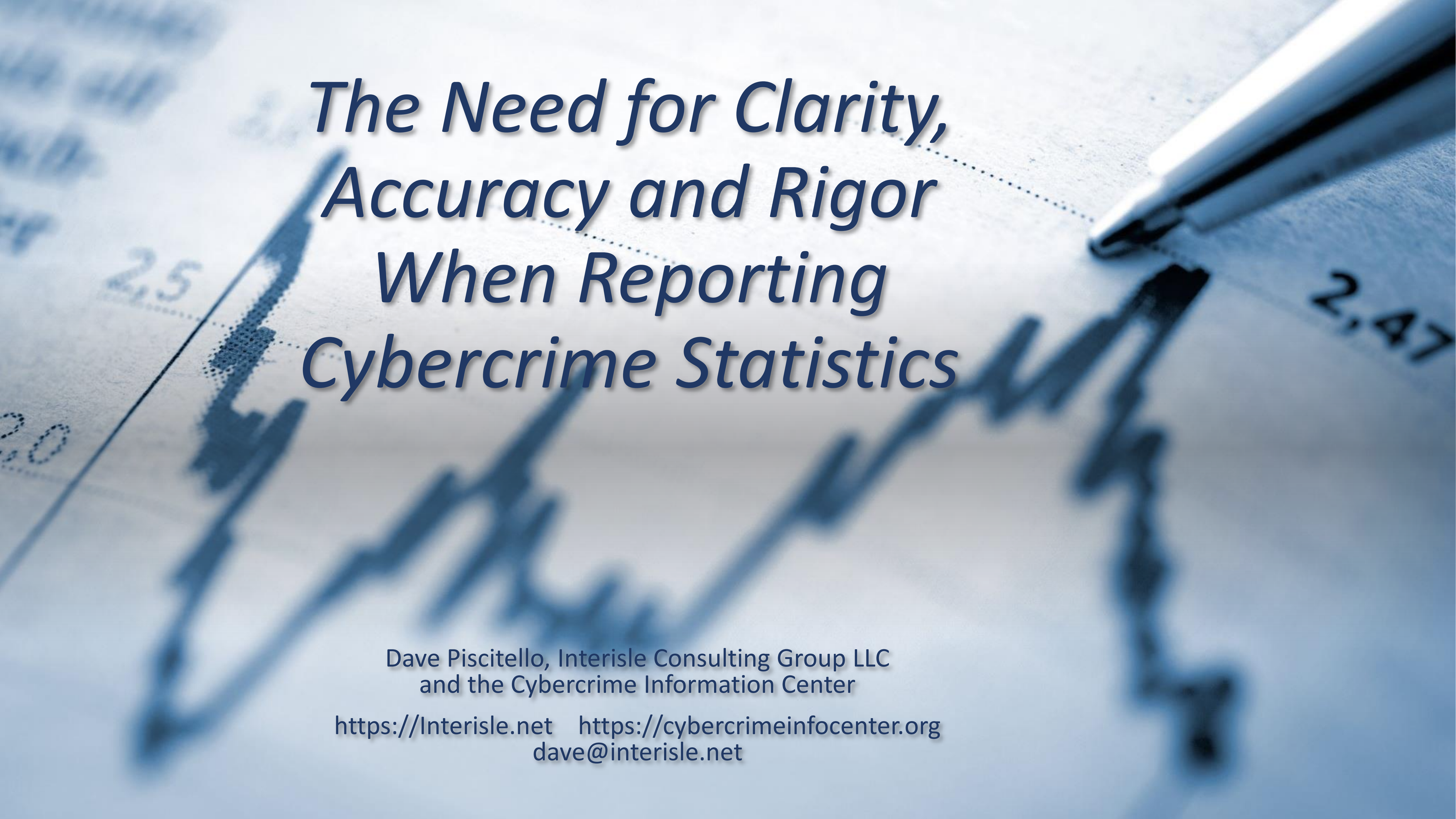


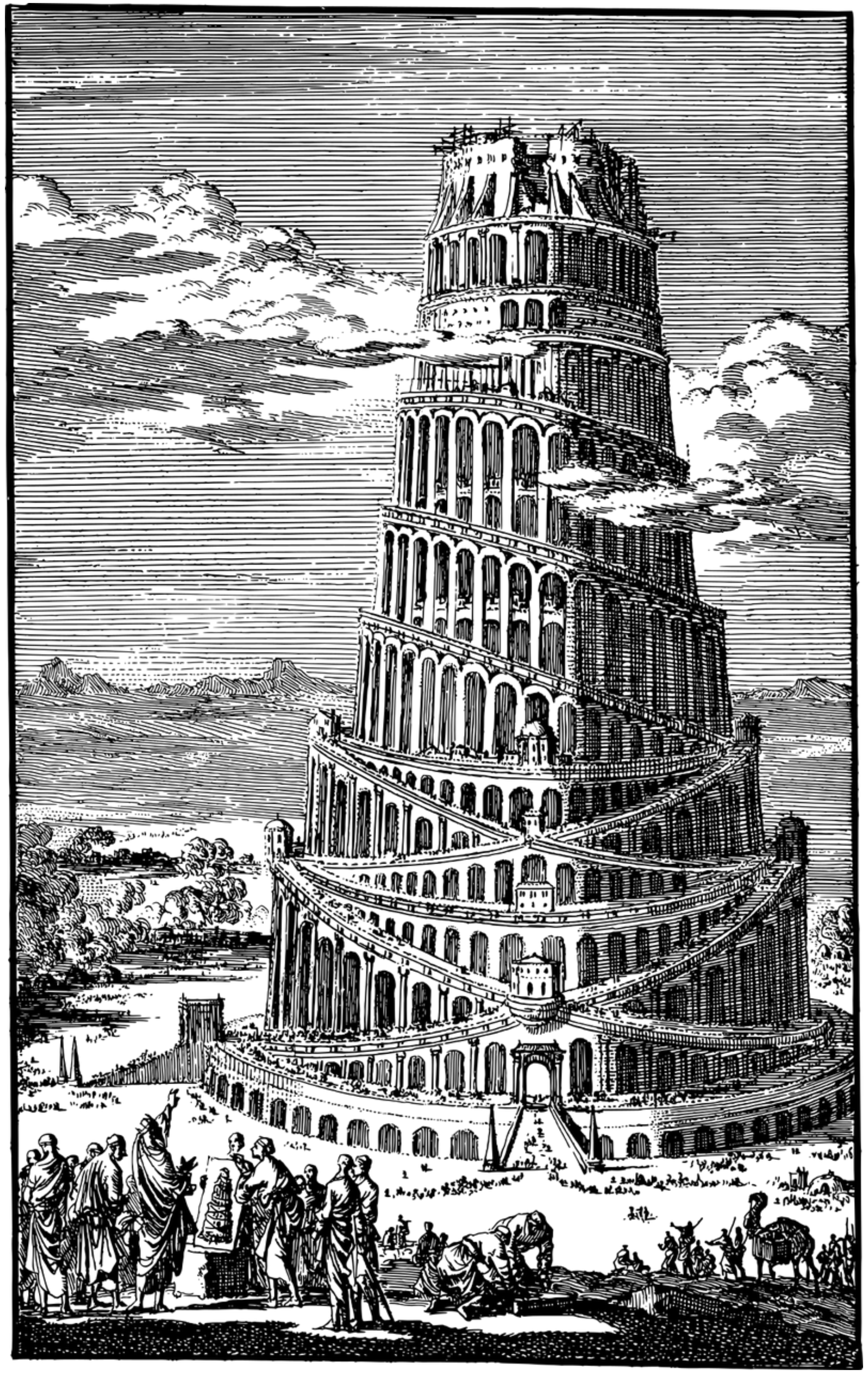
The Need for Clarity, Accuracy and Rigor When Reporting Cybercrime Statistics

Dave Piscitello, Interisle Consulting Group LLC
and the Cybercrime Information Center

