Shareable Cyber Threat Intelligence Using Weak Anonymization

Lena Pons

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Document Markings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

DM19-0499

Overview

Sharing Cybersecurity Information is Challenging

Sharing is Worthwhile

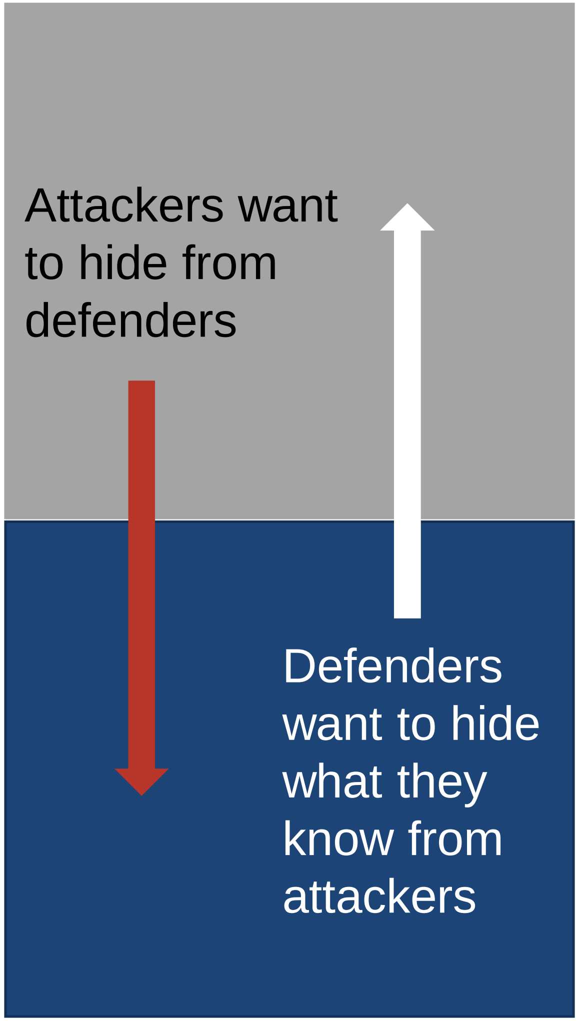
Barriers to Sharing Can Be Overcome

Tools to Enable Sharing

Shareable Cyber Threat Intelligence

Sharing Cybersecurity Information is Challenging

Cybersecurity is Adversarial


Attackers want to hide from defenders
Defenders want to hide what they know from attackers

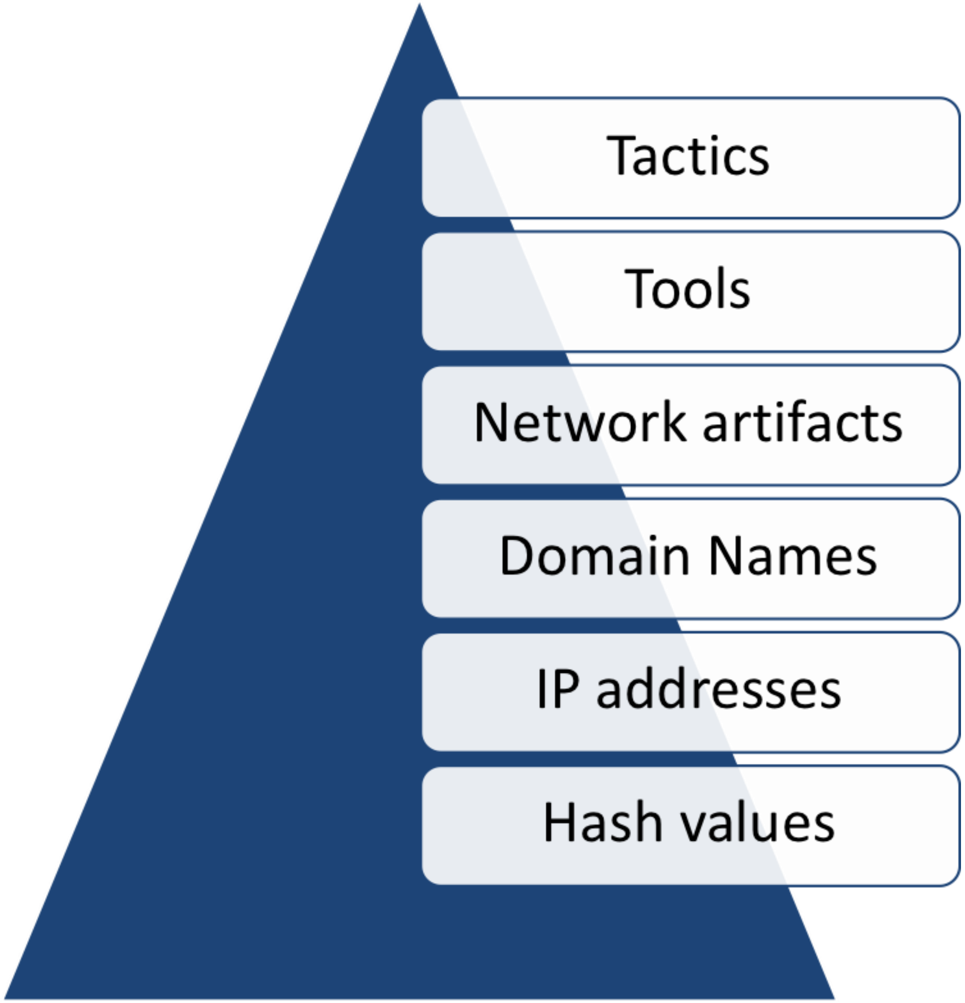
- Cyber attackers are continually updating techniques and infrastructure to evade detection
- Information is sensitive and valuable in this context
- Attackers want to stay ahead of defenders with the least effort & expense

