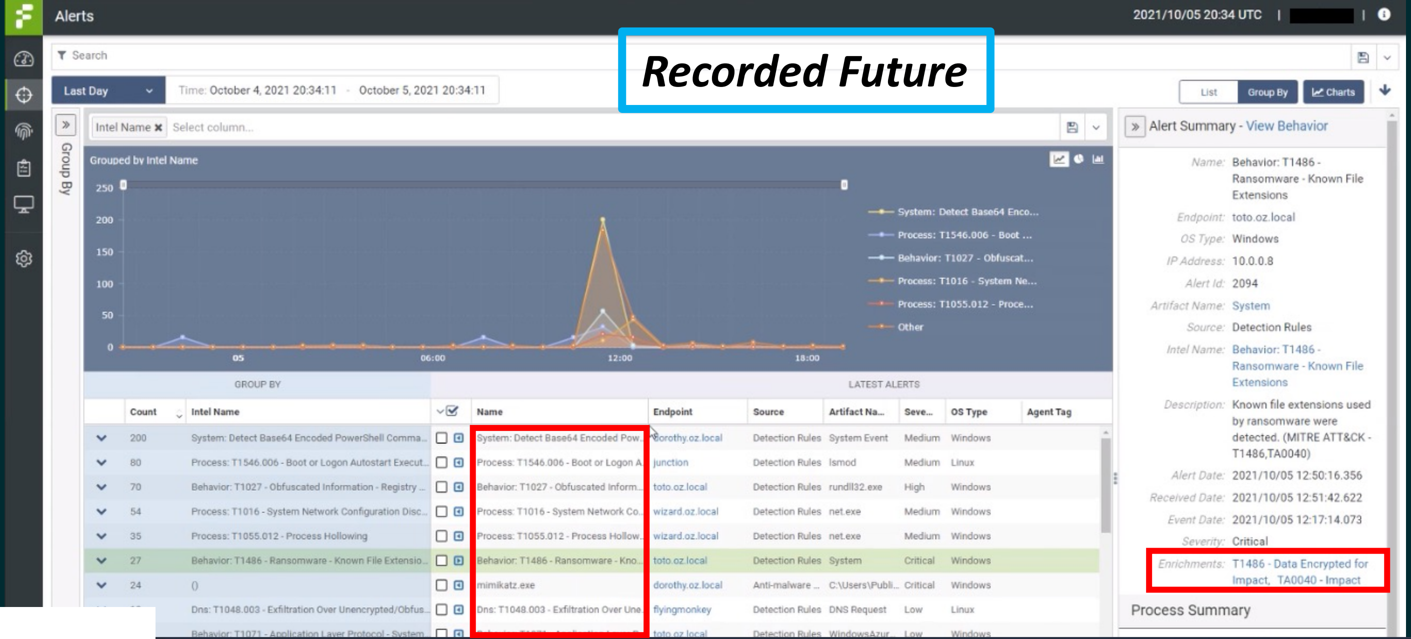
Adversary Emulation Library

[https://github.com/center-for-threat-informed-defense/adversary_emulation_library]



Full Emulation Plans	Intelligence Summary
FIN6	FIN6 is thought to be a financially motivated cyber-crime group. The group has aggressively targeted and compromised high-volume POS systems in the hospitality and retail sectors since at least 2015...
APT29	APT29 is thought to be an organized and well-resourced cyber threat actor whose collection objectives appear to align with the interests of the Russian Federation...
menuPass	menuPass is thought to be threat group motivated by collection objectives, with targeting that is consistent with Chinese strategic objectives...
⋮	⋮

TIDs in Tools

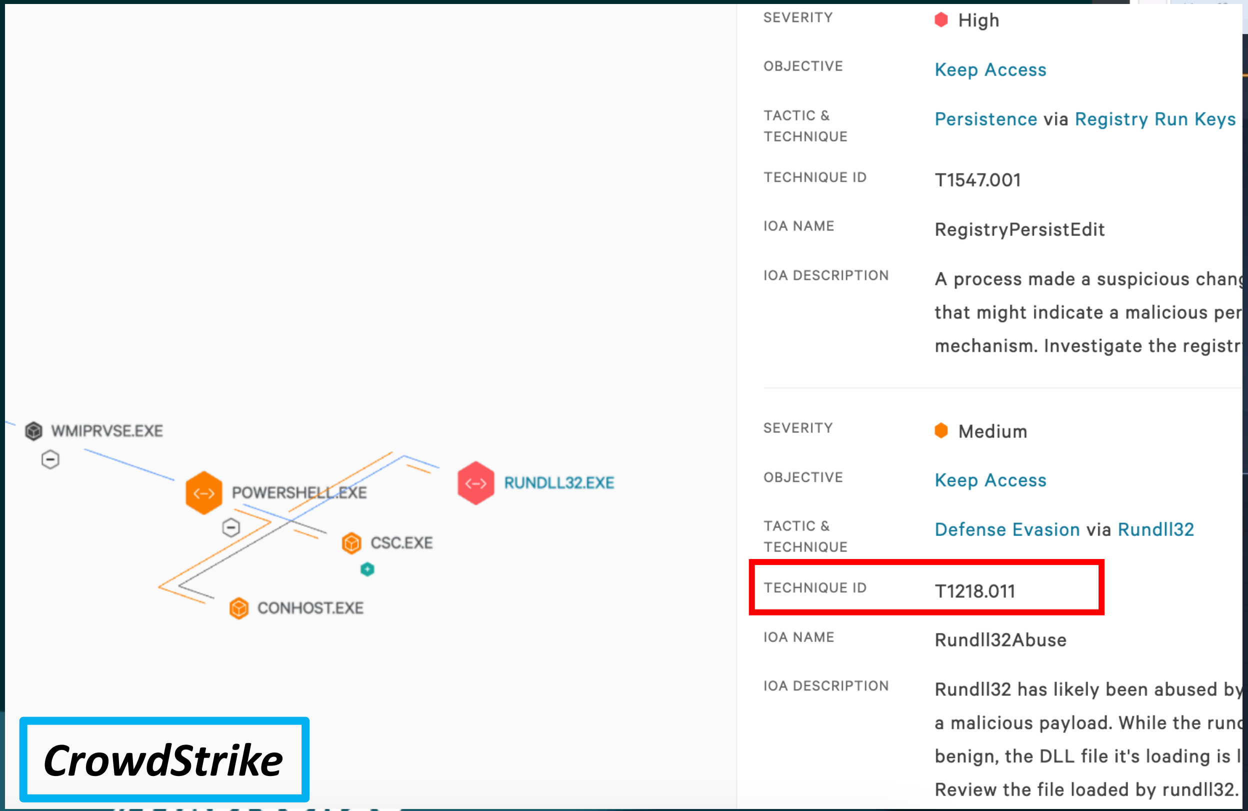

Recorded Future
Alerts
2021/10/05 20:34 UTC
Search
Last Day
Time: October 4, 2021 20:34:11 - October 5, 2021 20:34:11
List
Group By
Charts
Intel Name ✖
Select column...
Group By
Grouped by Intel Name
250
200
150
100
50
0
05
06:00
12:00
18:00
System: Detect Base64 Enco...
Process: T1546.006 - Boot ...
Behavior: T1027 - Obfuscat...
Process: T1016 - System Ne...
Process: T1055.012 - Proce...
Other

GROUP BY			LATEST ALERTS						
	Count	Intel Name	Name	Endpoint	Source	Artifact Na...	Seve...	OS Type	Agent Tag
⌄	200	System: Detect Base64 Encoded PowerShell Comma...	System: Detect Base64 Encoded Pow...	dorothy.oz.local	Detection Rules	System Event	Medium	Windows	
⌄	80	Process: T1546.006 - Boot or Logon Autostart Execut...	Process: T1546.006 - Boot or Logon A...	junction	Detection Rules	lsmod	Medium	Linux	
⌄	70	Behavior: T1027 - Obfuscated Information - Registry ...	Behavior: T1027 - Obfuscated Inform...	toto.oz.local	Detection Rules	rundll32.exe	High	Windows	
⌄	54	Process: T1016 - System Network Configuration Disc...	Process: T1016 - System Network Co...	wizard.oz.local	Detection Rules	net.exe	Medium	Windows	
⌄	35	Process: T1055.012 - Process Hollowing	Process: T1055.012 - Process Hollow...	wizard.oz.local	Detection Rules	net.exe	Medium	Windows	
⌄	27	Behavior: T1486 - Ransomware - Known File Extensio...	Behavior: T1486 - Ransomware - Kno...	toto.oz.local	Detection Rules	System	Critical	Windows	
⌄	24	()	mimikatz.exe	dorothy.oz.local	Anti-malware ...	C:\Users\Publi...	Critical	Windows	
		Dns: T1048.003 - Exfiltration Over Unencrypted/Obfus...	Dns: T1048.003 - Exfiltration Over Une...	flyingmonkey	Detection Rules	DNS Request	Low	Linux	
		Behavior: T1071 - Application Layer Protocol - System...		toto.oz.local	Detection Rules	WindowsAzur...	Low	Windows	