https://Interisle.net   https://cybercrimeinfocenter.org
dave@interisle.net

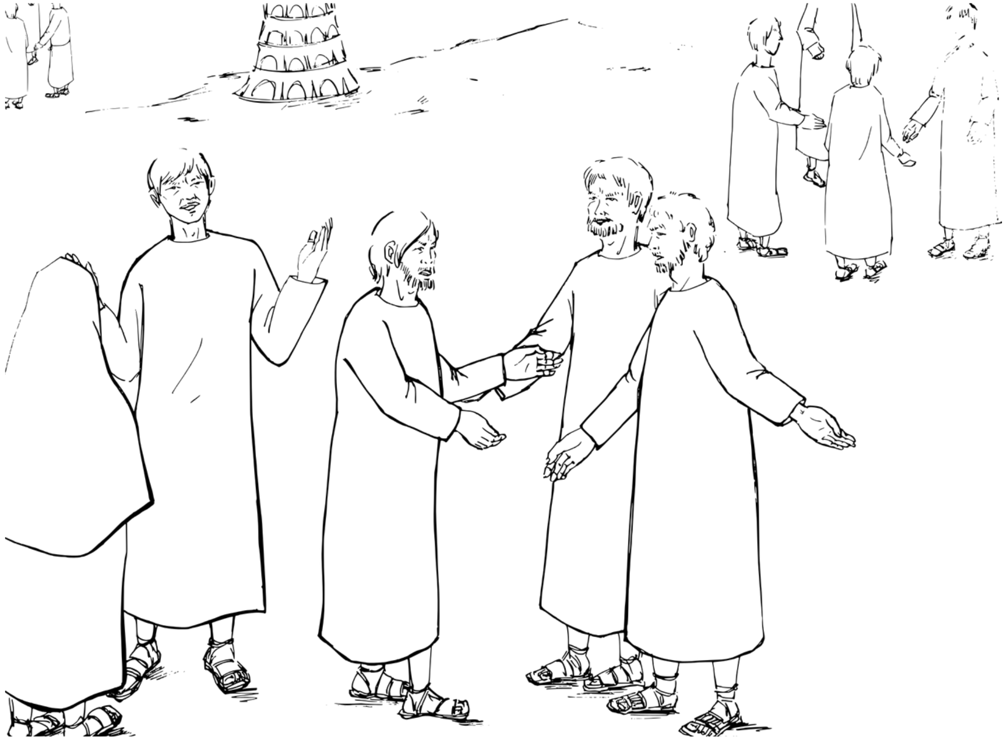
Numerous efforts to collect data to measure cybercrime exist today

Taxonomic conventions that exist elsewhere in public health and safety sectors are absent



