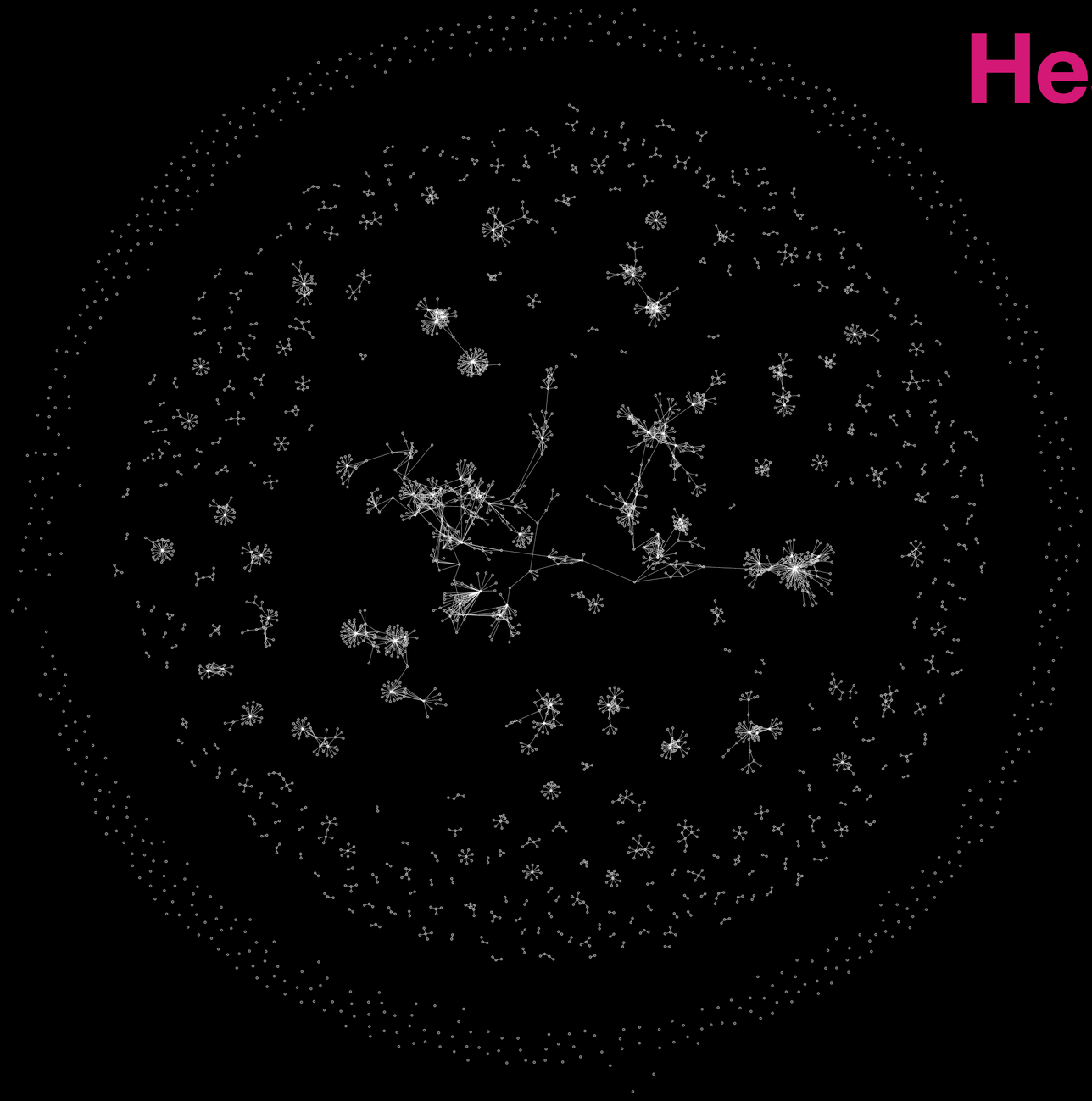


Hesperus is Phosphorus

Lessons from Mapping Threat Actor Naming Taxonomies at Scale

Juan Tapiador
Universidad Carlos III de Madrid

init.g()

Málaga — 30 Nov 2025

Hesperus is Phosphorus: Mapping Threat Actor Naming Taxonomies at Scale*

Gonzalo Roa, Manuel Suarez-Roman and Juan Tapiador*

Universidad Carlos III de Madrid, Avenida de la Universidad, 30, Leganés, 28911, Spain

ARTICLE INFO

Keywords:
Cyber threat intelligence
Threat actor naming

ABSTRACT

This paper studies the problem of Threat Actor (TA) naming convention inconsistency across leading Cyber Threat Intelligence (CTI) vendors. The current decentralized and proprietary nomenclature creates confusion and significant obstacles for researchers, including difficulties in integrating and correlating disparate CTI reports and TA profiles. This paper introduces HiP (Hesperus is Phosphorus, a reference to the classic question about the Morning and the Evening Star), a methodology for normalizing, integrating, and clustering TA names presumably corresponding to the same entity. Using HiP, we analyze a large dataset collected from 15 sources and spanning 13,371 CTI reports, 17 vendor taxonomies, 3,287 TA names, and 8 mappings between them. Our analysis of the resulting name graph provides insights on key features of the problem, such as the concentration of aliases on a relatively small subset of TAs, the evolution of this phenomenon over the years, and the factors that could explain TA name proliferation. We also report errors in the mappings and methodological pitfalls that contribute to make certain TA name clusters larger than they should be, including the use of temporary names for activity clusters, the existence of common tools and infrastructure, and overlapping operations. We conclude with a discussion on the inherent difficulties to adopt a TA naming standard, a quest fundamentally hampered by the need to share highly-sensitive telemetry that is private to each CTI vendor.

1. Introduction

Hesperus is the name that the ancient Greeks gave to the Evening Star, a bright object that appears in the west after the Sun sets. Phosphorus is the name they gave to the Morning Star, which can be seen in the eastern morning sky shortly before the Sun rises. At some point, ancient civilizations learned that these two celestial objects are one and the same —the planet we now know as Venus. In the philosophy of language, the sentence "Hesperus is Phosphorus" (HiP) refers to a famous puzzle often attributed to the philosopher and mathematician Gottlob Frege. The HiP debate gained significant attention after the work of Saul Kripke in the 1970s [1] and involves discussions about designation, reference of proper names, and the nature of identity [2, 3].

Enumeration [6], but to this day researchers need to resort to purposely created tools such as AVClass [7] or Euphony [8] to harmonize AV labels.

When it comes to attributing cyber attacks to state-sponsored threat actors (TAs) —the so-called Advanced Persistent Threats (APTs)—, nomenclature is a HiP problem on steroids. CTI vendors have developed their own analysis and attribution methodologies and, more importantly, their own naming schemes and catalogs for TAs [9, 10, 11, 12]. After two decades of APT reporting by dozens of CTI shops, the current CTI landscape is fragmented and plagued with a multiplicity of catalogs for naming TAs. This problem creates the need for a resource that helps analysts to know whether the TA known as 'X' by one vendor is the same as what another vendor names 'Y' [13, 14]. These mappings



(C) BBC Sky at Night



G.H. Frezza (1704).
Lucifer (The Morning Star)



Evelyn De Morgan (1881).
Phosphorus and Hesperus.



Stanisław Wyspiański (1897).
Phosphoros, Eos, Helios, Hesperos

Attribution is hard.

Naming is harder.

APT 9

APT 15

APT 17

APT 19

APT 26

ROYAL SPIDER

AXIOM

BACKDOOR DIPLOMACY

BARIUM

BEIJING GROUP

WINNTI GROUP

BRONZE ATLAS

BRONZE EXPORT

BRONZE FIRESTONE

TITAN RAIN

PRIMITIVE BEAR

DEEP PANDA

ELDERWOOD

HIDDEN LYNX

KE3CHANG

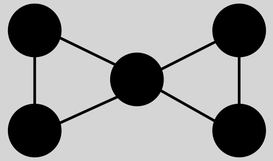
NIGHTSHADE PANDA

TURBINE PANDA

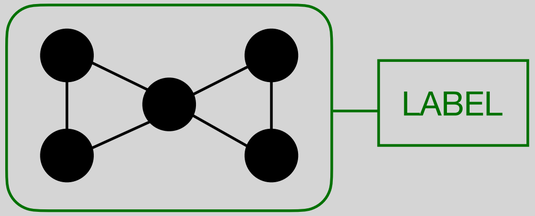
How have we come to this?

What gets named —and renamed—, and why

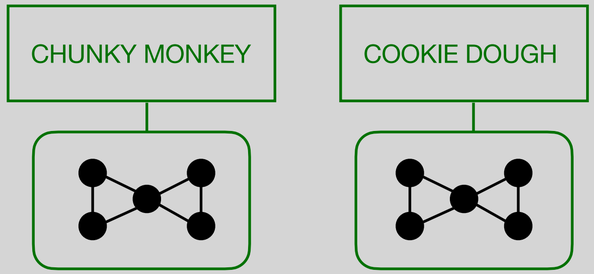
1. Analysis workflows involve **clustering** together threat activity / intrusion sets.