Alert Summary - View Behavior

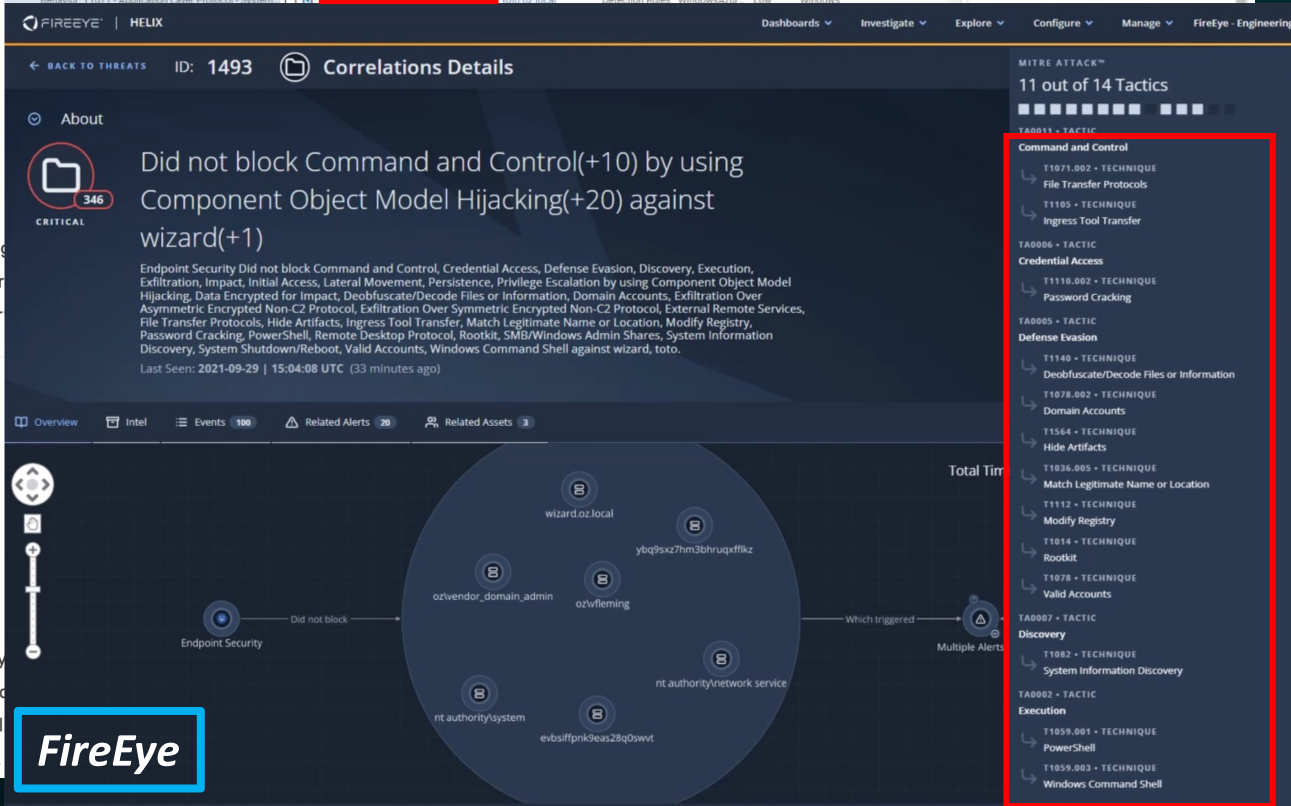
Name: Behavior: T1486 - Ransomware - Known File Extensions
Endpoint: toto.oz.local
OS Type: Windows
IP Address: 10.0.0.8
Alert Id: 2094
Artifact Name: System
Source: Detection Rules
Intel Name: Behavior: T1486 - Ransomware - Known File Extensions
Description: Known file extensions used by ransomware were detected. (MITRE ATT&CK - T1486,TA0040)
Alert Date: 2021/10/05 12:50:16.356
Received Date: 2021/10/05 12:51:42.622
Event Date: 2021/10/05 12:17:14.073
Severity: Critical
Enrichments: T1486 - Data Encrypted for Impact, TA0040 - Impact

Process Summary



SEVERITY: High
OBJECTIVE: Keep Access
TACTIC & TECHNIQUE: Persistence via Registry Run Keys
TECHNIQUE ID: T1547.001
IOA NAME: RegistryPersistEdit
IOA DESCRIPTION: A process made a suspicious chang
that might indicate a malicious per
mechanism. Investigate the registr

SEVERITY: Medium
OBJECTIVE: Keep Access
TACTIC & TECHNIQUE: Defense Evasion via Rundll32
TECHNIQUE ID: T1218.011
IOA NAME: Rundll32Abuse
IOA DESCRIPTION: Rundll32 has likely been abused by
a malicious payload. While the rund
benign, the DLL file it's loading is l
Review the file loaded by rundll32.



BACK TO THREATS ID: 1493 Correlations Details

About

Did not block Command and Control(+10) by using Component Object Model Hijacking(+20) against wizard(+1)

Endpoint Security Did not block Command and Control, Credential Access, Defense Evasion, Discovery, Execution, Exfiltration, Impact, Initial Access, Lateral Movement, Persistence, Privilege Escalation by using Component Object Model Hijacking, Data Encrypted for Impact, Deobfuscate/Decode Files or Information, Domain Accounts, Exfiltration Over Asymmetric Encrypted Non-C2 Protocol, Exfiltration Over Symmetric Encrypted Non-C2 Protocol, External Remote Services, File Transfer Protocols, Hide Artifacts, Ingress Tool Transfer, Match Legitimate Name or Location, Modify Registry, Password Cracking, PowerShell, Remote Desktop Protocol, Rootkit, SMB/Windows Admin Shares, System Information Discovery, System Shutdown/Reboot, Valid Accounts, Windows Command Shell against wizard, toto.

Last Seen: 2021-09-29 | 15:04:08 UTC (33 minutes ago)

MITRE ATTACK™

11 out of 14 Tactics

TA0011 • TACTIC
Command and Control
- T1071.002 • TECHNIQUE
  File Transfer Protocols
- T1105 • TECHNIQUE
  Ingress Tool Transfer

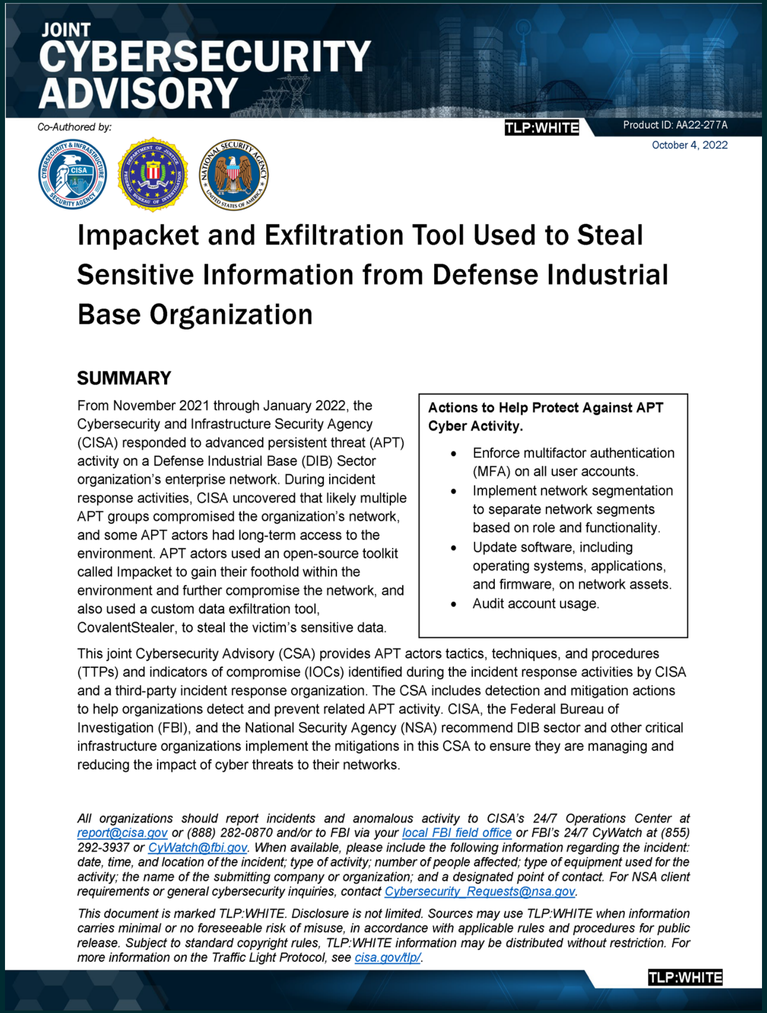
TA0006 • TACTIC
Credential Access
- T1110.002 • TECHNIQUE
  Password Cracking

TA0005 • TACTIC
Defense Evasion
- T1140 • TECHNIQUE
  Deobfuscate/Decode Files or Information
- T1078.002 • TECHNIQUE
  Domain Accounts
- T1564 • TECHNIQUE
  Hide Artifacts
- T1036.005 • TECHNIQUE
  Match Legitimate Name or Location
- T1112 • TECHNIQUE
  Modify Registry
- T1014 • TECHNIQUE
  Rootkit
- T1078 • TECHNIQUE
  Valid Accounts

TA0007 • TACTIC
Discovery
- T1082 • TECHNIQUE
  System Information Discovery

TA0002 • TACTIC
Execution
- T1059.001 • TECHNIQUE
  PowerShell
- T1059.003 • TECHNIQUE
  Windows Command Shell

TIDs in CTI Reports



JOINT CYBERSECURITY ADVISORY

Co-Authored by:

TLP:WHITE

Product ID: AA22-277A

October 4, 2022

Impacket and Exfiltration Tool Used to Steal Sensitive Information from Defense Industrial Base Organization

SUMMARY

From November 2021 through January 2022, the Cybersecurity and Infrastructure Security Agency (CISA) responded to advanced persistent threat (APT) activity on a Defense Industrial Base (DIB) Sector organization's enterprise network. During incident response activities, CISA uncovered that likely multiple APT groups compromised the organization's network, and some APT actors had long-term access to the environment. APT actors used an open-source toolkit called Impacket to gain their foothold within the environment and further compromise the network, and also used a custom data exfiltration tool, CovalentStealer, to steal the victim's sensitive data.

Actions to Help Protect Against APT Cyber Activity.

- Enforce multifactor authentication (MFA) on all user accounts.
- Implement network segmentation to separate network segments based on role and functionality.
- Update software, including operating systems, applications, and firmware, on network assets.
- Audit account usage.