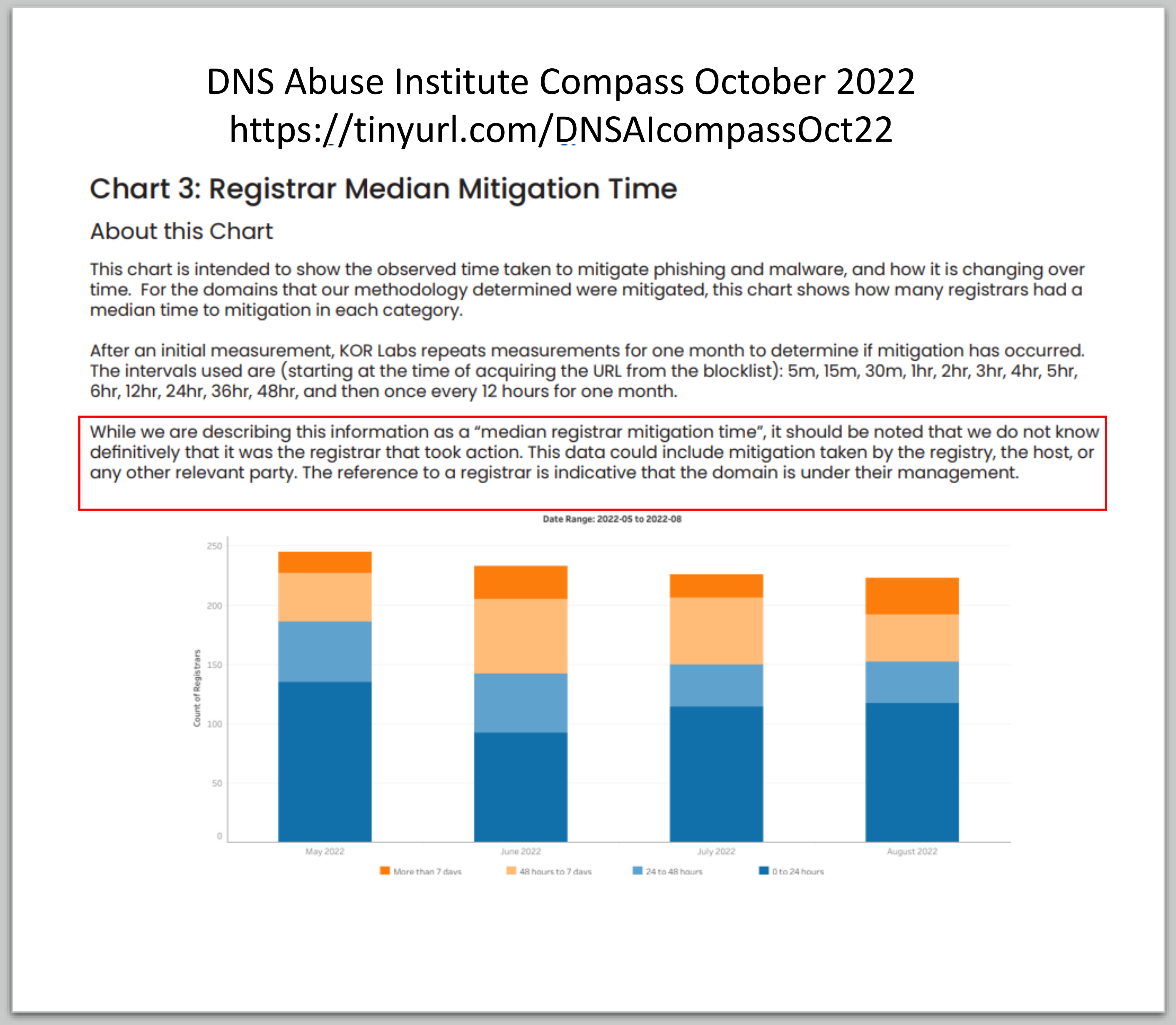
Result: We’re babbling

- Poorly defined measurements misinform audiences
- Snapshots often conflict with long term trends
- Narrowly scoped studies diminish value of findings
- Conflation or misuse of terminology misleads audiences



Poorly defined measurements misinforms audiences

- *Registrar Median Mitigation Time* credits registrars with mitigation activities they may not have performed
- Cannot use measurement to identify where new or different mitigation efforts could be applied



DNS Abuse Institute Compass October 2022
https://tinyurl.com/DNSAIcompassOct22

Chart 3: Registrar Median Mitigation Time

About this Chart

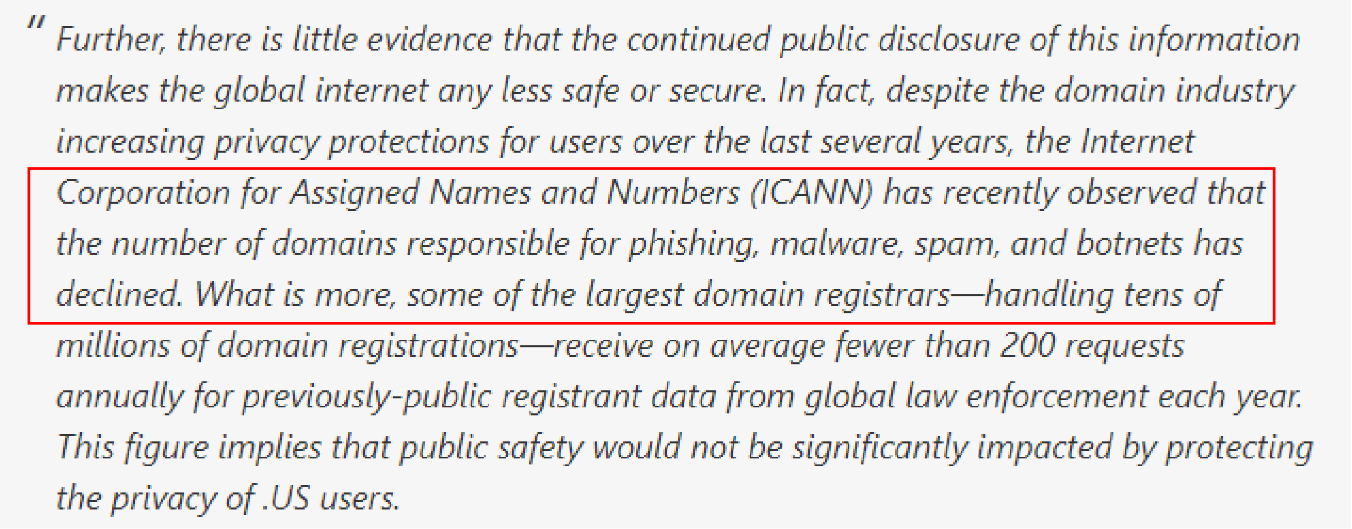
This chart is intended to show the observed time taken to mitigate phishing and malware, and how it is changing over time. For the domains that our methodology determined were mitigated, this chart shows how many registrars had a median time to mitigation in each category.