2. Clusters **get names** for internal cataloging and for future correlation.



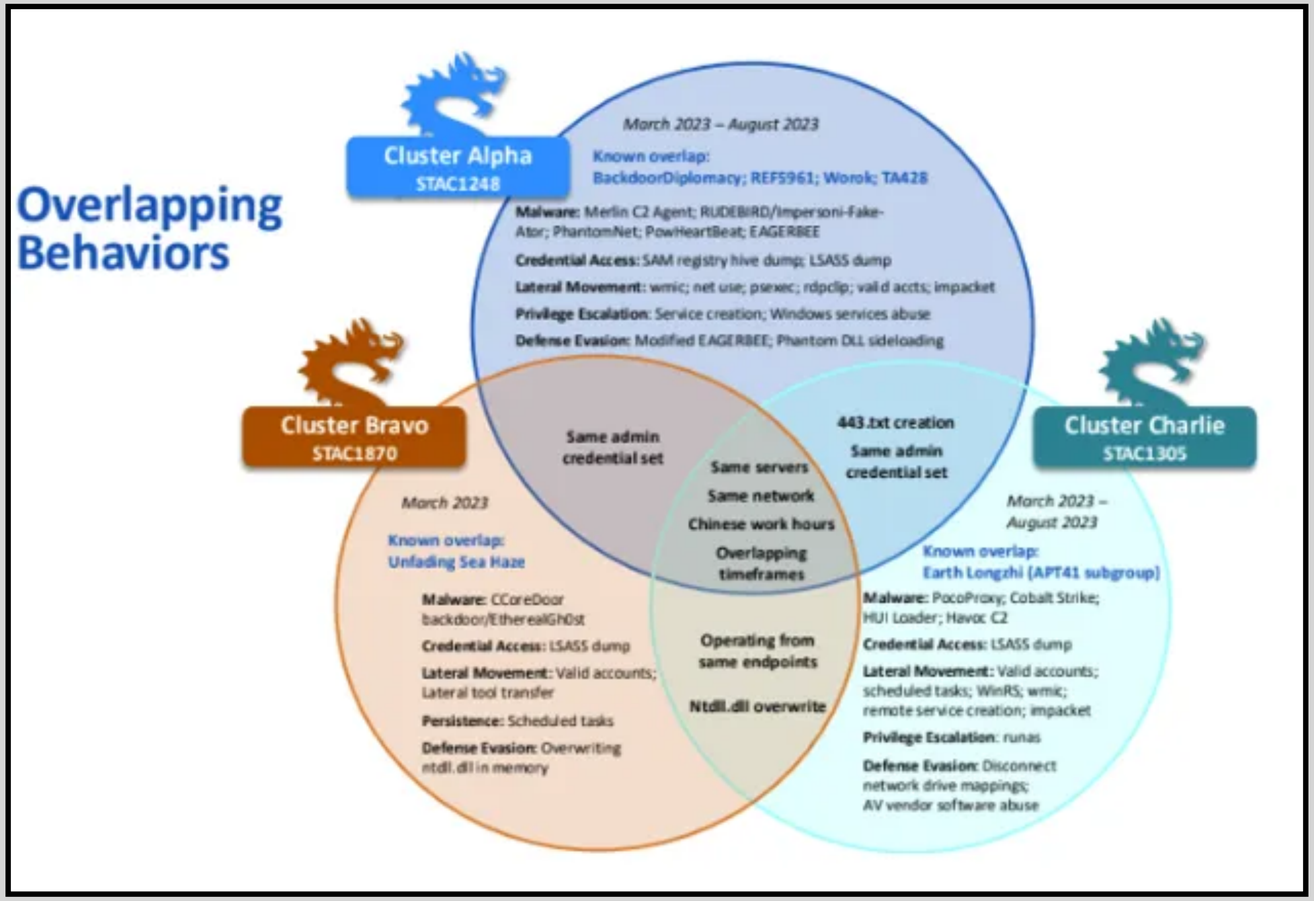
3. Because the process is approximately similar across vendors, naming schemes are roughly similar in structure, but **each vendor has one**.

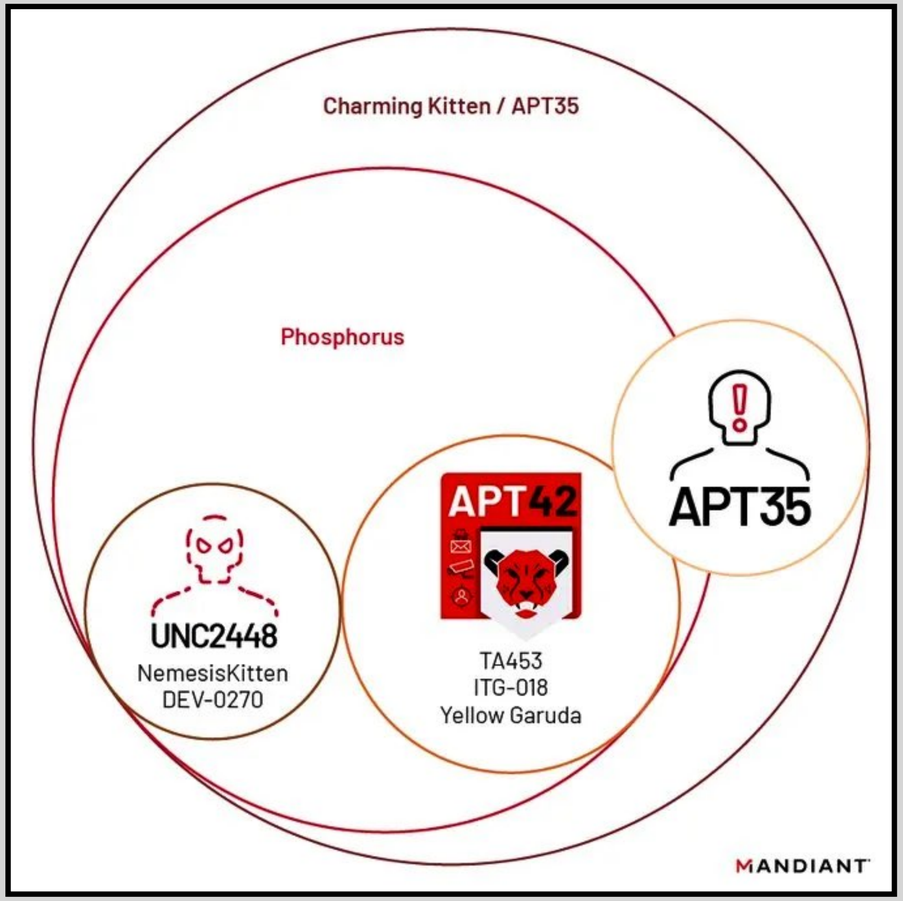


4. During **CTI dissemination** vendors use their internal names.



Overlapping intrusion sets or, what is a TA?

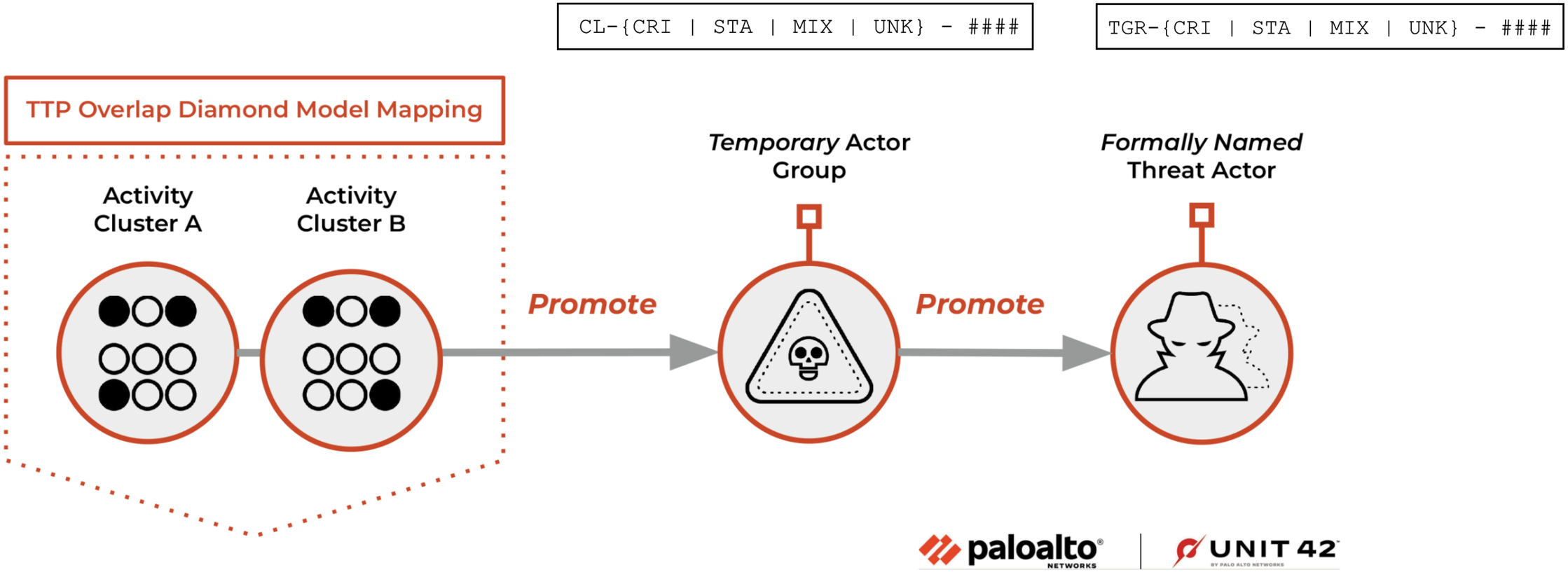




https://medium.com/@Shinigami42/what-a-cluster-how-industry-groups-and-names-threat-activity-clusters-0c65a7128162

Namespaces

Temporary names & the graduation of a TA



https://unit42.paloaltonetworks.com/from-activity-to-formal-naming/

(New) MSFT naming scheme

Blizzard	Typhoon	Sandstorm	Sleet	Dust	Gale	Derecho
Russia	**China**	**Iran**	**North Korea**	**Türkiye**	**Germany**	**Spain**
Cyclone	Rain	Hail	Whirlwind	Squall	Haze	Frost
Vietnam	**Lebanon**	**South Korea**	**Pakistan**	**Singapore**	**Syria**	**Ukraine**
Monsoon	Flood	Storm	Tempest	Tsunami	Lightning	Tornado
India	**Influence operations**	**Groups in development**	**Financially motivated**	**Private sector offensive actor**	**Palestinian Authority**	**United States**

CrowdStrike naming scheme

eCrime

LIGHTNING SPIDER
HOOK SPIDER
SINFUL SPIDER
HOLIDAY SPIDER
SAMBA SPIDER
LUNAR SPIDER
OCULAR SPIDER
RADIANT SPIDER
PLUMP SPIDER
CURLY SPIDER
CHATTY SPIDER
RENAISSANCE SPIDER
PUNK SPIDER
RECESS SPIDER