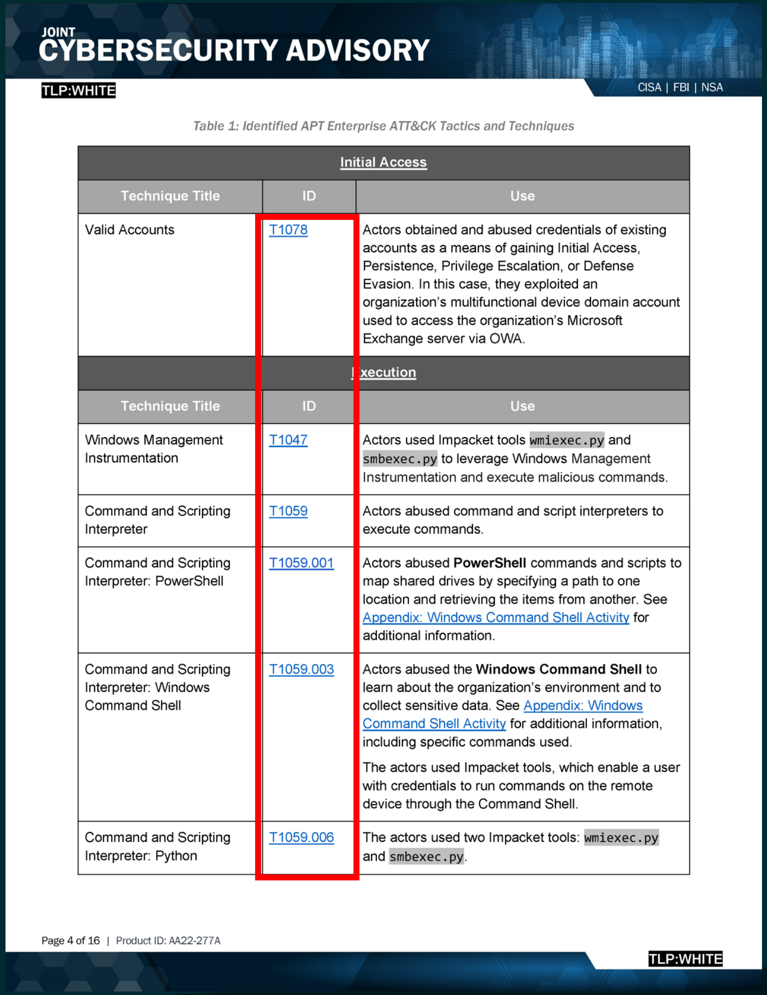
This joint Cybersecurity Advisory (CSA) provides APT actors tactics, techniques, and procedures (TTPs) and indicators of compromise (IOCs) identified during the incident response activities by CISA and a third-party incident response organization. The CSA includes detection and mitigation actions to help organizations detect and prevent related APT activity. CISA, the Federal Bureau of Investigation (FBI), and the National Security Agency (NSA) recommend DIB sector and other critical infrastructure organizations implement the mitigations in this CSA to ensure they are managing and reducing the impact of cyber threats to their networks.

All organizations should report incidents and anomalous activity to CISA's 24/7 Operations Center at report@cisa.gov or (888) 282-0870 and/or to FBI via your local FBI field office or FBI's 24/7 CyWatch at (855) 292-3937 or CyWatch@fbi.gov. When available, please include the following information regarding the incident: date, time, and location of the incident; type of activity; number of people affected; type of equipment used for the activity; the name of the submitting company or organization; and a designated point of contact. For NSA client requirements or general cybersecurity inquiries, contact Cybersecurity_Requests@nsa.gov.

This document is marked TLP:WHITE. Disclosure is not limited. Sources may use TLP:WHITE when information carries minimal or no foreseeable risk of misuse, in accordance with applicable rules and procedures for public release. Subject to standard copyright rules, TLP:WHITE information may be distributed without restriction. For more information on the Traffic Light Protocol, see cisa.gov/tlp/.

TLP:WHITE

...



JOINT CYBERSECURITY ADVISORY

TLP:WHITE

CISA | FBI | NSA

Table 1: Identified APT Enterprise ATT&CK Tactics and Techniques

Initial Access		
Technique Title	**ID**	**Use**
Valid Accounts	T1078	Actors obtained and abused credentials of existing accounts as a means of gaining Initial Access, Persistence, Privilege Escalation, or Defense Evasion. In this case, they exploited an organization's multifunctional device domain account used to access the organization's Microsoft Exchange server via OWA.
Execution		
Technique Title	**ID**	**Use**
Windows Management Instrumentation	T1047	Actors used Impacket tools `wmiexec.py` and `smbexec.py` to leverage Windows Management Instrumentation and execute malicious commands.
Command and Scripting Interpreter	T1059	Actors abused command and script interpreters to execute commands.
Command and Scripting Interpreter: PowerShell	T1059.001	Actors abused **PowerShell** commands and scripts to map shared drives by specifying a path to one location and retrieving the items from another. See Appendix: Windows Command Shell Activity for additional information.
Command and Scripting Interpreter: Windows Command Shell	T1059.003	Actors abused the **Windows Command Shell** to learn about the organization's environment and to collect sensitive data. See Appendix: Windows Command Shell Activity for additional information, including specific commands used. The actors used Impacket tools, which enable a user with credentials to run commands on the remote device through the Command Shell.
Command and Scripting Interpreter: Python	T1059.006	The actors used two Impacket tools: `wmiexec.py` and `smbexec.py`.

Page 4 of 16 | Product ID: AA22-277A

TLP:WHITE

...

Answer to the Research Question

Not a Complete Yes but Very Promising Results

Two Approaches

- TF-IDF

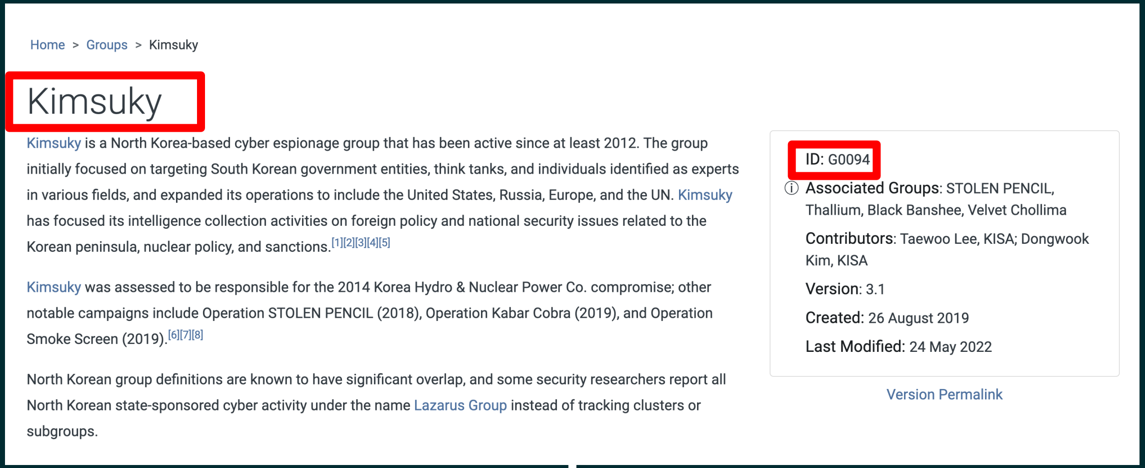
- Decision Tree

TF-IDF: Term Frequency–Inverse Document Frequency
TID: (ATT&CK) Technique IDentifier

TF-IDF

TF-IDF: Term Frequency–Inverse Document Frequency

ATT&CK Group



Home > Groups > Kimsuky

Kimsuky

Kimsuky is a North Korea-based cyber espionage group that has been active since at least 2012. The group initially focused on targeting South Korean government entities, think tanks, and individuals identified as experts in various fields, and expanded its operations to include the United States, Russia, Europe, and the UN. Kimsuky has focused its intelligence collection activities on foreign policy and national security issues related to the Korean peninsula, nuclear policy, and sanctions.[1][2][3][4][5]

Kimsuky was assessed to be responsible for the 2014 Korea Hydro & Nuclear Power Co. compromise; other notable campaigns include Operation STOLEN PENCIL (2018), Operation Kabar Cobra (2019), and Operation Smoke Screen (2019).[6][7][8]

North Korean group definitions are known to have significant overlap, and some security researchers report all North Korean state-sponsored cyber activity under the name Lazarus Group instead of tracking clusters or subgroups.

ID: G0094

ⓘ Associated Groups: STOLEN PENCIL, Thallium, Black Banshee, Velvet Chollima

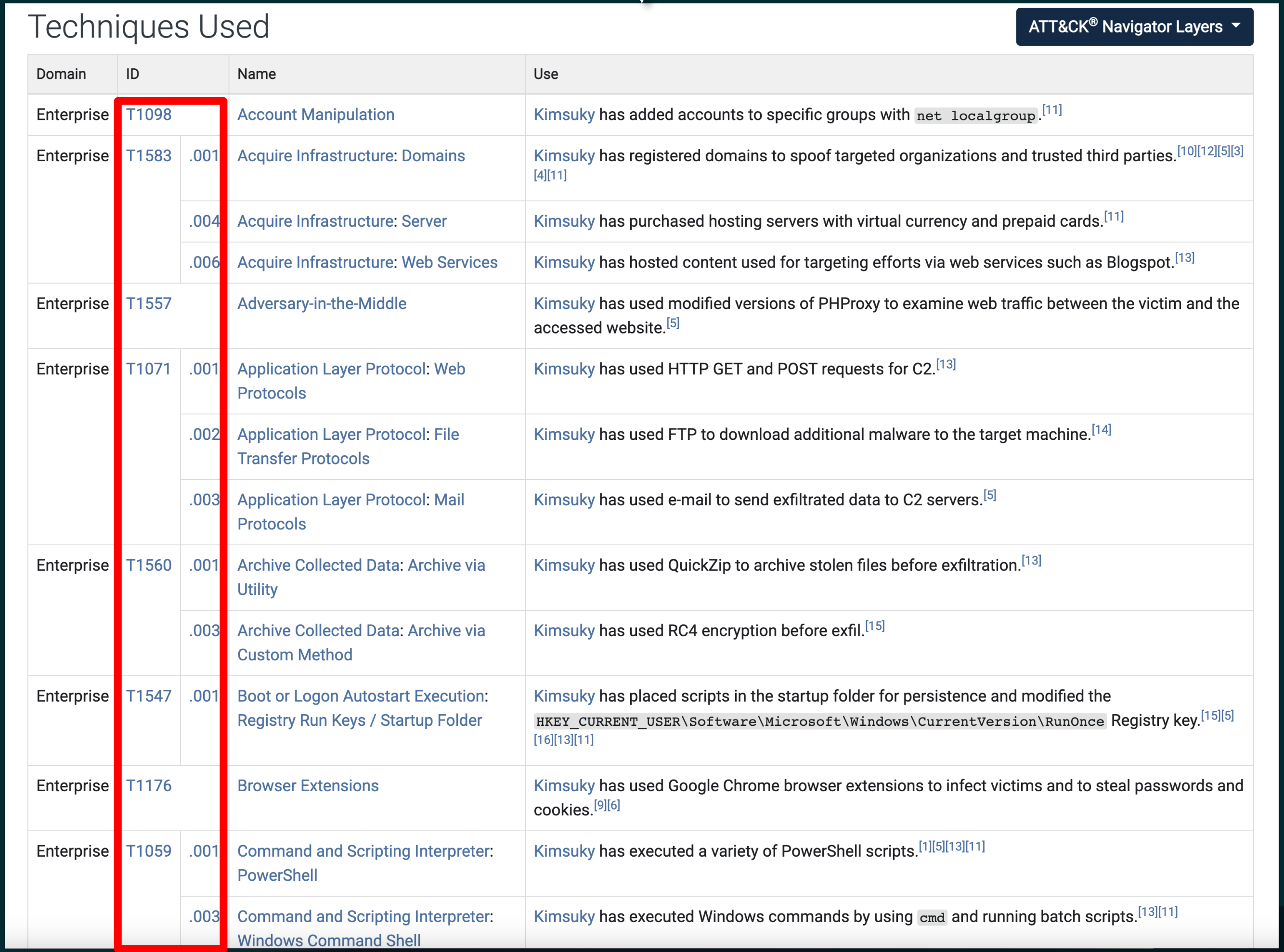
Contributors: Taewoo Lee, KISA; Dongwook Kim, KISA