After an initial measurement, KOR Labs repeats measurements for one month to determine if mitigation has occurred. The intervals used are (starting at the time of acquiring the URL from the blocklist): 5m, 15m, 30m, 1hr, 2hr, 3hr, 4hr, 5hr, 6hr, 12hr, 24hr, 36hr, 48hr, and then once every 12 hours for one month.

While we are describing this information as a “median registrar mitigation time”, it should be noted that we do not know definitively that it was the registrar that took action. This data could include mitigation taken by the registry, the host, or any other relevant party. The reference to a registrar is indicative that the domain is under their management.

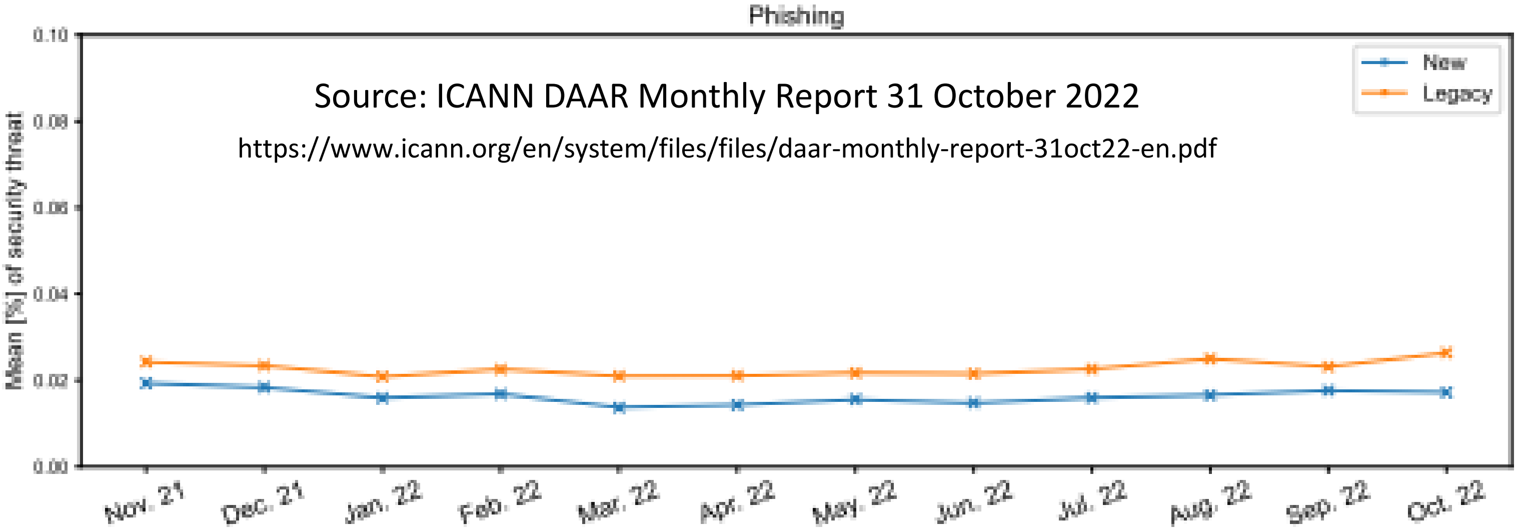
Snapshots often conflict with long term trends

Wyden Letter to NTIA on Privacy in .US Registry
https://tinyurl.com/wydenletterusregistry

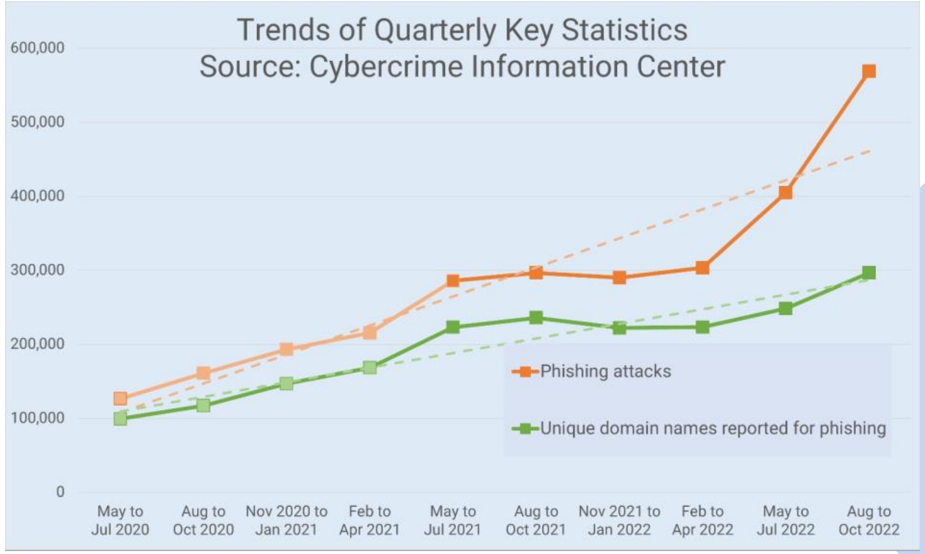


" Further, there is little evidence that the continued public disclosure of this information makes the global internet any less safe or secure. In fact, despite the domain industry increasing privacy protections for users over the last several years, the Internet Corporation for Assigned Names and Numbers (ICANN) has recently observed that the number of domains responsible for phishing, malware, spam, and botnets has declined. What is more, some of the largest domain registrars—handling tens of millions of domain registrations—receive on average fewer than 200 requests annually for previously-public registrant data from global law enforcement each year. This figure implies that public safety would not be significantly impacted by protecting the privacy of .US users.