BRAIN SPIDER
SCATTERED SPIDER
AVIATOR SPIDER
PROPHET SPIDER
GRACEFUL SPIDER
SCULLY SPIDER
CLOCKWORK SPIDER
SCION SPIDER
APOTHECARY SPIDER
MASKED SPIDER
HONEY SPIDER
VICE SPIDER
SOLAR SPIDER
TRAVELING SPIDER

BUTLER SPIDER
HAZARD SPIDER
FROZEN SPIDER
VAMPIRE SPIDER
BITWISE SPIDER
NIMBLE SPIDER
COOKIE SPIDER
ROBOT SPIDER
TUNNEL SPIDER
ODYSSEY SPIDER
ROYAL SPIDER
BLIND SPIDER
INDRIK SPIDER
WIZARD SPIDER

North Korea

FAMOUS CHOLLIMA
VELVET CHOLLIMA

Russian Federation

PRIMITIVE BEAR
COZY BEAR
FANCY BEAR
GOSSAMER BEAR

Iran

SPECTRAL KITTEN
STATIC KITTEN
PULSAR KITTEN

China

TREASURE PANDA
VAULT PANDA
CASCADE PANDA
KRYPTONITE PANDA
LOCKSMITH PANDA
SUNRISE PANDA
PHANTOM PANDA
VIXEN PANDA

OVERCAST PANDA
VERTIGO PANDA
AQUATIC PANDA
CIRCUIT PANDA
MUSTANG PANDA
HORDE PANDA
OPERATOR PANDA
ENVOY PANDA

Hacktivism

BOUNTY JACKAL
CRUEL JACKAL

Egypt

WATCHFUL SPHINX

Pakistan

MYTHIC LEOPARD

MITRE ATT&CK naming scheme

Groups: 170

ID	Name	Associated Groups	Description
G0018	admin@338		admin@338 is a China-based cyber threat group. It has previously used newsworthy events as lures to deliver malware and has primarily targeted organizations involved in financial, economic, and trade policy, typically using publicly available RATs such as PoisonIvy, as well as some non-public backdoors.
G1030	Agrius	Pink Sandstorm, AMERICIUM, Agonizing Serpens, BlackShadow	Agrius is an Iranian threat actor active since 2020 notable for a series of ransomware and wiper operations in the Middle East, with an emphasis on Israeli targets. Public reporting has linked Agrius to Iran's Ministry of Intelligence and Security (MOIS).
G0130	Ajax Security Team	Operation Woolen-Goldfish, AjaxTM, Rocket Kitten, Flying Kitten, Operation Saffron Rose	Ajax Security Team is a group that has been active since at least 2010 and believed to be operating out of Iran. By 2014 Ajax Security Team transitioned from website defacement operations to malware-based cyber espionage campaigns targeting the US defense industrial base and Iranian users of anti-censorship technologies.
G1024	Akira	GOLD SAHARA, PUNK SPIDER, Howling Scorpius	Akira is a ransomware variant and ransomware deployment entity active since at least March 2023. Akira uses compromised credentials to access single-factor external access mechanisms such as VPNs for initial access, then various publicly-available tools and techniques for lateral movement. Akira operations are associated with "double extortion" ransomware activity, where data is exfiltrated from victim environments prior to encryption, with threats to publish files if a ransom is not paid. Technical analysis of Akira ransomware indicates variants capable of targeting Windows or VMWare ESXi hypervisors and multiple overlaps with Conti ransomware.

Mappings

Lists of aliases

- Ideally mappings between two known namespaces:

$$\mathbf{taxon_A{:}name_1 = taxon_B{:}name_2}$$

- But in reality is more like

$$\mathbf{taxon_A{:}name_1 = [name_2, name_3, name_4]}$$

It's up to you to figure out where those names (aliases) come from.

Exhibit 1: MSFT mapping


```
{
        "Threat actor name": "Diamond Sleet",
        "Origin/Threat": "North Korea",
        "Other names": "ZINC, LABYRINTH CHOLLIMA, Black Artemis, Lazarus"
},
{
        "Threat actor name": "Emerald Sleet",
        "Origin/Threat": "North Korea",
        "Other names": "THALLIUM, VELVET CHOLLIMA, RGB-D5, Black Banshee, Kimsuky, Greendinosa"
},
{
        "Threat actor name": "Fallow Squall",
        "Origin/Threat": "Singapore",
        "Other names": "PLATINUM, PARASITE, RUBYVINE, GINGERSNAP"
},
{
        "Threat actor name": "Flax Typhoon",
        "Origin/Threat": "China",
        "Other names": "Storm-0919, ETHEREAL PANDA"
},
{
        "Threat actor name": "Forest Blizzard",
        "Origin/Threat": "Russia",
        "Other names": "STRONTIUM, FANCY BEAR, Sednit, ATG2, Sofacy, Blue Athena, Z-Lom Team, Operation Pawn Storm, Tsar Team, CrisisFour, HELLFIRE, APT28"
},
{
        "Threat actor name": "Ghost Blizzard",
        "Origin/Threat": "Russia",
        "Other names": "BROMINE, BERSERK BEAR, TG-4192, Koala Team, Blue Kraken, Crouching Yeti, Dragonfly"
},
{
        "Threat actor name": "Gingham Typhoon",
        "Origin/Threat": "China",
        "Other names": "GADOLINIUM, KRYPTONITE PANDA, TEMP.Periscope, Leviathan, JJDoor, APT40, Feverdream"
},
```

Exhibit 2: MISP galaxy


```
 "values": [
    {
      "description": "PLA Unit 61398 (Chinese: 61398部队, Pinyin: 61398 bùduì) is the
Military Unit Cover Designator …"
      …
      "synonyms": [
         "COMMENT PANDA",
         "PLA Unit 61398",
         "Comment Crew",
         "Byzantine Candor",
         "Group 3",
         "TG-8223",
         "Comment Group",
         "Brown Fox",
         "GIF89a",
         "ShadyRAT",
         "G0006"
      ]
     …
      "value": "APT1"
    },
```

Issues with existing mappings

- **Namespace not always available for all aliases.**
  - It's up to you to find out where each alias comes from.
  - Sometimes you can only say UNK.

- **Often not reciprocal.**
  - Source A says **A:X** has **B:Y** as an alias.
  - Source B does not say **B:Y** has **A:X** an alias.

- **Sources may:**
  - Change their naming scheme.
  - Split a TA into multiple subgroups, or vice versa.
  - Use different syntax conventions.

THE NAMING MESS

RISKY BUSINESS NEWS

Risky Bulletin: Security firms will attempt to clean up their own mess—APT name taxonomies

In other news: NSO asks for new WhatsApp trial; EMR-ISAC shuts down; zero-days patched in Chrome and Qualcomm chips.



Catalin Cimpanu
04 Jun 2025 — 14 min read

Ransomware, Threat Intelligence

Microsoft, CrowdStrike look to harmonize threat actor terminology

June 2, 2025

By Shaun Nichols



Threat actor naming conventions: a big mess!

Jul 17, 2023

The task of adversary attribution and how it should be done is a controversial topic in the *cyber threat intelligence* community. This is because there is no standardized way of doing attribution, starting with naming the threat actors.

Each security company has its own telemetry, data, standards, procedures and confidence levels. This is the main reason why most CTI teams use their own naming scheme.

WIRED SECURITY POLITICS THE BIG STORY BUSINESS SCIENCE MORE SIGN IN

ANDY GREENBERG SECURITY APR 23, 2023 7:00 AM

Hacker Group Names Are Now Absurdly Out of Control

Pumpkin Sandstorm. Spandex Tempest. Charming Kitten. Is this really how we want to name the hackers wreaking havoc worldwide?



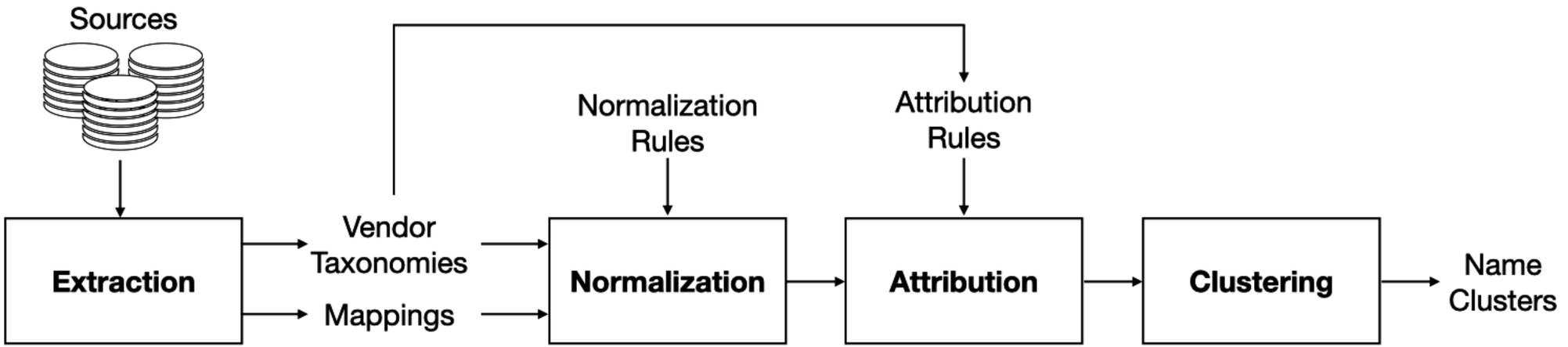
ILLUSTRATION: JACQUI VANLIEW; GETTY IMAGES

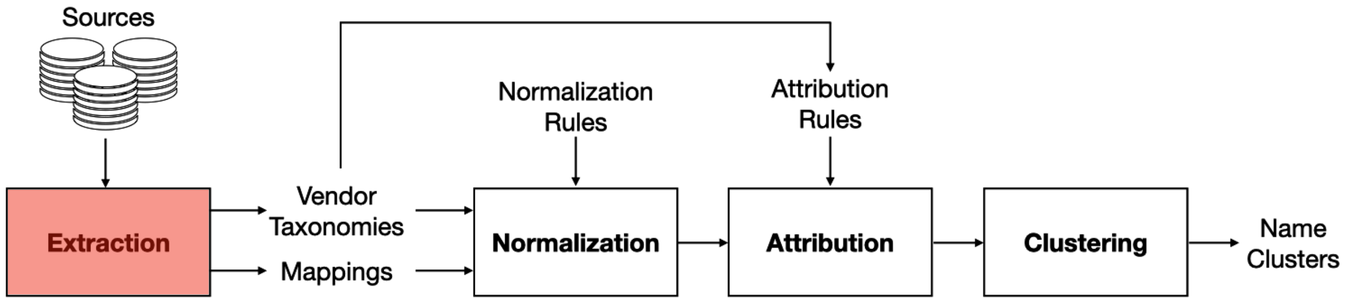
Let’s do something about it.