Version: 3.1

Created: 26 August 2019

Last Modified: 24 May 2022

Version Permalink



Techniques Used

Domain	ID		Name	Use
Enterprise	T1098		Account Manipulation	Kimsuky has added accounts to specific groups with `net localgroup`.[11]
Enterprise	T1583	.001	Acquire Infrastructure: Domains	Kimsuky has registered domains to spoof targeted organizations and trusted third parties.[10][12][5][3][4][11]
		.004	Acquire Infrastructure: Server	Kimsuky has purchased hosting servers with virtual currency and prepaid cards.[11]
		.006	Acquire Infrastructure: Web Services	Kimsuky has hosted content used for targeting efforts via web services such as Blogspot.[13]
Enterprise	T1557		Adversary-in-the-Middle	Kimsuky has used modified versions of PHProxy to examine web traffic between the victim and the accessed website.[5]
Enterprise	T1071	.001	Application Layer Protocol: Web Protocols	Kimsuky has used HTTP GET and POST requests for C2.[13]
		.002	Application Layer Protocol: File Transfer Protocols	Kimsuky has used FTP to download additional malware to the target machine.[14]
		.003	Application Layer Protocol: Mail Protocols	Kimsuky has used e-mail to send exfiltrated data to C2 servers.[5]
Enterprise	T1560	.001	Archive Collected Data: Archive via Utility	Kimsuky has used QuickZip to archive stolen files before exfiltration.[13]
		.003	Archive Collected Data: Archive via Custom Method	Kimsuky has used RC4 encryption before exfil.[15]
Enterprise	T1547	.001	Boot or Logon Autostart Execution: Registry Run Keys / Startup Folder	Kimsuky has placed scripts in the startup folder for persistence and modified the `HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\RunOnce` Registry key.[15][5][16][13][11]
Enterprise	T1176		Browser Extensions	Kimsuky has used Google Chrome browser extensions to infect victims and to steal passwords and cookies.[9][6]
Enterprise	T1059	.001	Command and Scripting Interpreter: PowerShell	Kimsuky has executed a variety of PowerShell scripts.[1][5][13][11]
		.003	Command and Scripting Interpreter: Windows Command Shell	Kimsuky has executed Windows commands by using `cmd` and running batch scripts.[13][11]

Group TID Vectors

1 – TID in "Techniques Used" of the Group
0 – TID not in "Techniques Used" of the Group

TID Group	…	T1053.005	T1055	T1055.001	T1055.002	T1055.012	T1055.013	T1056.001	T1056.002	T1057	T1059	T1059.001	T1059.003	T1059.006	…
G0094 (Kimsuky)	…	0	1	0	0	0	0	1	0	0	0	1	0	0	…
G0040 (Patchwork)	…	1	0	0	0	1	0	0	0	0	0	1	1	0	…
G0074 (Dragonfly 2.0)	…	1	0	0	0	0	0	0	0	0	1	1	1	1	…
G0072 (Honeybee)	…	0	1	0	0	0	0	0	0	1	0	0	1	0	…
G0050 (APT32)	…	1	1	0	0	0	0	1	0	0	1	1	1	0	…
G0043 (Group5)	…	0	0	0	0	0	0	1	0	0	0	0	0	0	…
G0100 (Inception)	…	0	0	0	0	0	0	0	0	1	0	1	0	0	…
G0080 (Cobalt Group)	…	1	1	0	0	0	0	0	0	0	0	1	1	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Group TF-IDF Vectors

- Applied TF-IDF with TIDs as terms and Groups as documents

TID Group	…	T1053.005	T1055	T1055.001	T1055.002	T1055.012	T1055.013	T1056.001	T1056.002	T1057	T1059	T1059.001	T1059.003	T1059.006	…
G0094 (Kimsuky)	…	0	0.235	0	0	0	0	0.189	0	0	0	0.129	0	0	…
G0040 (Patchwork)	…	0.125	0	0	0	0.227	0	0	0	0	0	0.100	0.105	0	…
G0074 (Dragonfly 2.0)	…	0.106	0	0	0	0	0	0	0	0	0.147	0.085	0.089	0.177	…
G0072 (Honeybee)	…	0	0.239	0	0	0	0	0	0	0.181	0	0	0.137	0	…
G0050 (APT32)	…	0.083	0.122	0	0	0	0	0.098	0	0	0.116	0.067	0.070	0	…
G0043 (Group5)	…	0	0	0	0	0	0	0.457	0	0	0	0	0	0	…
G0100 (Inception)	…	0	0	0	0	0	0	0	0	0.185	0	0.133	0	0	…
G0080 (Cobalt Group)	…	0.126	0.185	0	0	0	0	0	0	0	0	0.101	0.106	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

TF-IDF: Term Frequency–Inverse Document Frequency

Evaluation – Using a CTI Report

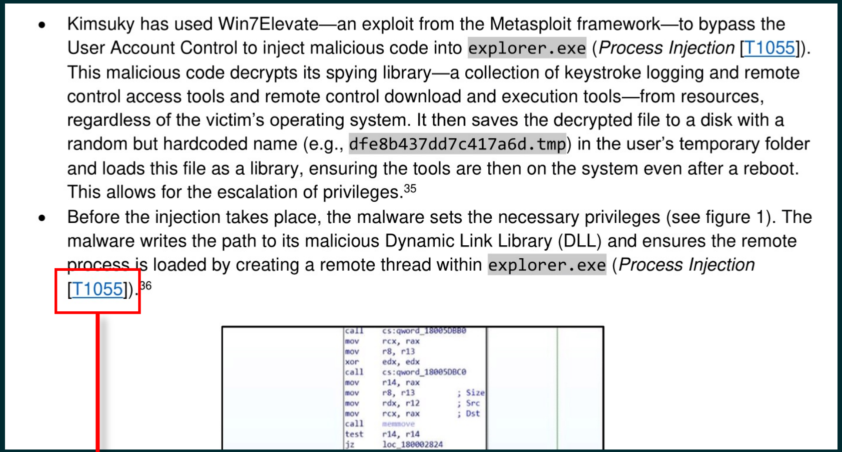
JOINT CYBERSECURITY ADVISORY

North Korean Advanced Persistent Threat Focus: Kimsuky

AA20-301A

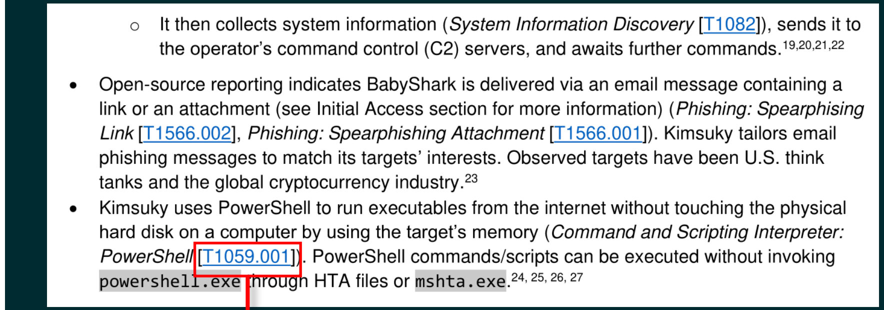
October 27, 2020

…



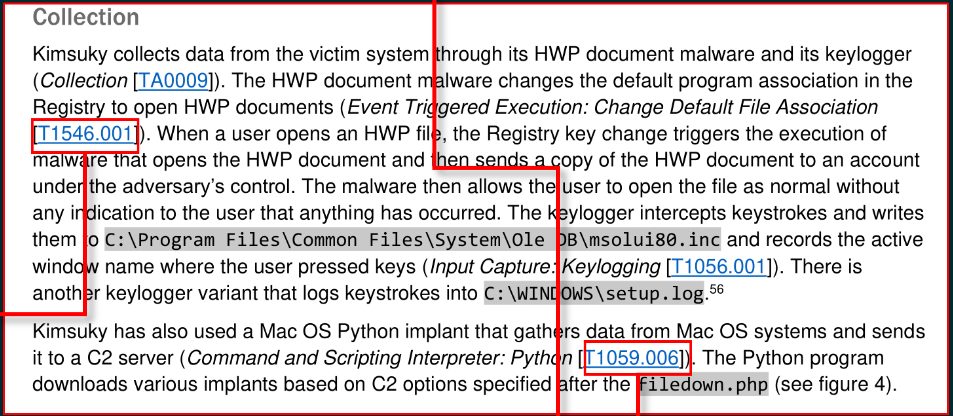
- Kimsuky has used Win7Elevate—an exploit from the Metasploit framework—to bypass the User Account Control to inject malicious code into `explorer.exe` (*Process Injection* [T1055]). This malicious code decrypts its spying library—a collection of keystroke logging and remote control access tools and remote control download and execution tools—from resources, regardless of the victim's operating system. It then saves the decrypted file to a disk with a random but hardcoded name (e.g., `dfe8b437dd7c417a6d.tmp`) in the user's temporary folder and loads this file as a library, ensuring the tools are then on the system even after a reboot. This allows for the escalation of privileges.[35]
- Before the injection takes place, the malware sets the necessary privileges (see figure 1). The malware writes the path to its malicious Dynamic Link Library (DLL) and ensures the remote process is loaded by creating a remote thread within `explorer.exe` (*Process Injection* [T1055]).[36]

```
call    cs:qword_18005DBB0
mov     rcx, rax
mov     r8, r13
xor     edx, edx
call    cs:qword_18005DBC0
mov     r14, rax
mov     r8, r13         ; Size
mov     rdx, r12        ; Src
mov     rcx, rax        ; Dst
call    memmove
test    r14, r14
jz      loc_180002824
```