The most recent ICANN monthly report on phishing suggests a downward trend that contradicts the trendline and results in an erroneous finding and implication.



Source: ICANN DAAR Monthly Report 31 October 2022
https://www.icann.org/en/system/files/files/daar-monthly-report-31oct22-en.pdf

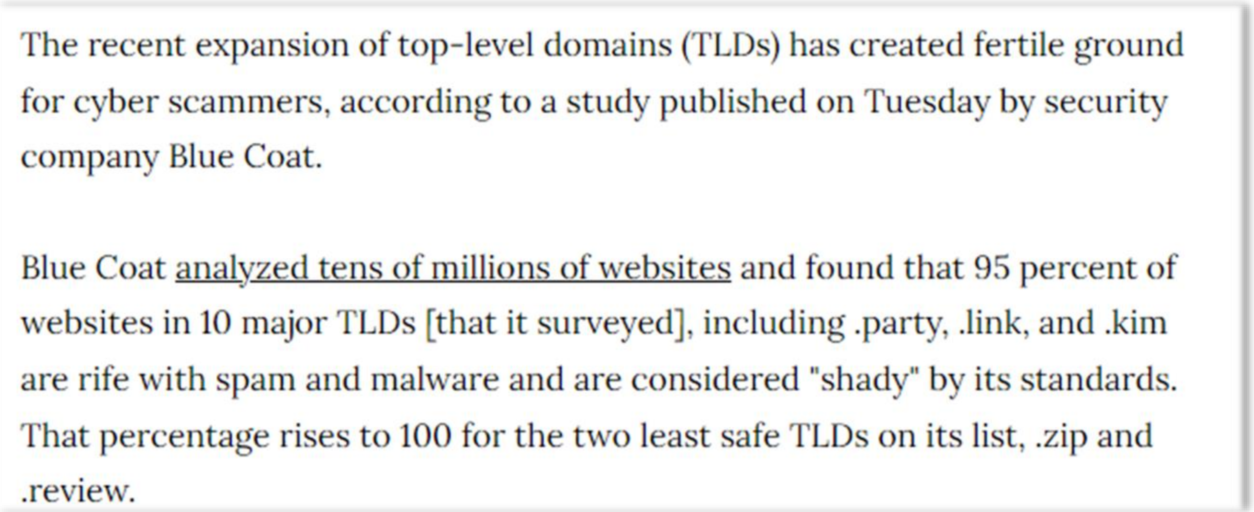


Trends of Quarterly Key Statistics
Source: Cybercrime Information Center

Poorly scoped studies diminish value of findings

Study: 100% of Websites in These Two Top-Level Domains Are 'Shady'

https://tinyurl.com/BlueCoatStudy



The recent expansion of top-level domains (TLDs) has created fertile ground for cyber scammers, according to a study published on Tuesday by security company Blue Coat.

Blue Coat analyzed tens of millions of websites and found that 95 percent of websites in 10 major TLDs [that it surveyed], including .party, .link, and .kim are rife with spam and malware and are considered "shady" by its standards. That percentage rises to 100 for the two least safe TLDs on its list, .zip and .review.

Sampling bias, lack of rigorous methodology led to generalizations and caused controversy.

Subsequent and more scientifically conducted studies are viewed with suspicion

Conflation or misuse of terminology misleads audiences

Failing to distinguish or qualify

- A cybercrime or DNS Abuse
  - phishing, malware, spam, counterfeiting
- FROM a means of perpetrating a crime or abuse
  - Phishing or ransomware *attack*, malware *download*, spam *campaign*
- FROM resources employed in the perpetration of a crime or abuse
  - Domain Name, URL, IP address, hosting account, name or mail server
- FROM the outcomes following execution of a crime
  - Infections, harms, losses, victim counts

Causes confusion and hampers comparisons across similar studies

Today, data collection and measurements efforts operate disjointly.

Reporting organizations use *ad hoc* conventions for classification and measurements.