Develop a tool to defragment TA namespaces, integrating as many existing mappings as possible.

Then use it to gain a more empirical understanding of the problem.

HiP

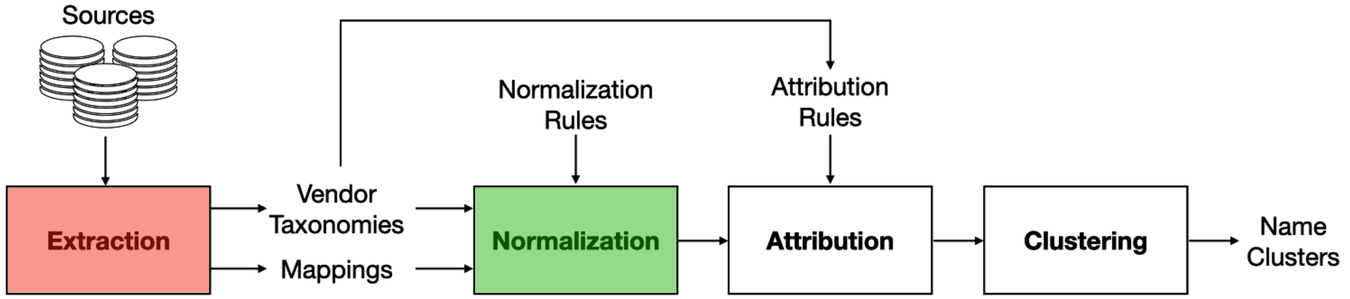

Sources
Extraction
Vendor
Taxonomies
Mappings
Normalization
Rules
Normalization
Attribution
Rules
Attribution
Clustering
Name
Clusters



Source	Description
APTGO [19]	The APT Groups and Operations is a community-driven knowledge base compiling multiple vendor taxonomies and mappings between them.
Malpedia [20]	Repository of malware families and related CTI maintained by Fraunhofer FKIE. It offers standardized metadata, including TA aliases, associations to malware families, and links to reports.
Microsoft Mapping [12]	Microsoft's official description of its TA taxonomy and some known aliases for each threat actor.
MISP Galaxy Clusters [21]	Multiple clusters contained in the MISP Galaxy focused on TA names and their aliases.
ORKL [22]	Large collection of open-source CTI reports annotated with relevant metadata, including TA names and aliases. ORKL aggregates data from 11 distinct sources, including MITRE ATT&CK data [23], APTnotes [24], CyberMonitor's APT & Cybercriminals Campaign Collection [25], ETDA's Threat Group Cards [26], Alienvault's Open Threat Exchange (OTX) [27], Secureworks Threat Profiles [28], ORKL Community Contributed Content [22], and VX Underground [29]

Mapping	Source	Size	Avg.	Max.
APTGO	APTGO [19]	237	4.24	26
Paloalto	Paloalto [18]	50	4.34	17
Malpedia	Malpedia [20]	314	3.85	41
ETDA	ORKL [22]	270	4.98	46
MITRE	ORKL [22]	175	2.93	15
Secureworks	ORKL [22]	106	4.75	17
MISP Galaxy	MISP Galaxy [21]	341	3.55	39
Microsoft	Microsoft [12]	114	3.64	14

Vendor	Taxonomy size
UNK	1,980
CrowdStrike	241
Microsoft	200
Secureworks	193
Mandiant	155
Microsoft (old)	123
MITRE	110
Palo Alto Unit 42	71
Thales	47
CERT-UA	35
Recorded Future	32
360	27
FireEye	19
Kaspersky	17
Symantec	14
Cisco Talos	14
Tencent	7
NSA	2
Total	3,287

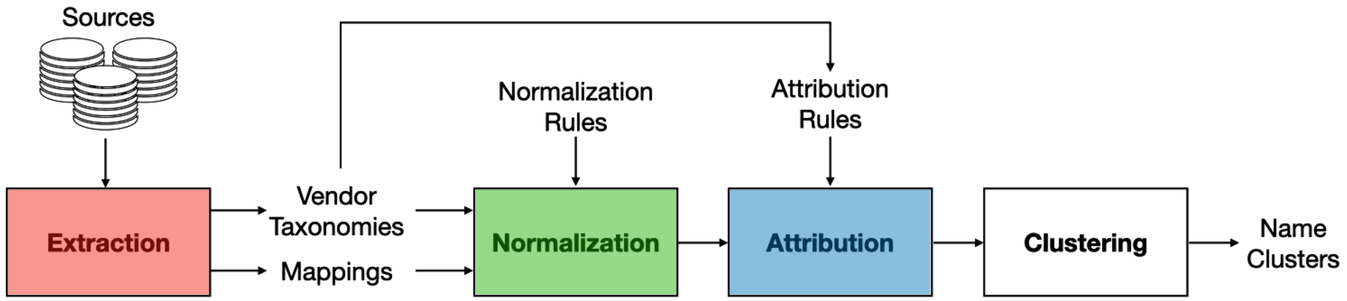


1	**Misspellings**	Agoniznig Serpens vs. Agonizing Serpens Kimsuki vs. Kimsuky Red Bald Knight vs. Red Bald Night
2	**Group variants**	Black Energy vs. Black Energy Group vs. Black Energy (Group) Budminer vs. Budminer Cyberespionage Group Callisto vs. Callisto Group
3	**Prefixes**	The Gorgon Group vs. Gorgon Group The Lamberts vs. Lamberts The Mask vs. Mask
4	**Vendor suffixes**	APTC38 vs. APTC38 (Qianxin) Earth Preta vs. Earth Preta (Trendmicro) Hive 0081 vs. Hive 0081 (IBM)
5	**Cases and separators**	scarred_manticore vs. Scarred Manticore apt32 vs. APT 32 vs. APT-32 scarleteel vs. SCARLETEEL vs. ScarletEel

Syntax normalization

1. Remove all spaces, dashes, underscore and slash symbols.
2. No prefixes or suffixes
3. Convert to upper case

APT32, BLACKENERGY
HIVE0081, THELAMBERTS



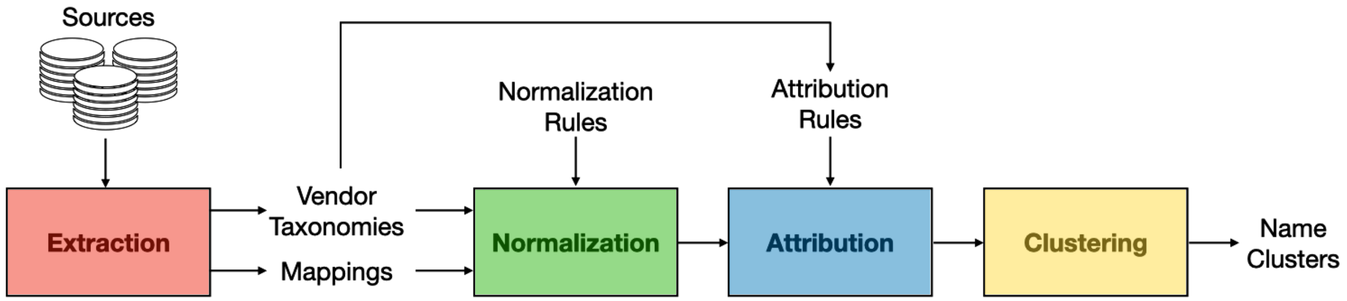
From **NAME** to **TAXONOMY:NAME**

1. Reverse search & matching
2. Special taxonomy (UNK) for unmatched names

TRENDMICRO:EARTHPRETA

CROWDSRTRIKE:COZYBEAR

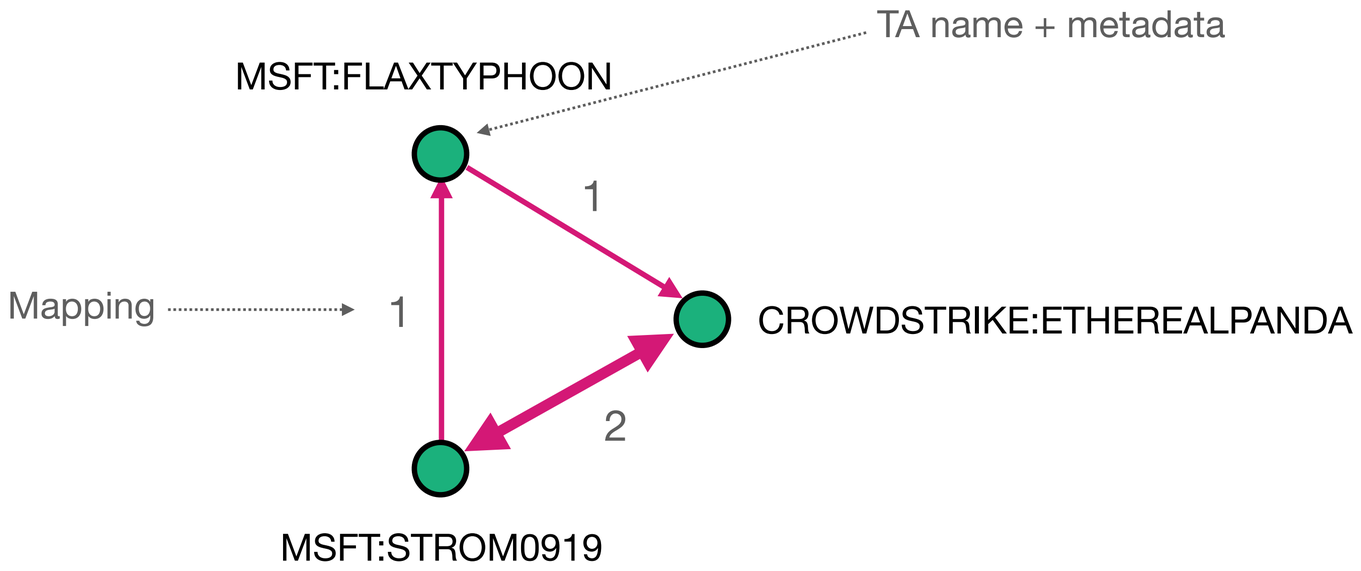
Limited to accuracy of information in sources, but errors are easy to fix manually