Cybersecurity is Rapidly Changing

Attackers are continually changing observables that can be easily changed

Defenders seek to build information at higher levels of difficulty to change – this gives them advantage

Moving up the hierarchy gives defenders more time to operate



Cyber Threat Intelligence

- Cyber threat intelligence is a combination of observable information and prose descriptions
- Much of this information is currently shared through networks of individual contributions
- Some large scale open source information is available, e.g. databases of malware hashes, blogs, etc.
- Information Sharing and Analysis Centers provide sharing for cybersecurity information within sectors

Challenges

Observables aren’t always informative across sectors

Information gets stale quickly

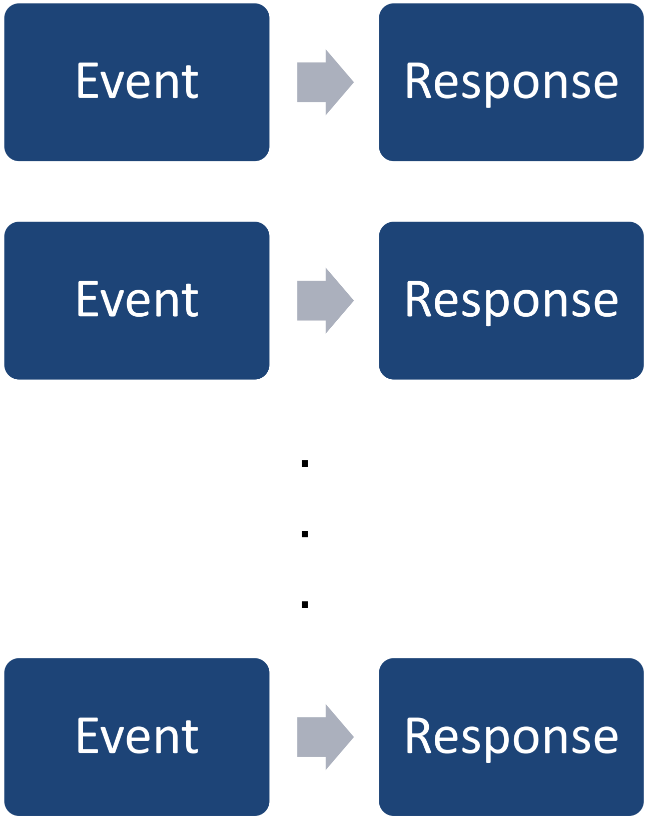
Information too voluminous to store efficiently