Diverse metadata and tagging

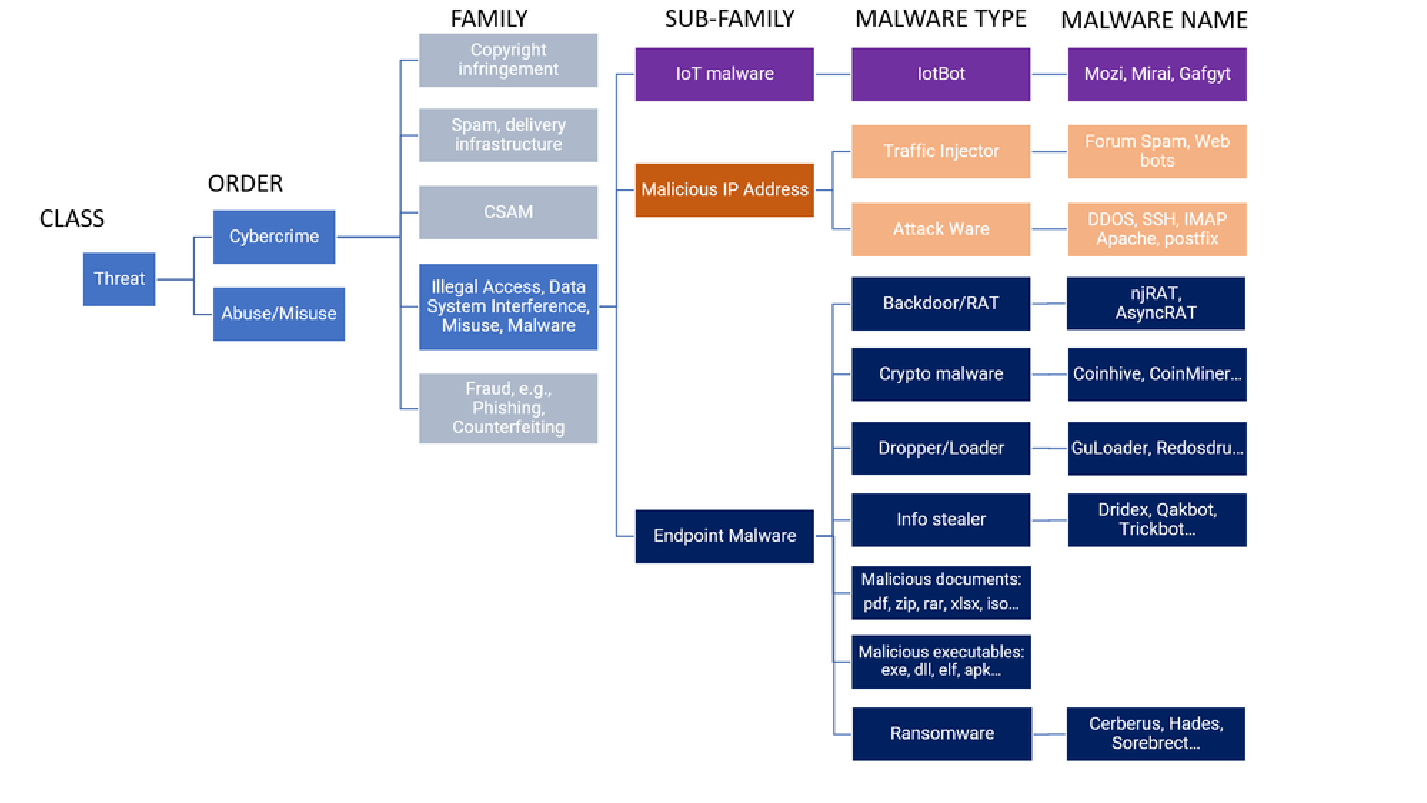
Malware URL	URLhaus
Trojan Linux CoinMiner	Malware_download/32,CoinMiner,exe,Tofsee
Backdoor Linux Gafgyt.A!MTB	Malware_download/ddos,gafgyt,mirai
Trojan FluBot	malware_download/Flubot
Trojan Qakbot	malware_download/Qakbot,qbot,Quakbot,xlsb
Trojan Emotet	malware_download/doc,emotet,epoch4,heodo
Trojan AgentTesla	malware_download/AgenTesla,AgentTesla,exe

The metadata in all cases is valuable.

Only certain feeds provide metadata.

Lack of conventional tagging results in (ad hoc) normalization

Ad hoc classification systems



CLASS · ORDER · FAMILY · SUB-FAMILY · MALWARE TYPE · MALWARE NAME

- Threat
  - Cybercrime
    - Copyright infringement
    - Spam, delivery infrastructure
    - CSAM
    - Illegal Access, Data System Interference, Misuse, Malware
      - IoT malware
        - IotBot
          - Mozi, Mirai, Gafgyt
      - Malicious IP Address
        - Traffic Injector
          - Forum Spam, Web bots
        - Attack Ware
          - DDOS, SSH, IMAP Apache, postfix
      - Endpoint Malware
        - Backdoor/RAT
          - njRAT, AsyncRAT
        - Crypto malware
          - Coinhive, CoinMiner…
        - Dropper/Loader
          - GuLoader, Redosdru…
        - Info stealer
          - Dridex, Qakbot, Trickbot…
        - Malicious documents: pdf, zip, rar, xlsx, iso…
        - Malicious executables: exe, dll, elf, apk…
        - Ransomware
          - Cerberus, Hades, Sorebrect…
    - Fraud, e.g., Phishing, Counterfeiting
  - Abuse/Misuse

A malware classification using the Computer Antivirus Research Organization (CARO) as a baseline to create a taxonomic ranking, where:

Class = *Threat*
Order = *Cybercrime*
Family = *Crime Type*
Sub-family = *Target or Origin*
Genus = *Malware Type*

Source: Cybercrime Information Center

How we interpret data makes comparisons challenging

Are these URLs part of one phishing attack or 10 attacks?

https://noorgate.com/dhl/clients/5Iw6LW.php?verification

https://noorgate.com/dhl/clients/gMKOF1.php?verification

https://noorgate.com/dhl/clients/GVBVPu.php?verification

https://noorgate.com/dhl/clients/20rxJL.php?verification

https://noorgate.com/dhl/clients/M1WVS8.php?verification

https://noorgate.com/dhl/clients/fVJRfR.php?verification

https://noorgate.com/dhl/clients/3IHxTa.php?verification

https://noorgate.com/dhl/clients/wzoPci.php?verification

https://noorgate.com/dhl/clients/JluNhK.php?verification

https://noorgate.com/dhl/clients/RD8zan.php?verification

Are these URLs part of one phishing attack or 10 attacks?