  ○ It then collects system information (*System Information Discovery* [T1082]), sends it to the operator's command control (C2) servers, and awaits further commands.[19,20,21,22]

- Open-source reporting indicates BabyShark is delivered via an email message containing a link or an attachment (see Initial Access section for more information) (*Phishing: Spearphishing Link* [T1566.002], *Phishing: Spearphishing Attachment* [T1566.001]). Kimsuky tailors email phishing messages to match its targets' interests. Observed targets have been U.S. think tanks and the global cryptocurrency industry.[23]
- Kimsuky uses PowerShell to run executables from the internet without touching the physical hard disk on a computer by using the target's memory (*Command and Scripting Interpreter: PowerShell* [T1059.001]). PowerShell commands/scripts can be executed without invoking `powershell.exe` through HTA files or `mshta.exe`.[24, 25, 26, 27]



Collection

Kimsuky collects data from the victim system through its HWP document malware and its keylogger (*Collection* [TA0009]). The HWP document malware changes the default program association in the Registry to open HWP documents (*Event Triggered Execution: Change Default File Association* [T1546.001]). When a user opens an HWP file, the Registry key change triggers the execution of malware that opens the HWP document and then sends a copy of the HWP document to an account under the adversary's control. The malware then allows the user to open the file as normal without any indication to the user that anything has occurred. The keylogger intercepts keystrokes and writes them to `C:\Program Files\Common Files\System\Ole DB\msolui80.inc` and records the active window name where the user pressed keys (*Input Capture: Keylogging* [T1056.001]). There is another keylogger variant that logs keystrokes into `C:\WINDOWS\setup.log`.[56]

Kimsuky has also used a Mac OS Python implant that gathers data from Mac OS systems and sends it to a C2 server (*Command and Scripting Interpreter: Python* [T1059.006]). The Python program downloads various implants based on C2 options specified after the `filedown.php` (see figure 4).

CTI TID Vector)

TID / Report	…	T1053.005	T1055	T1055.001	T1055.002	T1055.012	T1055.013	T1056.001	T1056.002	T1057	T1059	T1059.001	T1059.003	T1059.006	…
aa20-301a	…	0	1	0	0	0	0	1	0	0	0	1	0	1	…

Determining Likely Threat Actors (Groups)

TID Group	…	T1053.005	T1055	T1055.001	T1055.002	T1055.012	T1055.013	T1056.001	T1056.002	T1057	T1059	T1059.001	T1059.003	T1059.006	…	< , X>
G0094 (Kimsuky)	…	0	0.235	0	0	0	0	0.189	0	0	0	0.129	0	0	…	0.498
G0040 (Patchwork)	…	0.125	0	0	0	0.227	0	0	0	0	0	0.100	0.105	0	…	0.316
G0074 (Dragonfly 2.0)	…	0.106	0	0	0	0	0	0	0	0	0.147	0.085	0.089	0.177	…	0.313
G0072 (Honeybee)	…	0	0.239	0	0	0	0	0	0	0.181	0	0	0.137	0	…	0.305
G0050 (APT32)	…	0.083	0.122	0	0	0	0	0.098	0	0	0.116	0.067	0.070	0	…	0.291
G0043 (Group5)	…	0	0	0	0	0	0	0.457	0	0	0	0	0	0	…	0.256
G0100 (Inception)	…	0	0	0	0	0	0	0	0	0.185	0	0.133	0	0	…	0.247
G0080 (Cobalt Group)	…	0.126	0.185	0	0	0	0	0	0	0	0	0.101	0.106	0	…	0.243
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Inner Product

CTI TID Vector (Normalized)

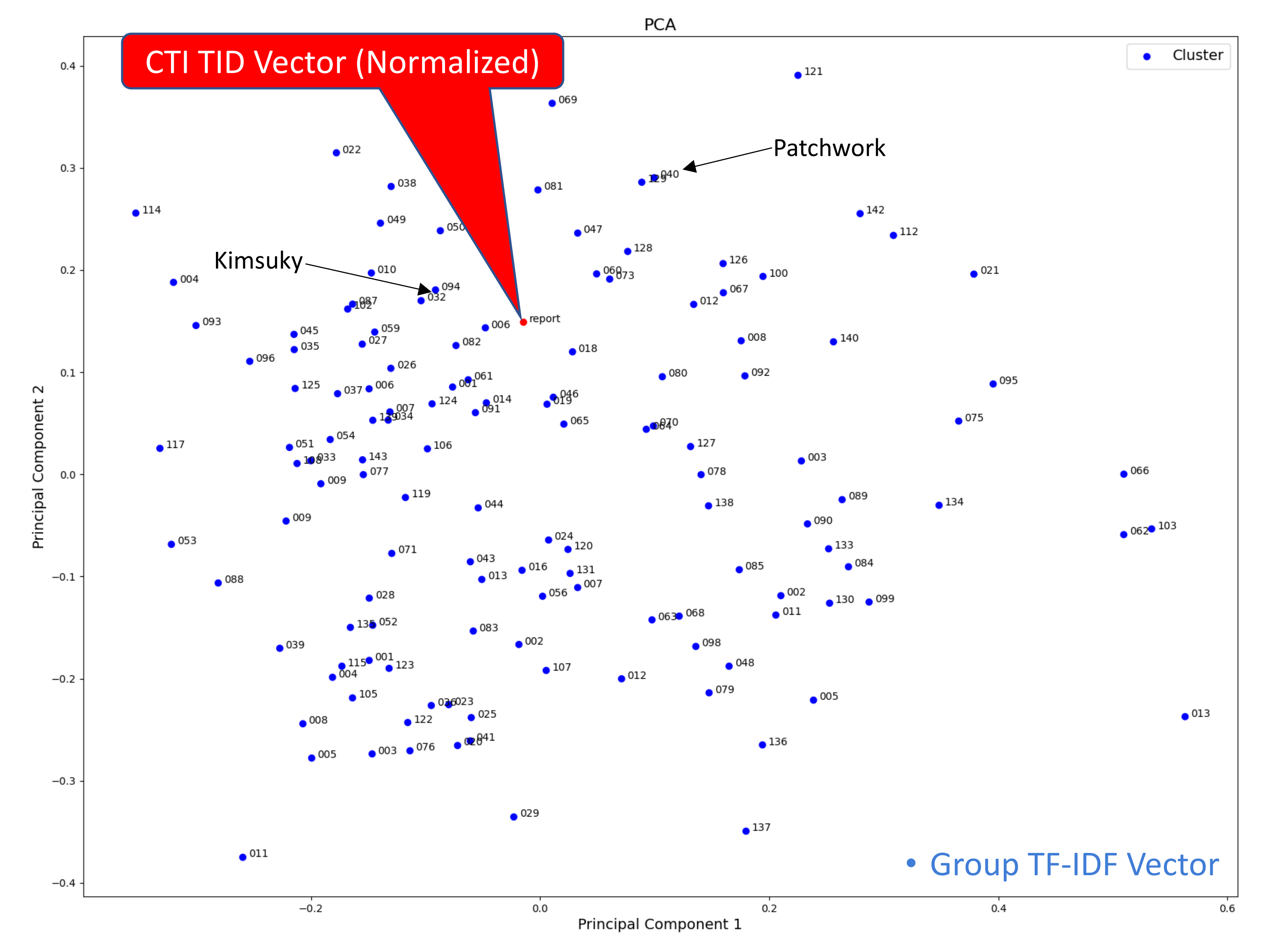
TID Report	…	T1053.005	T1055	T1055.001	T1055.002	T1055.012	T1055.013	T1056.001	T1056.002	T1057	T1059	T1059.001	T1059.003	T1059.006	…
aa20-301a	…	0	0.189	0	0	0	0	0.189	0	0	0	0.189	0	0.189	…

= X

= X

PCA to Visualize the Idea

- Applied Principal Component Analysis (PCA) to Group TF-IDF vectors
- Plotted the CTI TID vector in the vector space