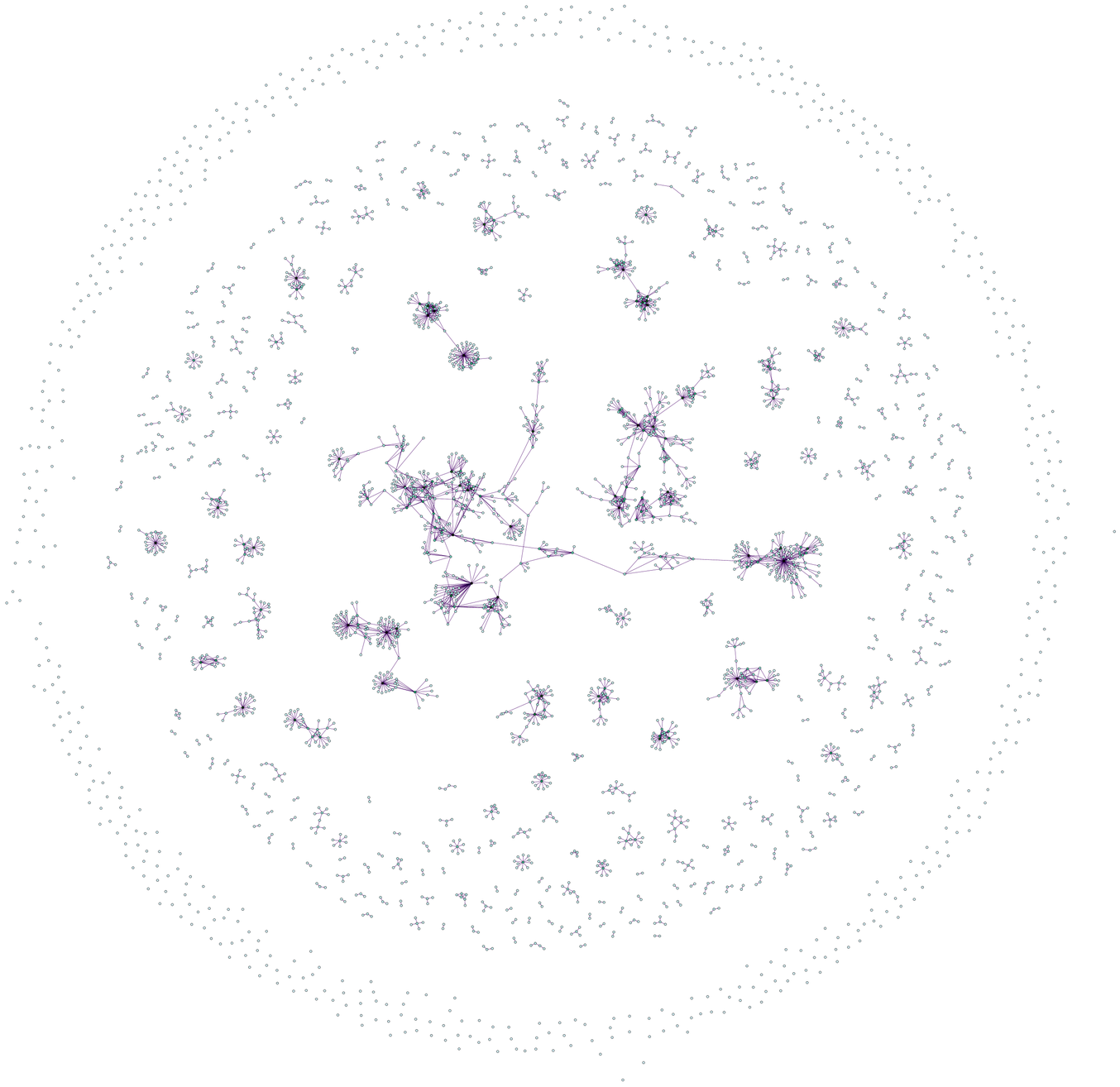


Threat Actor Name Alias Graph (TANAG)

```
{
  “source” = “MSTIC”
  “ta_canonical_name” = “Flax Typhoon”,
  “ta_aliases” = [“Storm-0919”, “ETHEREAL PANDA”],
  “country” = “China”,
  “reports” = [“url 1”, “url 2”, …],
  “tools” = [“tool 1”, “tool 2”, …],
   …
}
```

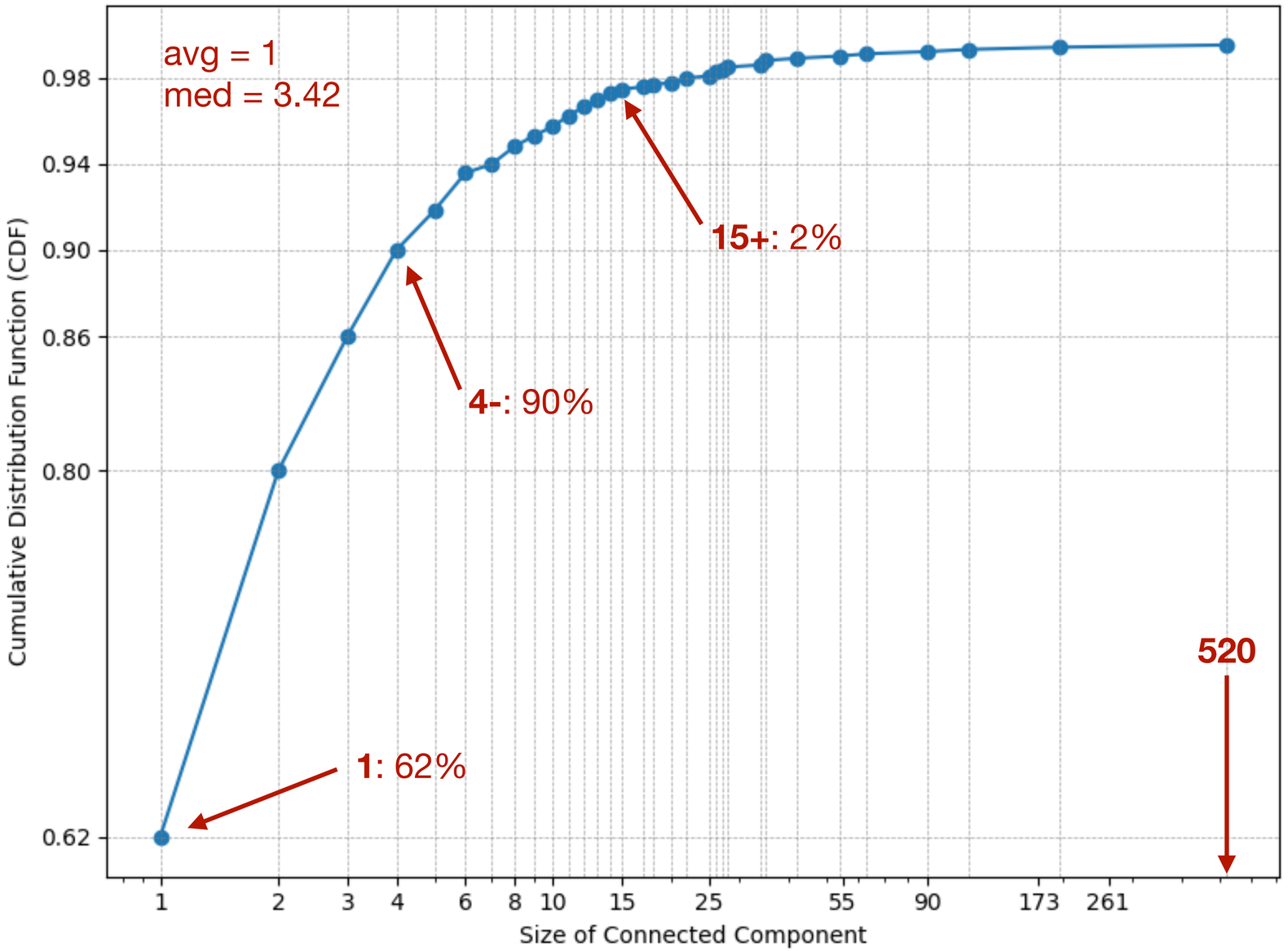
```
{
  “source” = “Malpedia”
  “ta_canonical_name” = “Ethereal Panda”,
  “ta_aliases” = [“Storm-0919”],
  “country” = “China”,
  “reports” = [“url 1”, “url 2”, …],
  “tools” = [“tool 1”, “tool 2”, …],
   …
}
```





Interactive explorer

Let’s measure it.