http://www[.]ekl-net.comnn-aocscsneisa.tqdvtw.top/jp.php

http://www[.]ekl-net.comm-ascaceesnea.tqdvtw.top/jp.php

http://www[.]ekl-net.comnn-aoaseeesaa.tqdvtw.top/jp.php

http://www[.]ekl-net.comm-ascaceeeeea.tqdvtw.top/jp.php

http://www[.]ekl-net.comm-ascaceeccea.tqdvtw.top/jp.php

http://www[.]ekl-net.comnn-aoascmesaa.tqdvtw.top/jp.php

http://www[.]eki-net.comn-aesceosneiesa.tqdvtw.top/jp.php

http://www[.]ekl-net.comnn-aoascmeoaa.tqdvtw.top/jp.php

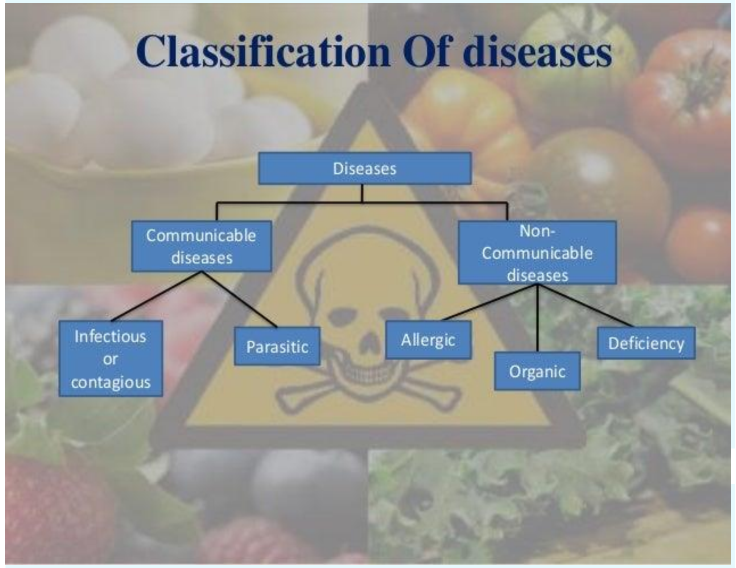
http://www[.]eki-net.comn-aesceesneiesa.tqdvtw.top/jp.php

http://www[.]ekl-net.comnn-aoasccesaa.tqdvtw.top/jp.php

We need taxonomic conventions

Follow the lead of public health

Develop a taxonomy and use conventionally applied metadata to better classify cybercrimes



Classification Of diseases

Diseases

Communicable diseases

Non-Communicable diseases

Infectious or contagious

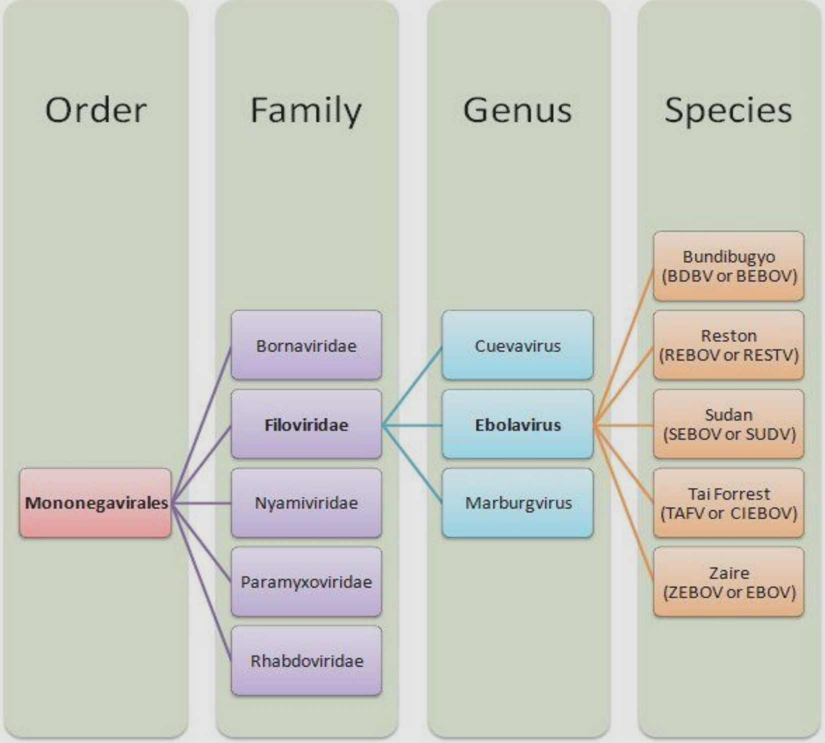
Parasitic

Allergic

Organic

Deficiency

Ebola hemorrhagic fever



Order

Family

Genus

Species

Mononegavirales

Bornaviridae

Filoviridae

Nyamiviridae

Paramyxoviridae

Rhabdoviridae

Cuevavirus

Ebolavirus

Marburgvirus

Bundibugyo (BDBV or BEBOV)

Reston (REBOV or RESTV)

Sudan (SEBOV or SUDV)