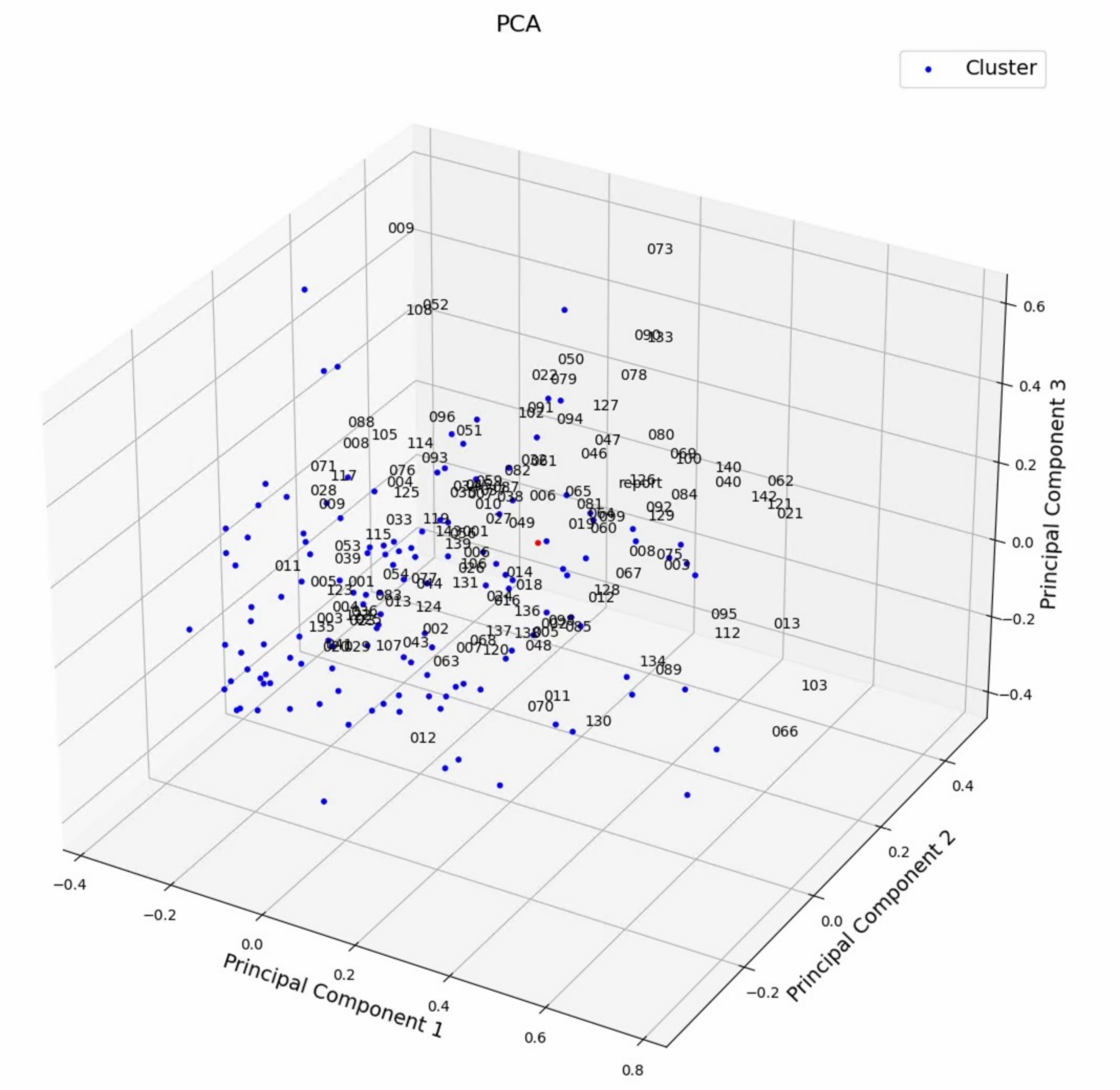
CTI TID Vector (Normalized)

Patchwork

Kimsuky

• Group TF-IDF Vector

In Three Dimension



TF-IDF Evaluations

- TF-IDF Results
  - CISA Kimsuky Report -> “Kimsuky”: No. 1
  - ESET OceanLotus Report –> “APT32/OceanLotus”: No. 2
- “APT32” not in top 30 for Mandiant APT32 Report
  - Problems caused by:
    - Groups with too few TIDs
    - Top 15 most sighted TIDs
  - Tuned algorithm moved “APT32/OceanLotus” to No. 10
- Note: Pitfalls with CTI reports
  - Ex. ATT&CK includes the CTI report -> Rewind the ATT&CK version

TF-IDF –
Say you don’t know
what you don’t know!

In Case of Mandiant APT43 Report

TID Group	score
G0112 (Windshift)	0.373
G0049 (OilRig)	0.360
G0021 (Molerats)	0.359
G0040 (Patchwork)	0.358
G0050 (APT32)	0.356
…	…

- Is G0112 still the winner?
- Actually this is a trick question …
  - APT43 is not in ATT&CK!

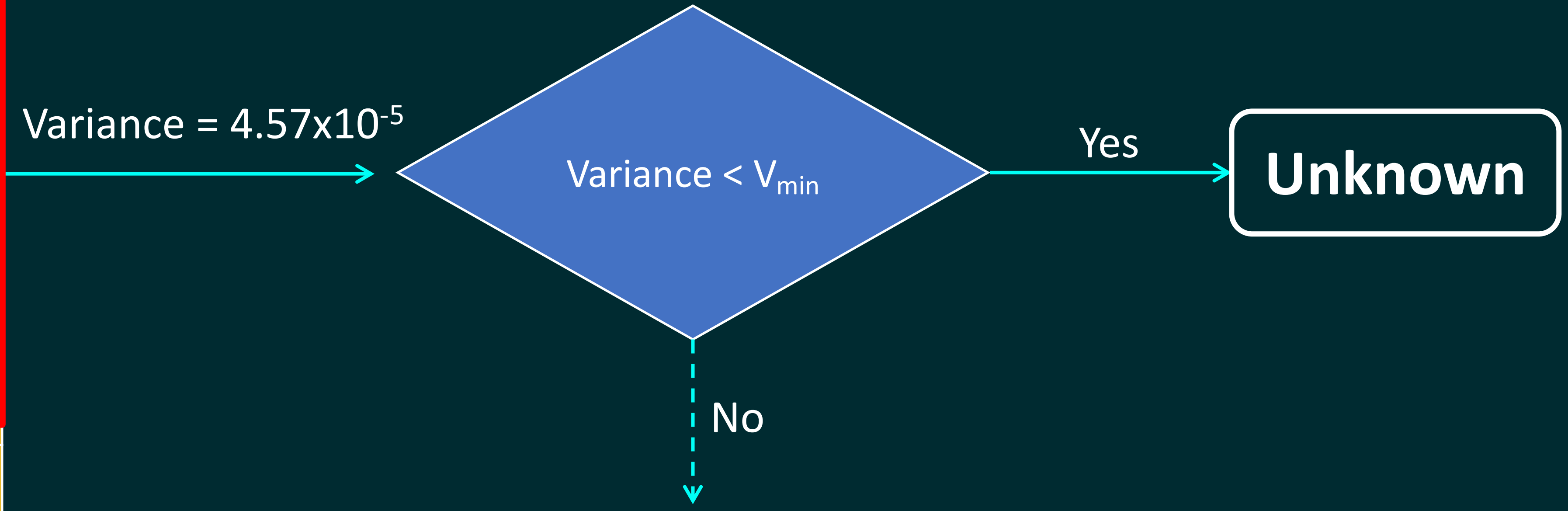
- Can we say we don’t know what we don’t know?

Mandiant APT43 Report:
https://www.mandiant.com/resources/blog/apt43-north-korea-cybercrime-espionage

Yes, we can … say we don’t know what we don’t know

TID Group	score
G0112 (Windshift)	0.373
G0049 (OilRig)	0.360
G0021 (Molerats)	0.359
G0040 (Patchwork)	0.358
G0050 (APT32)	0.356
…	…

- When the variance is small (below a given threshold, V_{min}), we say we don't know



Variance = $4.57x10^{-5}$

Variance < V_{min}

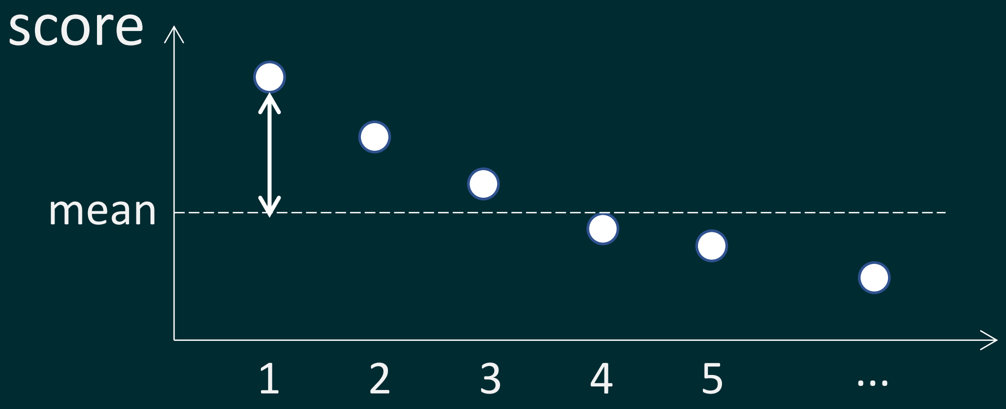
Yes

Unknown

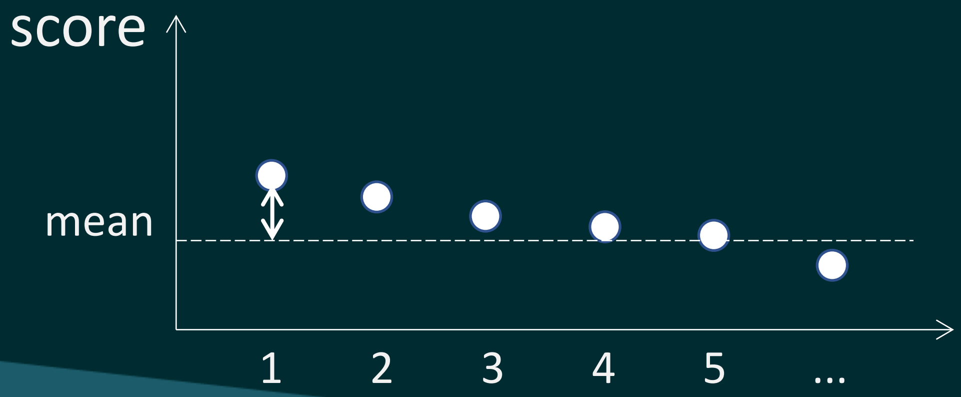
No

“**Unknown**” Can Be a Right Answer

- … and a responsible one when:
  - A new threat group unknown before pops up
  - A threat group sometimes changes their techniques