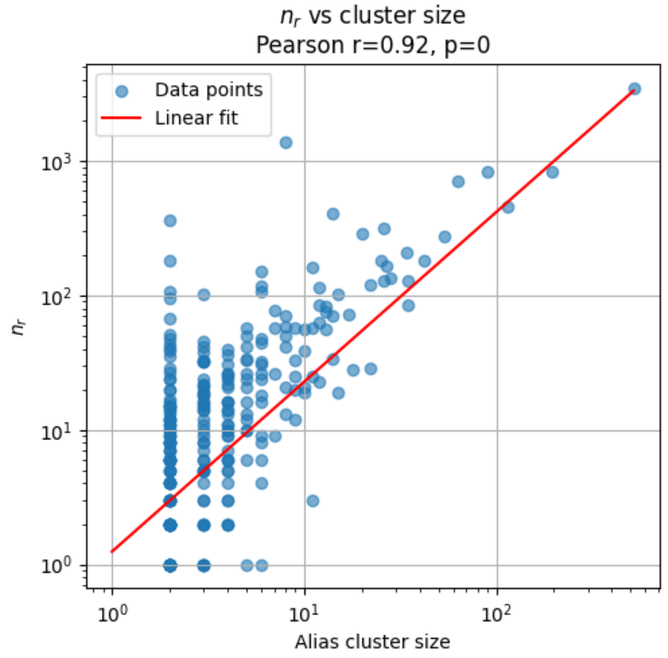

avg = 1
med = 3.42
15+: 2%
4-: 90%
1: 62%
520
0.98
0.94
0.90
0.86
0.80
0.62
Cumulative Distribution Function (CDF)
1
2
3
4
6
8
10
15
25
55
90
173
261
Size of Connected Component

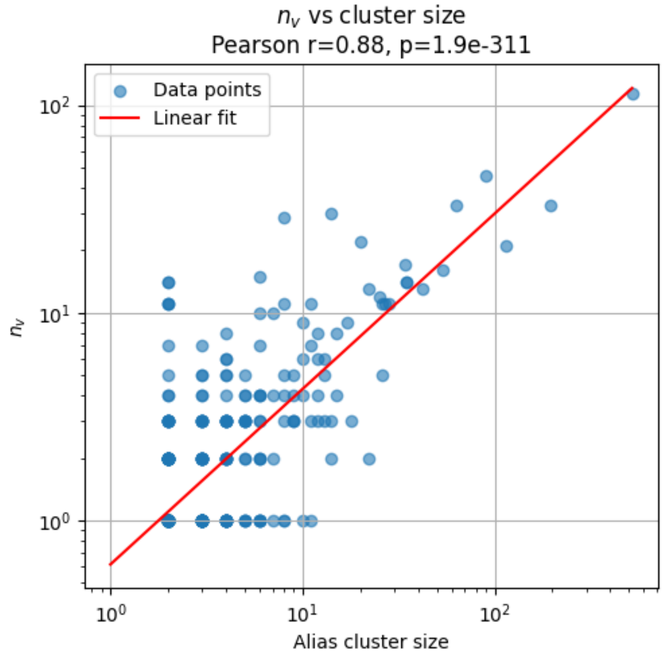
Name proliferation focused on a very reduced subset of known TAs. Why?

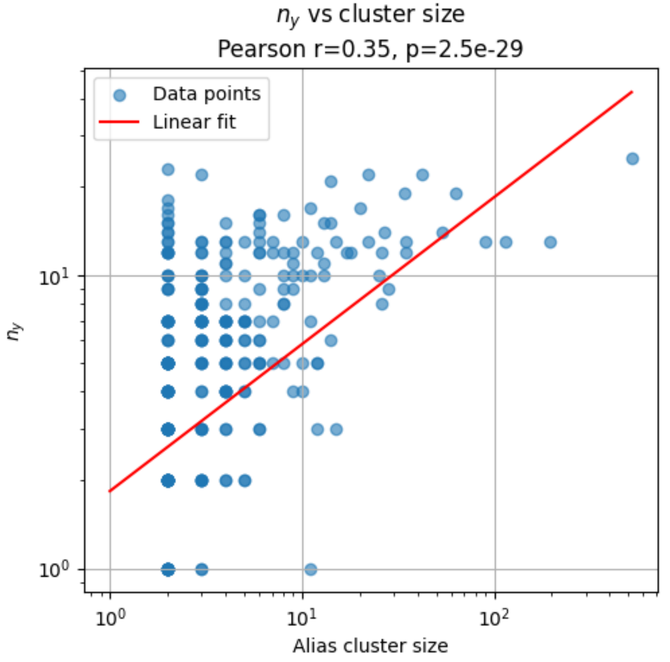
Correlation analysis

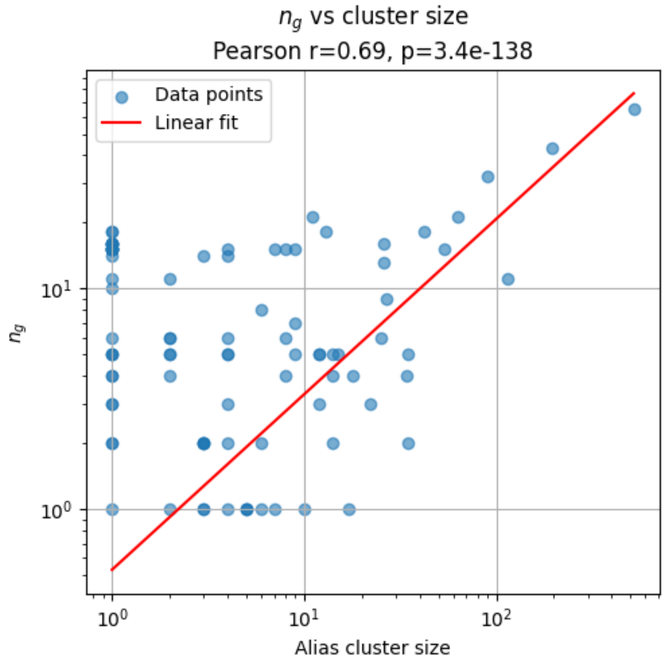
- n_r = Number of CTI reports in which any of the aliases in the cluster appear.
- n_v = Number of CTI vendors that have published a CTI report about the TA.
- n_y = Number of years that the TA has been studied (by any CTI vendor).
- n_g = Number of different geographies or world regions targeted by the TA.
- n_s = Number of victim sectors targeted by the TA.

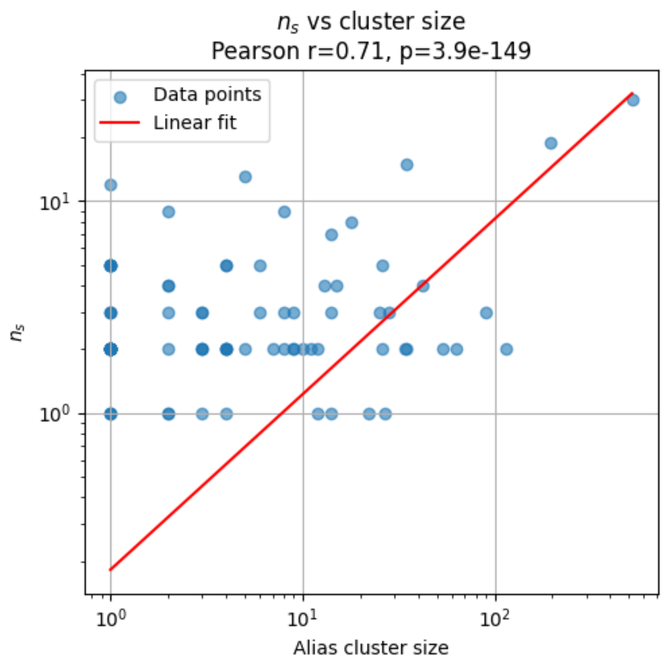
TA feature	Pearson	Spearman	Kendall	Interpretation
n_r	0.92 (p=0)	0.83 (p=6.59e-249)	0.77 (p=5.72e-163)	Very strong, highly linear
n_v	0.88 (p=1.94e-311)	0.83 (p=1.13e-246)	0.77 (p=1.11e-159)	Very strong, highly linear
n_y	0.35 (p=2.47e-29)	0.82 (p=3.40e-231)	0.75 (p=1.24e-151)	Strong, non-linear
n_g	0.70 (p=3.41e-138)	0.29 (p=1.66e-19)	0.27 (p=3.20e-19)	Inconclusive
n_s	0.71 (p=3.86e-149)	0.27 (p=2.27e-17)	0.25 (p=7.06e-17)	Inconclusive


n_r vs cluster size
Pearson r=0.92, p=0
Data points
Linear fit
n_r
Alias cluster size


n_v vs cluster size
Pearson r=0.88, p=1.9e-311
Data points
Linear fit
n_v
Alias cluster size


n_y vs cluster size
Pearson r=0.35, p=2.5e-29
Data points
Linear fit
n_y
Alias cluster size


n_g vs cluster size
Pearson r=0.69, p=3.4e-138
Data points
Linear fit
n_g
Alias cluster size