Data interoperability

Shareable Cyber Threat Intelligence

Sharing Cybersecurity Information is Worthwhile

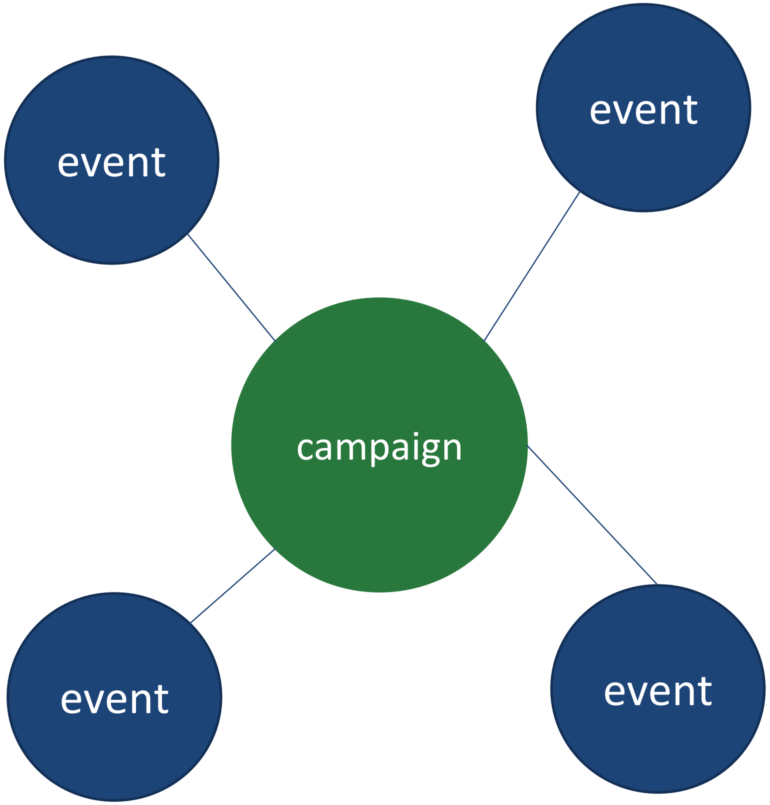
Defending Networks



Much of the current practice operates on a diagnose & treat model

Events are handled on an individual basis and patterns are hard to detect

Cyber epidemiology



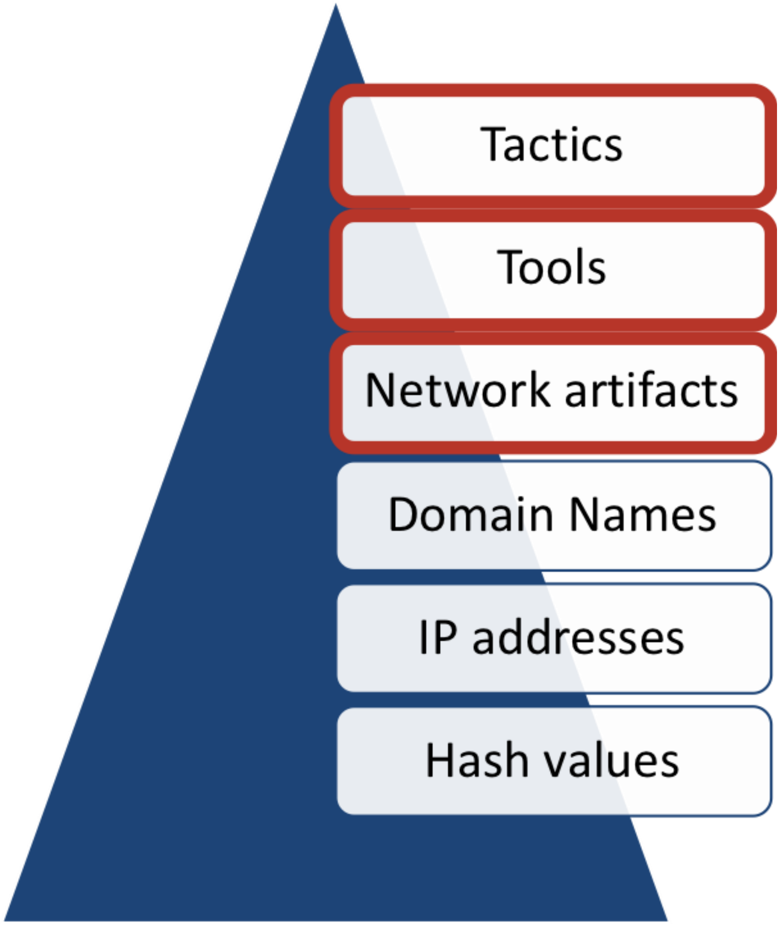
In the diagnose & treat model, it's hard to put events together

The goal of information sharing in cybersecurity is to detect events with similar observables

Requires up-to-date, actionable information

Machine Learning Runs on Data

- Can’t effectively learn models to identify higher level cyber observations without a large amount of data
- Complexity of the problem means large number of unhelpful training examples
- Expanding the available data is built on sharing



[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution.