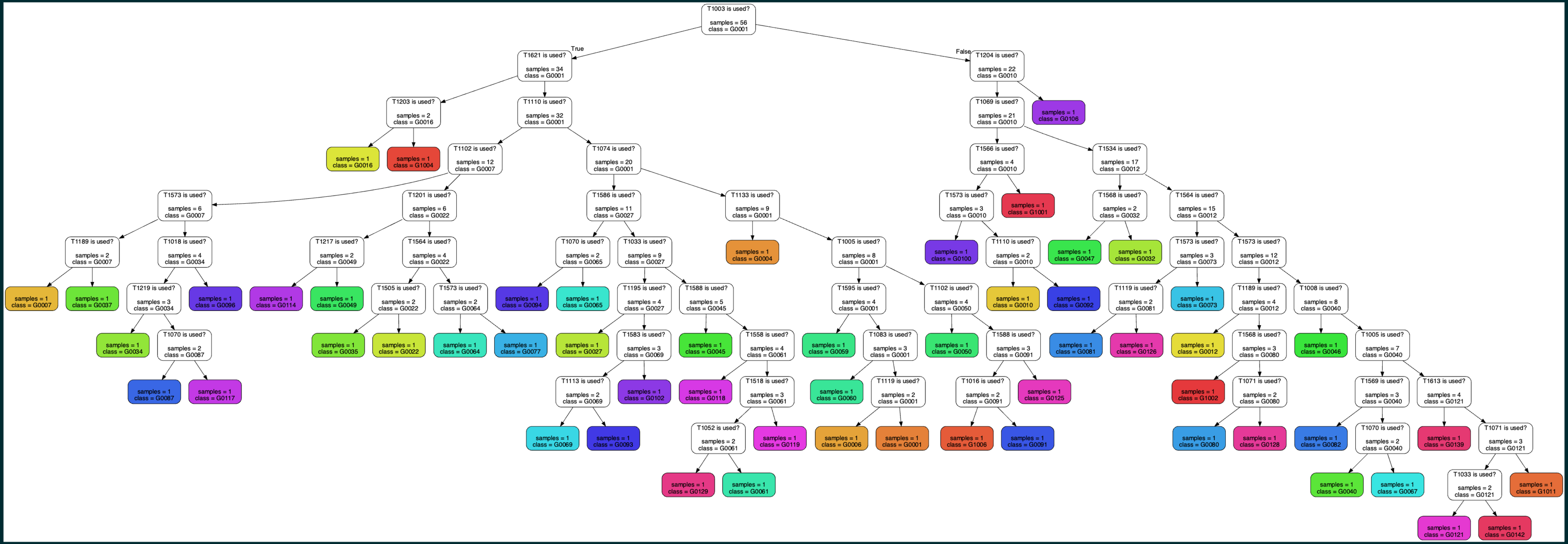
High variance of TF/IDF scores
-> Groups with high scores



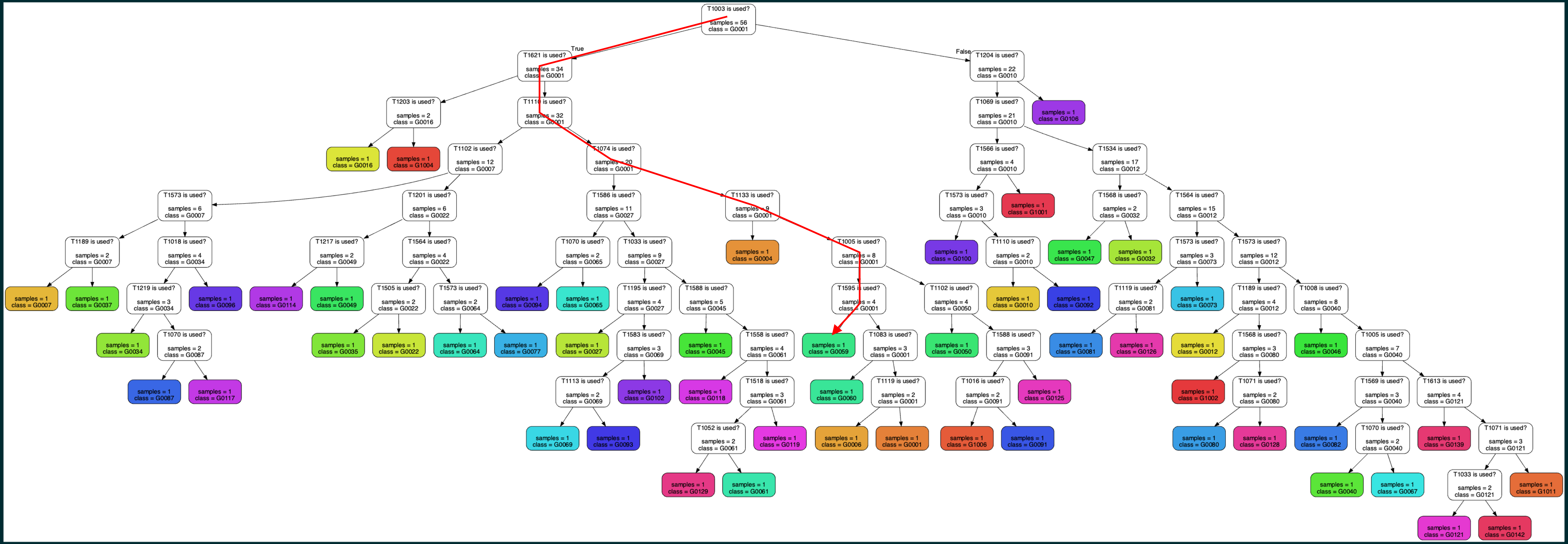
Low variance of TF/IDF scores
-> “Unknown”

Decision Tree

Decision Tree Example

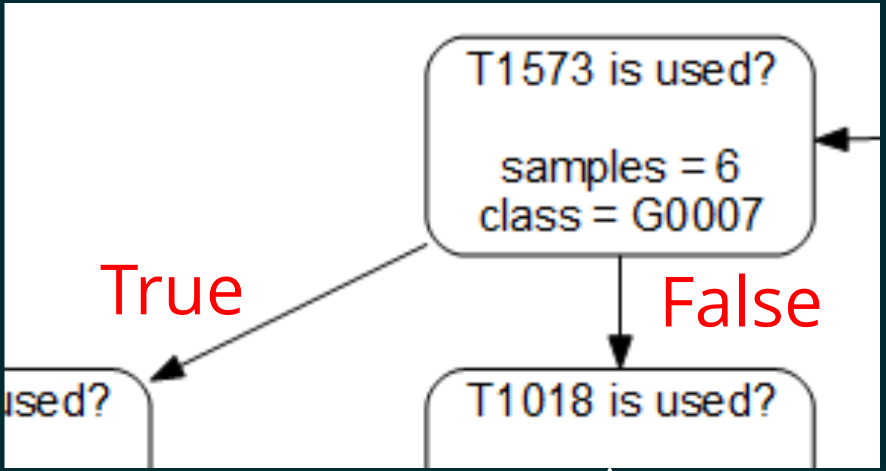

T1003 is used? samples = 56 class = G0001
True
False
T1621 is used? samples = 34 class = G0001
T1204 is used? samples = 22 class = G0010
T1203 is used? samples = 2 class = G0016
T1110 is used? samples = 32 class = G0001
T1069 is used? samples = 21 class = G0010
samples = 1 class = G0106
samples = 1 class = G0016
samples = 1 class = G1004
T1102 is used? samples = 12 class = G0007
T1074 is used? samples = 20 class = G0001
T1566 is used? samples = 4 class = G0010
T1534 is used? samples = 17 class = G0012
T1573 is used? samples = 6 class = G0007
T1201 is used? samples = 6 class = G0022
T1586 is used? samples = 11 class = G0027
T1133 is used? samples = 9 class = G0001
T1573 is used? samples = 3 class = G0010
samples = 1 class = G1001
T1568 is used? samples = 2 class = G0032
T1564 is used? samples = 15 class = G0012
T1189 is used? samples = 2 class = G0007
T1018 is used? samples = 4 class = G0034
T1217 is used? samples = 2 class = G0049
T1564 is used? samples = 4 class = G0022
T1070 is used? samples = 2 class = G0065
T1033 is used? samples = 9 class = G0027
samples = 1 class = G0004
T1005 is used? samples = 8 class = G0001
samples = 1 class = G0100
T1110 is used? samples = 2 class = G0010
samples = 1 class = G0047
samples = 1 class = G0032
T1573 is used? samples = 3 class = G0073
T1573 is used? samples = 12 class = G0012
samples = 1 class = G0007
samples = 1 class = G0037
T1219 is used? samples = 3 class = G0034
samples = 1 class = G0096
samples = 1 class = G0114
samples = 1 class = G0049
T1505 is used? samples = 2 class = G0022
T1573 is used? samples = 2 class = G0064
samples = 1 class = G0094
samples = 1 class = G0065
T1195 is used? samples = 4 class = G0027
T1588 is used? samples = 5 class = G0045
T1595 is used? samples = 4 class = G0001
T1102 is used? samples = 4 class = G0050
samples = 1 class = G0010
samples = 1 class = G0092
T1119 is used? samples = 2 class = G0081
samples = 1 class = G0073
T1189 is used? samples = 4 class = G0012
T1008 is used? samples = 8 class = G0040
samples = 1 class = G0034
T1070 is used? samples = 2 class = G0087
samples = 1 class = G0035
samples = 1 class = G0022
samples = 1 class = G0064
samples = 1 class = G0077
samples = 1 class = G0027
T1583 is used? samples = 3 class = G0069
samples = 1 class = G0045
T1558 is used? samples = 4 class = G0061
samples = 1 class = G0059
T1083 is used? samples = 3 class = G0001
samples = 1 class = G0050
T1588 is used? samples = 3 class = G0091
samples = 1 class = G0081
samples = 1 class = G0126
samples = 1 class = G0012
T1568 is used? samples = 3 class = G0080
samples = 1 class = G0046
T1005 is used? samples = 7 class = G0040
samples = 1 class = G0087
samples = 1 class = G0117
T1113 is used? samples = 2 class = G0069
samples = 1 class = G0102
samples = 1 class = G0118
T1518 is used? samples = 3 class = G0061
samples = 1 class = G0060
T1119 is used? samples = 2 class = G0001
T1016 is used? samples = 2 class = G0091
samples = 1 class = G0125
samples = 1 class = G1002
T1071 is used? samples = 2 class = G0080
T1569 is used? samples = 3 class = G0040
T1613 is used? samples = 4 class = G0121
samples = 1 class = G0069
samples = 1 class = G0093
T1052 is used? samples = 2 class = G0061
samples = 1 class = G0119
samples = 1 class = G0006
samples = 1 class = G0001
samples = 1 class = G1006
samples = 1 class = G0091
samples = 1 class = G0080
samples = 1 class = G0128
samples = 1 class = G0082
T1070 is used? samples = 2 class = G0040
samples = 1 class = G0139
T1071 is used? samples = 3 class = G0121
samples = 1 class = G0129
samples = 1 class = G0061
samples = 1 class = G0040
samples = 1 class = G0067
T1033 is used? samples = 2 class = G0121
samples = 1 class = G1011
samples = 1 class = G0121
samples = 1 class = G0142