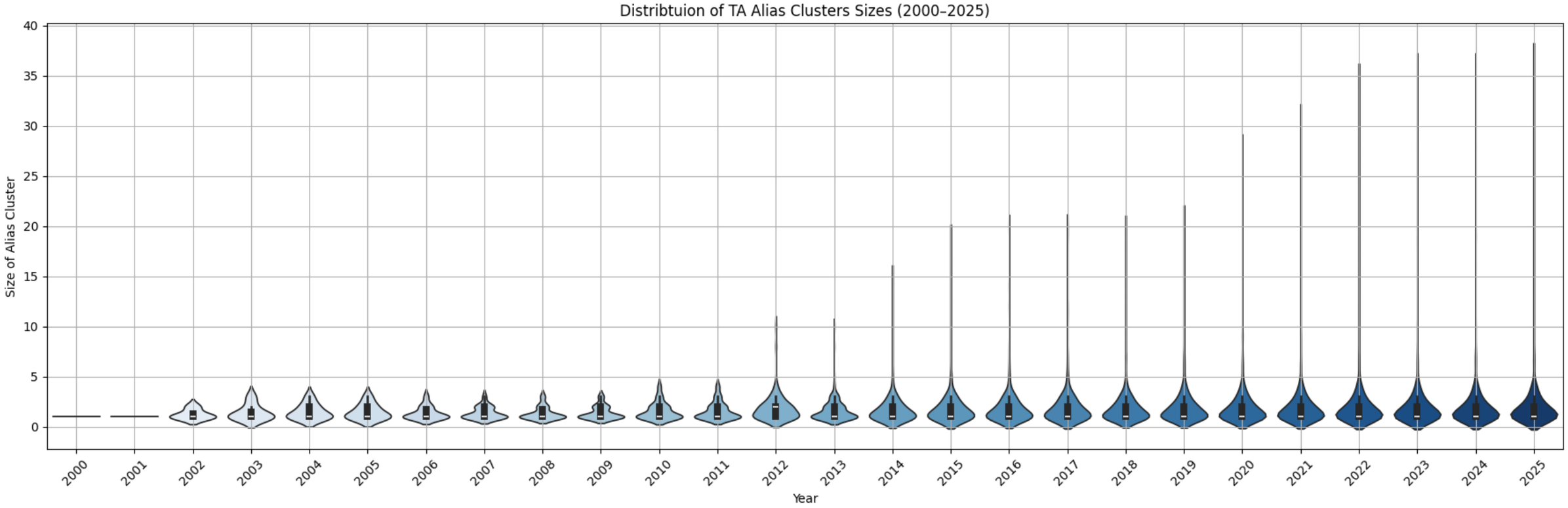

n_s vs cluster size
Pearson r=0.71, p=3.9e-149
Data points
Linear fit
n_s
Alias cluster size



Correlation
does
NOT
imply
causality

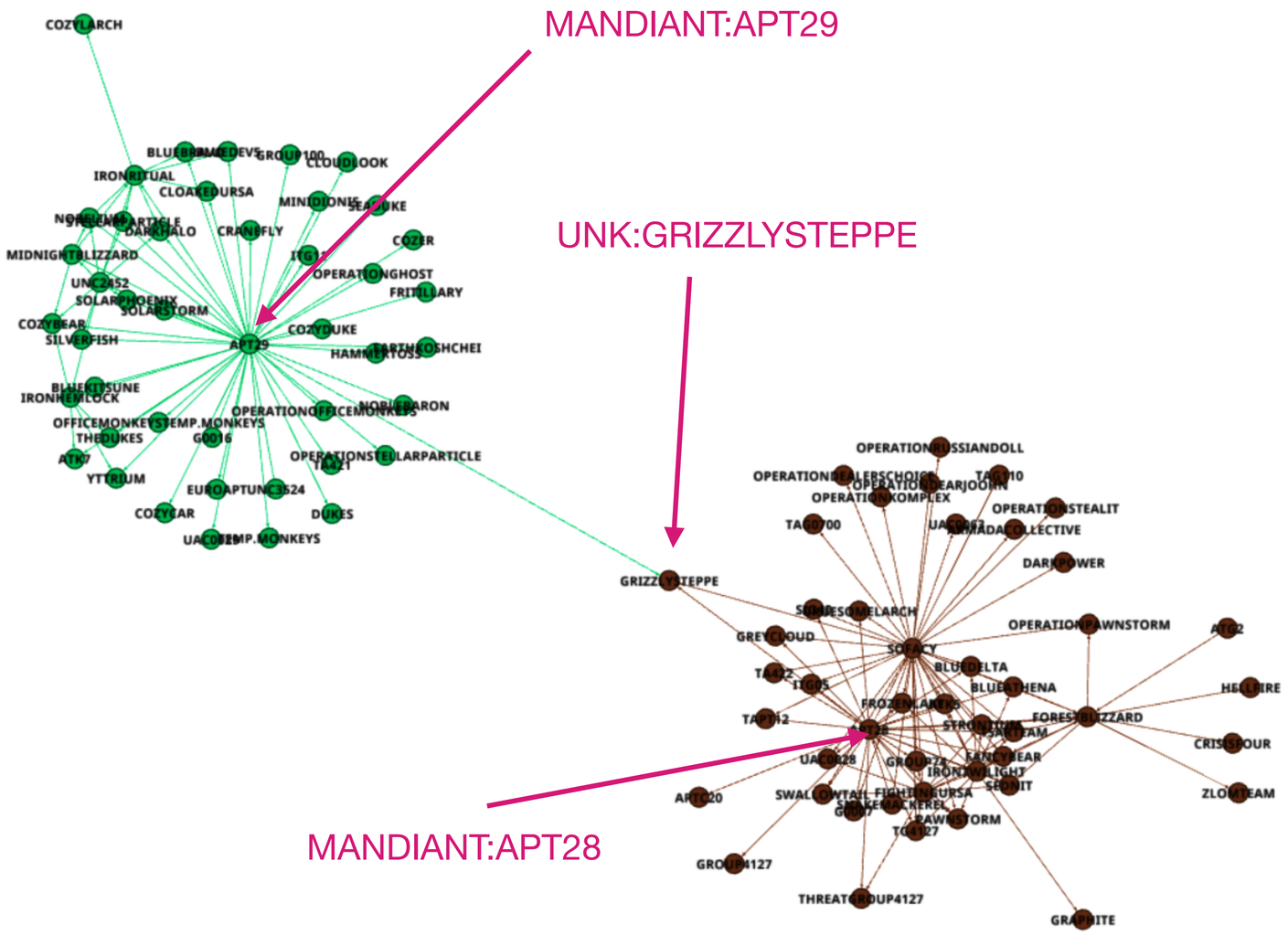
Evolution

Only for a subset of names due to limitations in available temporal data



Maybe the methodology is okay but, do we trust the data?

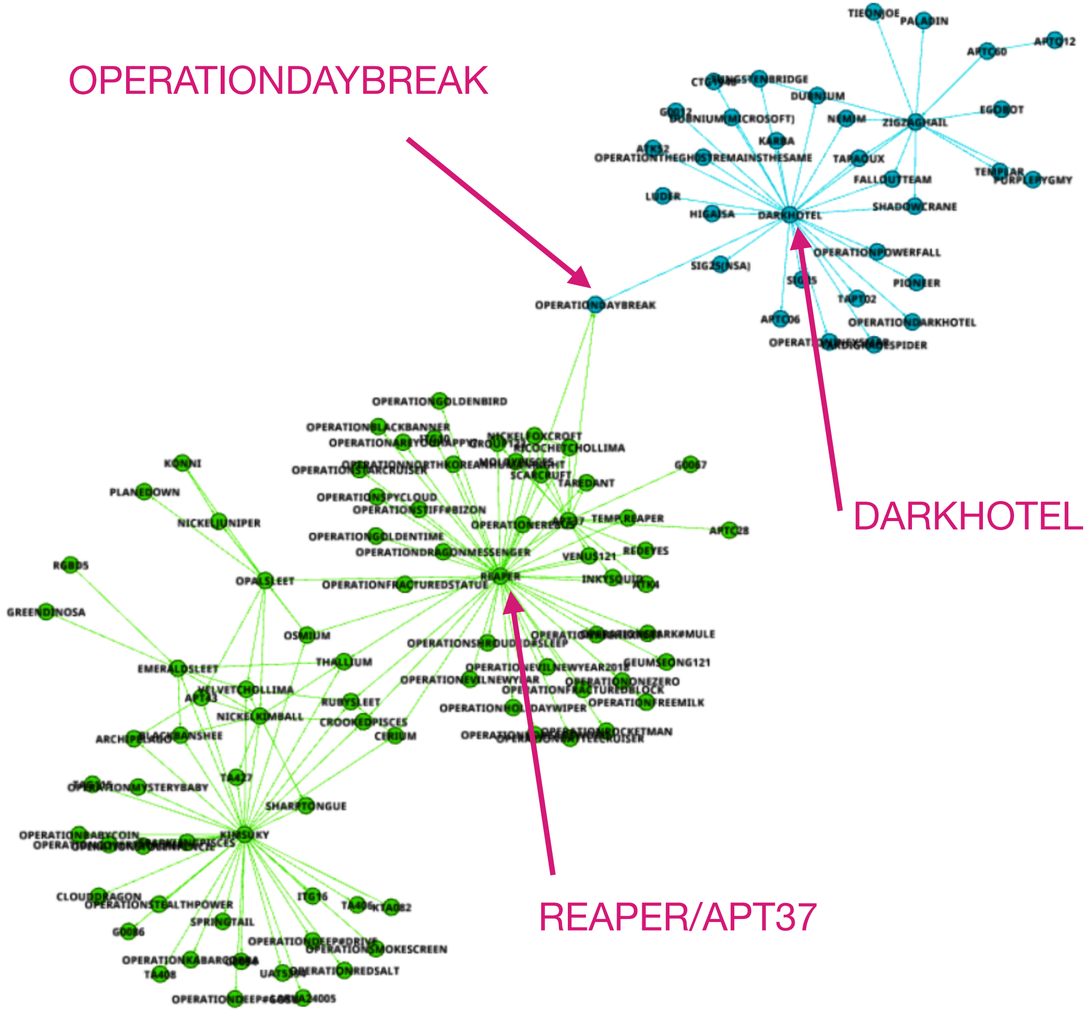
Sloppy mappings: The case of Grizzly Steppe and APT29



MISP maps APT28 APT29, and SOFACY to GRIZZLYSTEPPE

> *"The U.S. Government confirms that two different RIS actors participated in the intrusion into a U.S. political party. The first actor group, known as Advanced Persistent Threat (APT) 29, entered into the party's systems in summer 2015, while the second, known as APT28, entered in spring 2016"* **[CISA, 2016]**