Risks of Sharing are Real

- Giving away your defensive posture is just one element of hesitancy toward sharing
- PII protections mean holders of this type of information must apply certain safeguards
- Entities may not want to disclose that they’ve been affected by certain types of cyber attacks

Risks of Not Sharing are Real

- Entities are exposed to a large number of potential threats
- Hard to keep up - large networks are inundated by attacks, small networks usually do not have dedicated staff
- We cannot get ahead of tempo without building sharing relationships

Shareable Cyber Threat Intelligence

Barriers are Surmountable

Challenges (Recall)

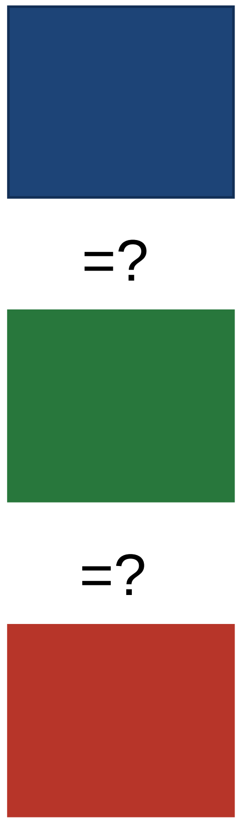
Observables aren’t always informative across sectors

Information gets stale quickly

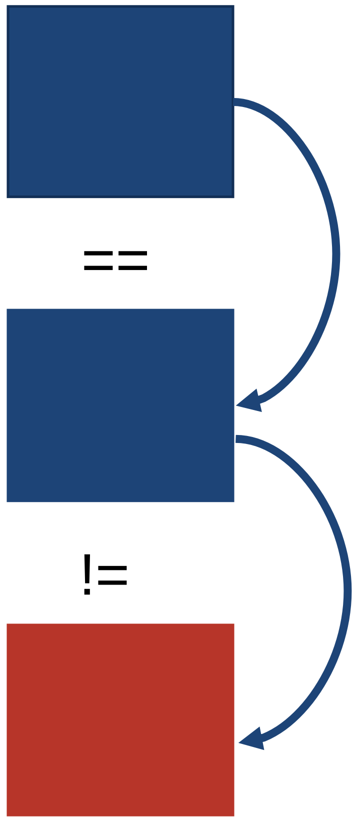
Information too voluminous to store efficiently

Data interoperability

Data Interoperability



Identifying whether an observable that is seen in one place is the same as one seen in another is frequently a challenge