Decision Tree for Attribution


T1003 is used? samples = 56 class = G0001
True
False
T1621 is used? samples = 34 class = G0001
T1204 is used? samples = 22 class = G0010
T1203 is used? samples = 2 class = G0016
T1110 is used? samples = 32 class = G0001
T1069 is used? samples = 21 class = G0010
samples = 1 class = G0106
samples = 1 class = G0016
samples = 1 class = G1004
T1102 is used? samples = 12 class = G0007
T1074 is used? samples = 20 class = G0001
T1566 is used? samples = 4 class = G0010
T1534 is used? samples = 17 class = G0012
T1573 is used? samples = 6 class = G0007
T1201 is used? samples = 6 class = G0022
T1586 is used? samples = 11 class = G0027
T1133 is used? samples = 9 class = G0001
T1573 is used? samples = 3 class = G0010
samples = 1 class = G1001
T1568 is used? samples = 2 class = G0032
T1564 is used? samples = 15 class = G0012
T1189 is used? samples = 2 class = G0007
T1018 is used? samples = 4 class = G0034
T1217 is used? samples = 2 class = G0049
T1564 is used? samples = 4 class = G0022
T1070 is used? samples = 2 class = G0065
T1033 is used? samples = 9 class = G0027
samples = 1 class = G0004
T1005 is used? samples = 8 class = G0001
samples = 1 class = G0100
T1110 is used? samples = 2 class = G0010
samples = 1 class = G0047
samples = 1 class = G0032
T1573 is used? samples = 3 class = G0073
T1573 is used? samples = 12 class = G0012
samples = 1 class = G0007
samples = 1 class = G0037
T1219 is used? samples = 3 class = G0034
samples = 1 class = G0096
samples = 1 class = G0114
samples = 1 class = G0049
T1505 is used? samples = 2 class = G0022
T1573 is used? samples = 2 class = G0064
samples = 1 class = G0094
samples = 1 class = G0065
T1195 is used? samples = 4 class = G0027
T1588 is used? samples = 5 class = G0045
T1595 is used? samples = 4 class = G0001
T1102 is used? samples = 4 class = G0050
samples = 1 class = G0010
samples = 1 class = G0092
T1119 is used? samples = 2 class = G0081
samples = 1 class = G0073
T1189 is used? samples = 4 class = G0012
T1008 is used? samples = 8 class = G0040
samples = 1 class = G0034
T1070 is used? samples = 2 class = G0087
samples = 1 class = G0035
samples = 1 class = G0022
samples = 1 class = G0064
samples = 1 class = G0077
samples = 1 class = G0027
T1583 is used? samples = 3 class = G0069
samples = 1 class = G0045
T1558 is used? samples = 4 class = G0061
samples = 1 class = G0059
T1083 is used? samples = 3 class = G0001
samples = 1 class = G0050
T1588 is used? samples = 3 class = G0091
samples = 1 class = G0081
samples = 1 class = G0126
samples = 1 class = G0012
T1568 is used? samples = 3 class = G0080
samples = 1 class = G0046
T1005 is used? samples = 7 class = G0040
samples = 1 class = G0087
samples = 1 class = G0117
T1113 is used? samples = 2 class = G0069
samples = 1 class = G0102
samples = 1 class = G0118
T1518 is used? samples = 3 class = G0061
samples = 1 class = G0060
T1119 is used? samples = 2 class = G0001
T1016 is used? samples = 2 class = G0091
samples = 1 class = G0125
samples = 1 class = G1002
T1071 is used? samples = 2 class = G0080
T1569 is used? samples = 3 class = G0040
T1613 is used? samples = 4 class = G0121
samples = 1 class = G0069
samples = 1 class = G0093
T1052 is used? samples = 2 class = G0061
samples = 1 class = G0119
samples = 1 class = G0006
samples = 1 class = G0001
samples = 1 class = G1006
samples = 1 class = G0091
samples = 1 class = G0080
samples = 1 class = G0128
samples = 1 class = G0082
T1070 is used? samples = 2 class = G0040
samples = 1 class = G0139
T1071 is used? samples = 3 class = G0121
samples = 1 class = G0129
samples = 1 class = G0061
samples = 1 class = G0040
samples = 1 class = G0067
T1033 is used? samples = 2 class = G0121
samples = 1 class = G1011
samples = 1 class = G0121
samples = 1 class = G0142

Attribution by Decision Tree

- Why you did not observe a technique?
  - You simply failed to observe it
  - The threat actor did not use it in their arsenal for this attack
  - The threat actor does not use it in general
- The absence of a technique observation may lead to a wrong conclusion

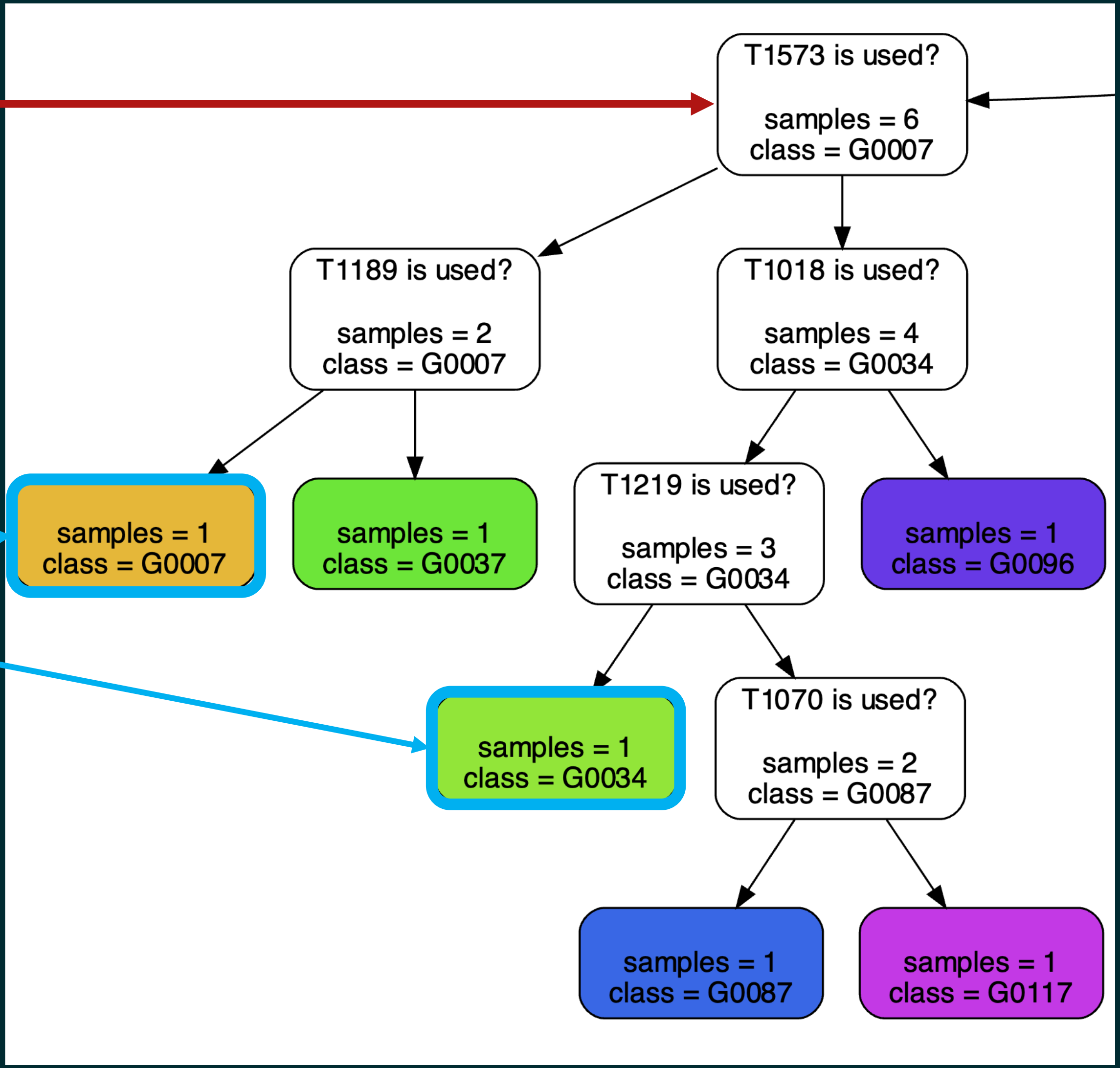


Threat Hunting in ACH Context

Threat Hunt for T1573 (Encrypted Channel)

Competing Hypotheses

- G0007 (APT28)
- G0034 (Sandworm)



ACH: Analysis of Competing Hypotheses

Conclusions

Our Journey of 2020

- We found the treasure where human and system defenders share

Road Behind
Signpost
Bridge
Where Human and System Defenders Share
Treasure
Sea of IoCs

We are back and …

- Found a way to extract treasure from MITRE ATT&CK
  - … an enabler for attribution

Summary

- Can We Tell the Threat Actor from Their ATT&CK TIDs?
  - ➢Not a Complete Yes but Very Promising Results

- Landscape changes enabled our approaches
  - Changes: Wider availability of observed ATT&CK TIDs
  - Approaches: TF-IDF and Decision Tree
- Making threat actor attribution accessible for many organizations
  - Improves your cyber defense, and
  - Increases the exposure risk for your adversary

Our Next Journey

1. Other evaluation methods
2. Finer-grained dataset like Campaigns instead of Groups
3. Additional elements like Software, titles/texts of CTI reports

Takeaway Message

- TIDs observed in a cyber attack should help you make more informed attribution of the cyber attack
- This capability makes your cyber defenses more proactive by knowing which threat actors are actively targeting you

Thank you

Ryusuke Masuoka: @rmasuoka

Toshitaka Satomi: @stmtstk

Koji Yamada: @ykoji8681

Thank you and see you in Fukuoka!

Ryusuke Masuoka: @rmasuoka

Toshitaka Satomi: @stmtstk

Koji Yamada: @ykoji8681

Q & A

#FIRSTCON24

36TH ANNUAL
FIRST CONFERENCE
FUKUOKA
JUNE 9-14, 2024 JAPAN