Name collisions: The Two Operations Daybreak



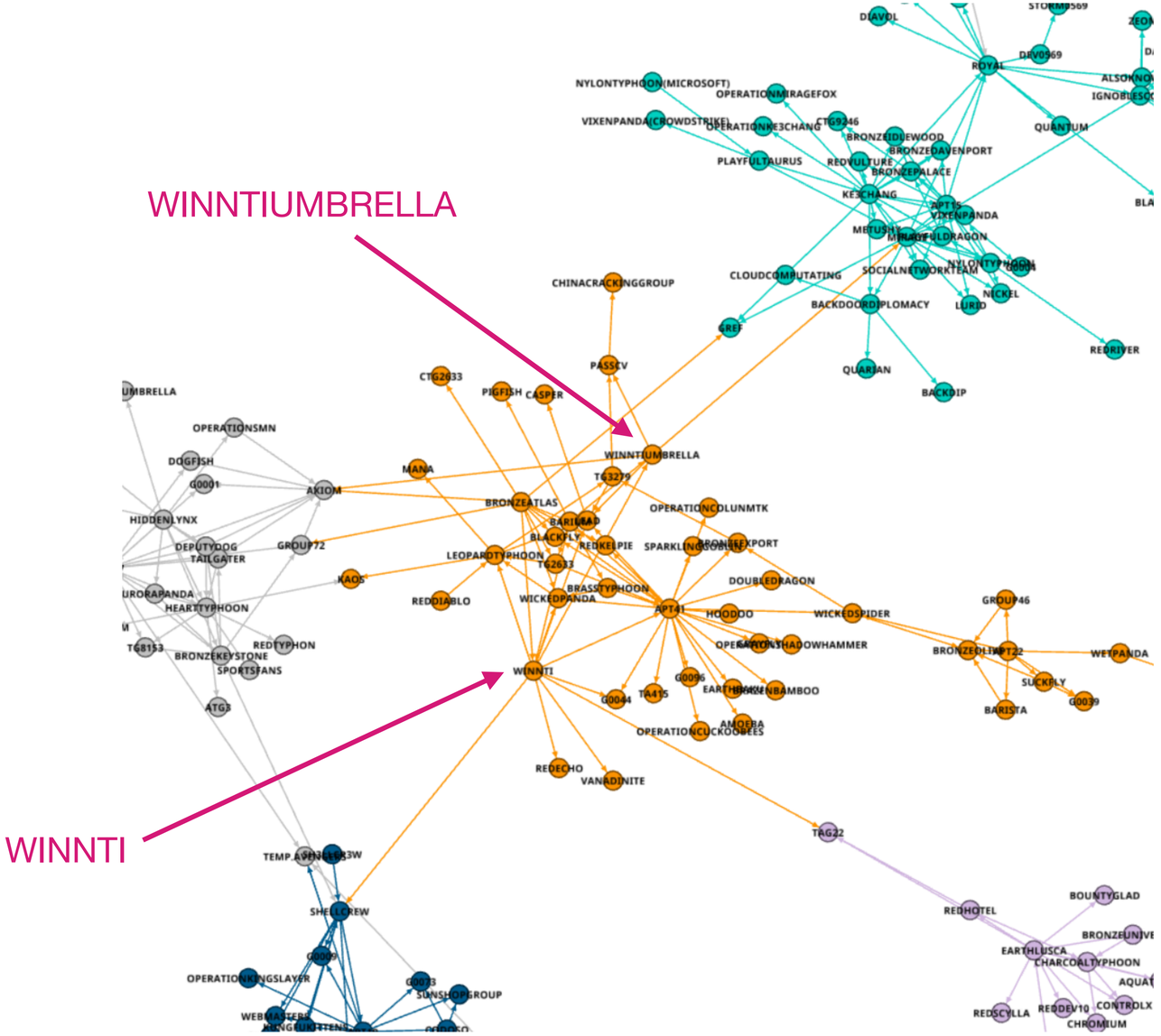
> *"Operation Daybreak Method: Uses Flash zero-day exploit for CVE-2015-8651. Note: not the same operation as Reaper, APT 37, Ricochet Chollima, ScarCruft's Operation Daybreak"*
> **[ETDA's DarkHotel profile, 2015]**

> *"Operation Daybreak. Target: High profile victims. Method: Previously unknown (0-day) Adobe Flash Player exploit. It is also possible that the group deployed another zero day exploit, CVE-2016-0147, which was patched in April. <https://securelist.com/operation-daybreak/75100/> Note: not the same operation as DarkHotel's Operation Daybreak"*
> **[ETDA's Reaper/APT37 profile, 2016]**

Tools vs. Groups, or Winnti is not a TA

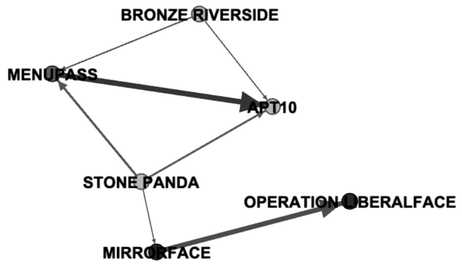
WINNTI: Trojan/RAT family heavily used by a number of Chinese APT groups. Initially (2011) used to target the gaming industry and later found in in attacks targeting organizations in other sectors.

Activity initially attributed to the Winnti group was later shown to involve multiple TAs reusing common tooling.

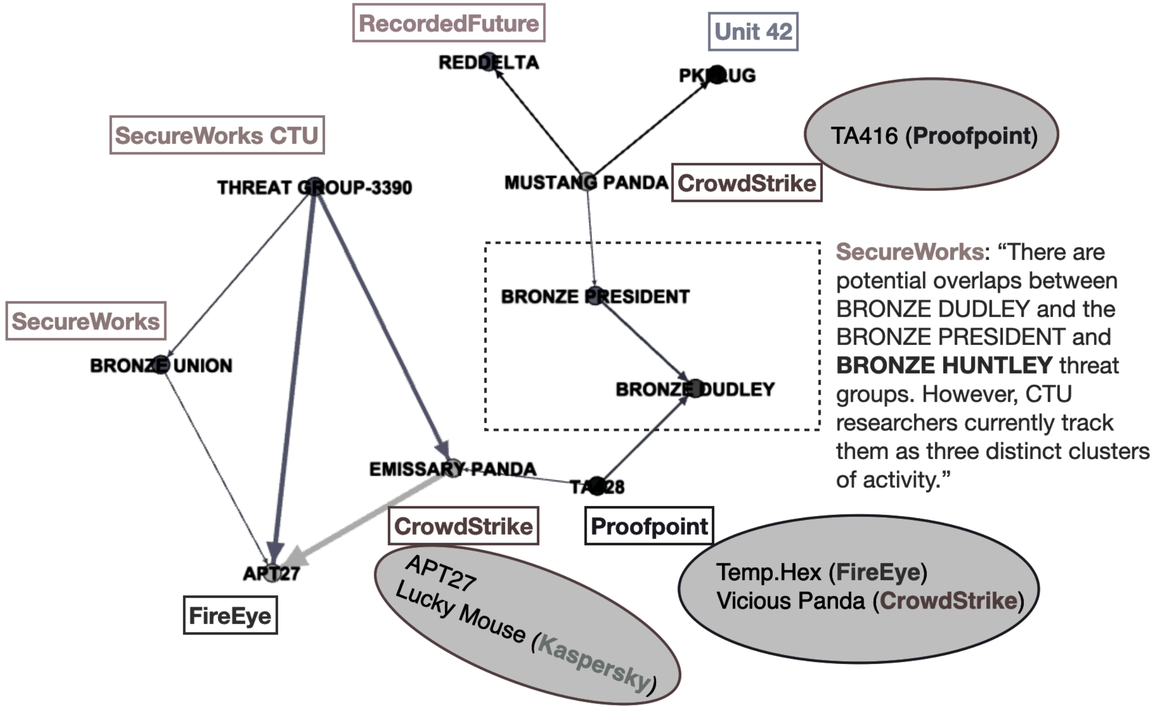


More insights

- Names conflated with operations
- Large clusters, weak associations
- Persistence of temporal names
- Different approaches to attribution



*"ESET researchers discovered a **spearphishing campaign**, launched in the weeks leading up to the Japanese House of Councillors election in July 2022, **by the APT group that ESET Research tracks as MirrorFace**. The campaign, which we have named **Operation LiberalFace**, targeted Japanese political entities"*



SecureWorks: "There are potential overlaps between BRONZE DUDLEY and the BRONZE PRESIDENT and **BRONZE HUNTLEY** threat groups. However, CTU researchers currently track them as three distinct clusters of activity."

Directions

MISP RFC

https://www.misp-standard.org/rfc/threat-actor-naming.html

Workgroup:	Network Working Group
Internet-Draft:	draft-00
Published:	21 December 2024
Intended Status:	Informational
Expires:	24 June 2025

Authors:	A. Dulaunoy	P. Bourmeau
	CIRCL	*Cubessa*

Recommendations on Naming Threat Actors

Abstract

This document provides advice on the naming of threat actors (also known as malicious actors). The objective is to provide practical advice for organizations such as security vendors or organizations attributing incidents to a group of threat actors. It also discusses the implications of naming a threat actor for intelligence analysts and threat intelligence platforms such as MISP [MISP-P].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Table of Contents ▲

1. Introduction
   1.1. Conventions and Terminology
2. Recommendations
   2.1. Reusing Threat Actor Names
   2.2. Uniqueness
   2.3. Format
   2.4. Encoding
   2.5. Avoid Confusing Actor Names with Malware Names
   2.6. Directory
3. Examples
4. Security Considerations
5. Acknowledgements
6. References
7. References
   7.1. Normative References
   7.2. Informative References

Authors' Addresses

It is unlikely to work

- Remember CARO
    - Syntax rules may be adopted or may be inspirational
    - But 2 vendors won't be able to agree on whether they are referring to the same TA **unless they share raw intelligence** (telemetry, samples)
    - **Systematically**, not on a case-by-case basis
- Are they willing to do it?
- Are they allowed to do it?

Naming is hard.

Correction:
Adopting a naming convention will be hard.

In the meantime, we need better mappings …



MALWARE & THREATS

Microsoft, CrowdStrike Lead Effort to Map Threat Actor Names

Microsoft and CrowdStrike are running a project that aims to align threat actor names, and Google and Palo Alto Networks will also contribute.

By Eduard Kovacs | June 3, 2025 (4:25 AM ET)

… and HiP :-)

FIN

Thank you for listening.


jestevez@inf.uc3m.es


0xjet.github.io


@0xjet