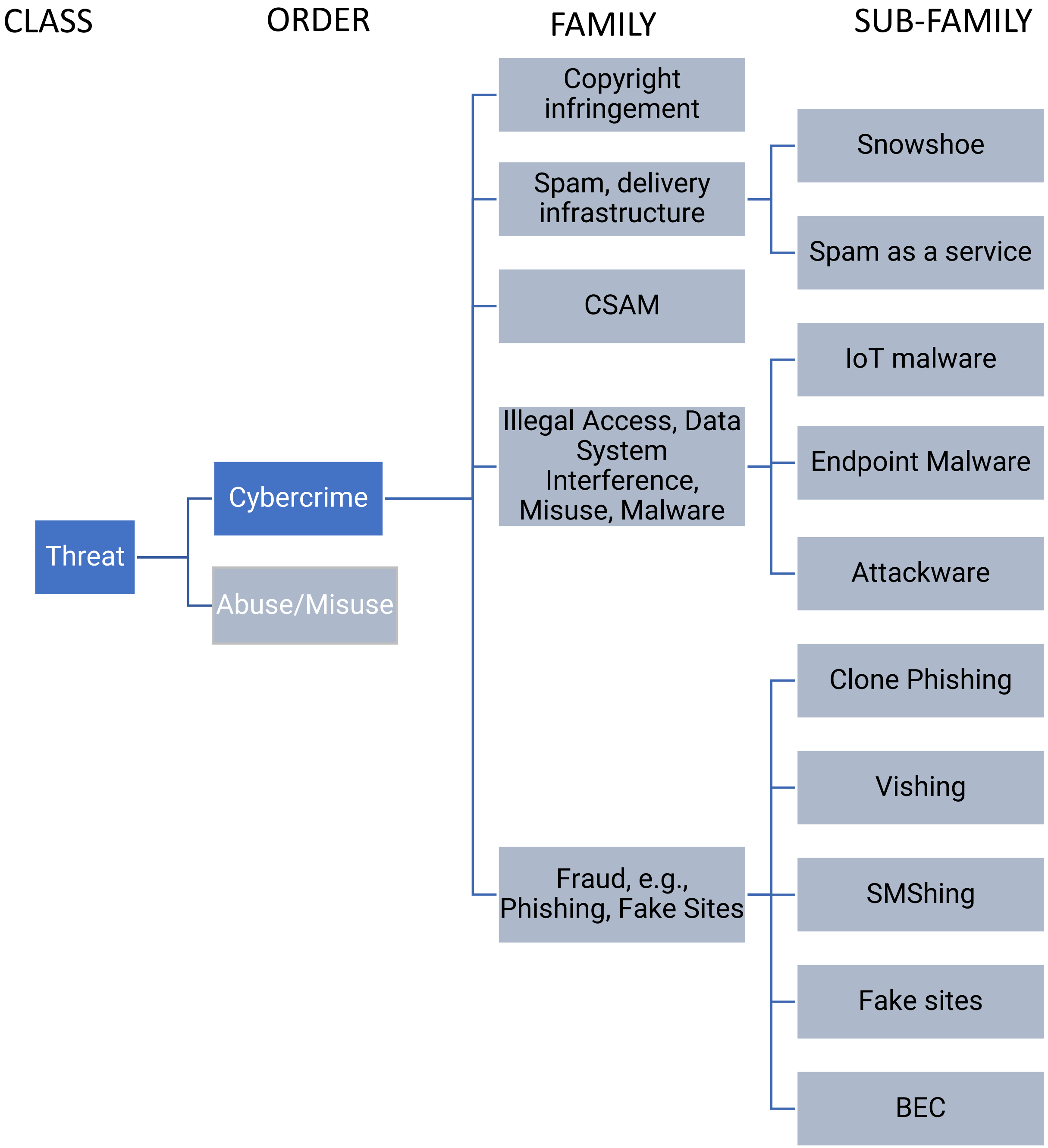
Tai Forrest (TAFV or CIEBOV)

Zaire (ZEBOV or EBOV)

Example: Follow Convention's Articles and Guidelines for fraud, network security, and copyright infringement

Budapest Convention	Cybercrime	Operational Security term
Articles 2, 6	Illegal access, misuse of device/software	Computer intrusion, unauthorized access, malware
Article 3, 21	Illegal interception, interception of content data	MITM attacks, web, DNS, mail redirection, data exfiltration
Articles 4 and 5	Data interference, System interference	DoS/DDoS attacks, destructive data breach, ransomware
Article 8	Computer related fraud	Phishing, Scam, Fake/counterfeit sites
Article 9	CSAM	Child pornography, Child abuse
Article 10	Copyrights infringement	Targeted Brands
Guidance Note #8	Spam content, act of sending, and infrastructure	Spam emitters, botnet C&C

Can we develop and adopt cybercrime taxonomies?



CLASS

ORDER

FAMILY

SUB-FAMILY

- Threat
  - Cybercrime
    - Copyright infringement
    - Spam, delivery infrastructure
      - Snowshoe
      - Spam as a service
    - CSAM
    - Illegal Access, Data System Interference, Misuse, Malware
      - IoT malware
      - Endpoint Malware
      - Attackware
    - Fraud, e.g., Phishing, Fake Sites
      - Clone Phishing
      - Vishing
      - SMShing
      - Fake sites
      - BEC
  - Abuse/Misuse

For measuring and reporting, set industry wide conventions; for example…

- Total number of reports collected from feeds
- {phishing, spam, malware…} attacks reported (URL similarity)
- Unique domain names reported for {phishing, spam, malware…}
- Maliciously registered domains reported for {phishing, spam, malware…}
- Top-level Domains (TLDs) where we observed {phishing, spam, malware…}
- Registrars that had domains under management reported for {phishing, spam, malware…}
- ASNs where {phishing, spam, malware…} sites were reported
- Operators (ASNs under common admin) where {phishing, spam, malware…} sites were reported
- Subdomain resellers where {phishing, spam, malware…} sites were reported

Other opportunities where conventions and uniformity will benefit public safety

- Industries and brands targeted by phishers and fake sites
- Goods most targeted
- Indicators of abusive intent in the composition of suspicious or abusive domain names
- Sinkhole operations
- Malware naming and typing
- Indicators of Compromise
- Conventions for registrar names, subdomain reseller names, brand names, and similar free-form metadata

Call for action

Proposal

Create reporting and archiving conventions for cybercrime machine event data

- Record schema for archiving events
- Precisely and rigorously define
  - Classification systems for cybercrimes and abuses of identifier systems
  - Event types (e.g., attacks, downloads)
  - Event records and their data elements
  - Criteria for data fidelity or confidence
  - Resources appropriated for events (domains, addresses, hosting executables, scripts)

The how and where…

- Begin with APWG working party to draft framework
- Continue at APWG EU event (tentatively May 2023)
- Invite comments
- Lather. Rinse. Repeat