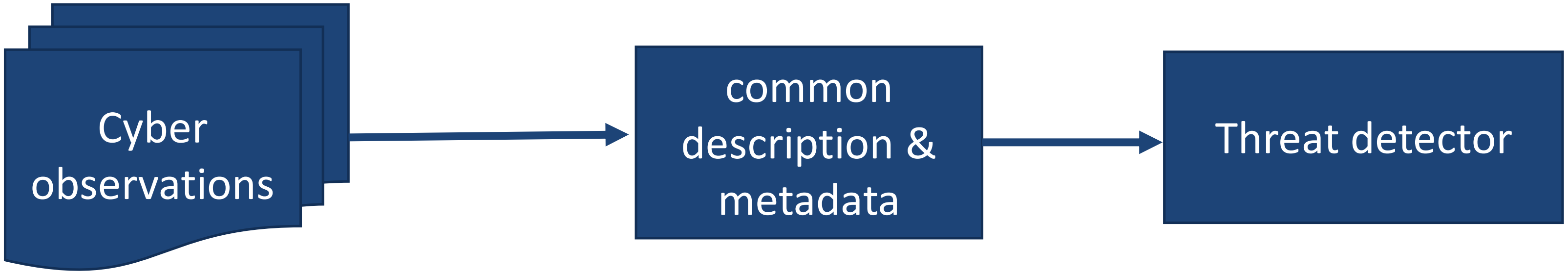


Common threat representation

Scale of Data

Information too voluminous to store efficiently

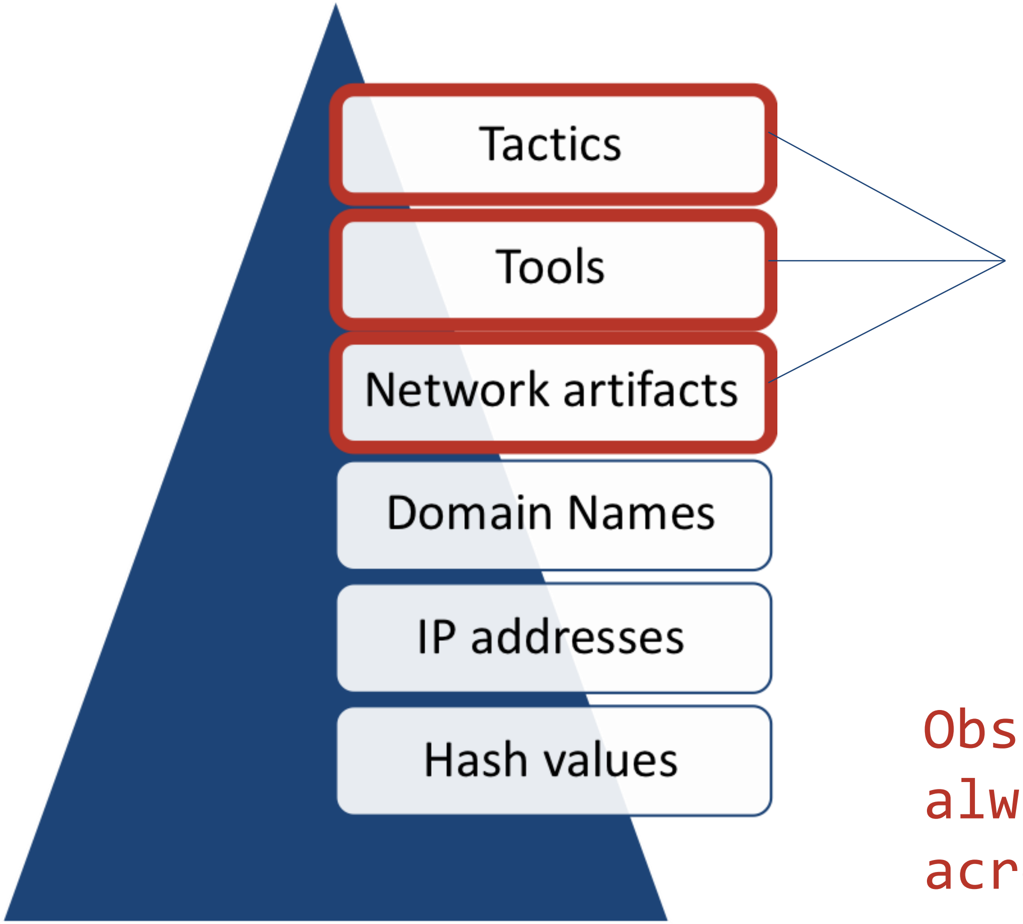
Higher level description means many observations => one



Too much data to ever store / process

Now this observation can be used to look for similar, not exact same

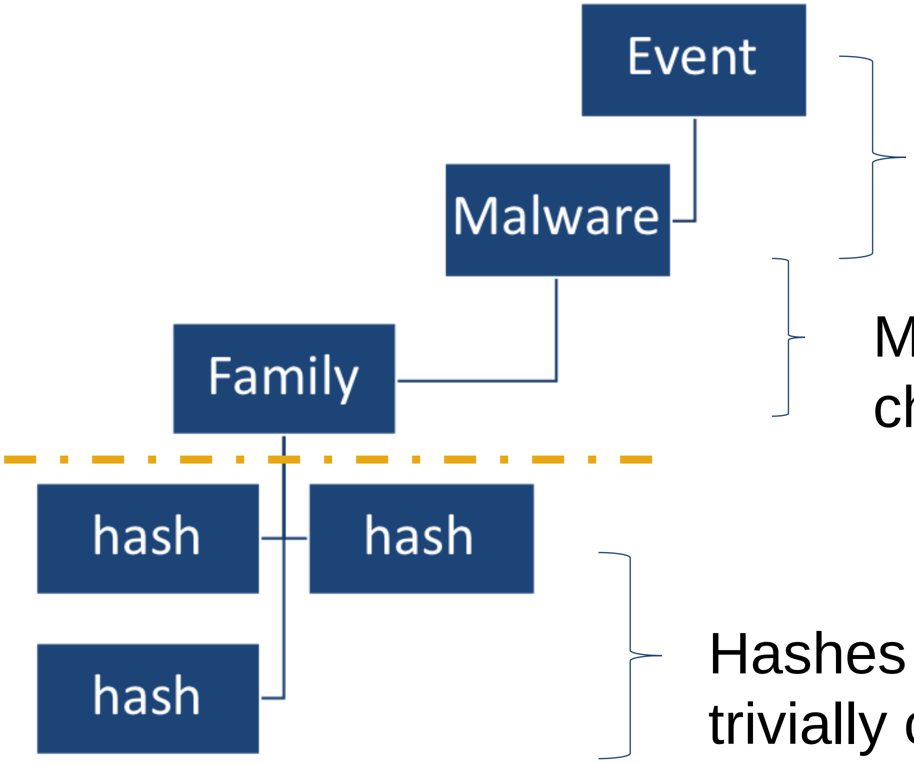
Information Value By Sector



Constructing observations at this level are more useful for looking at broad patterns

Observables aren’t always informative across sectors

Hierarchal Representations



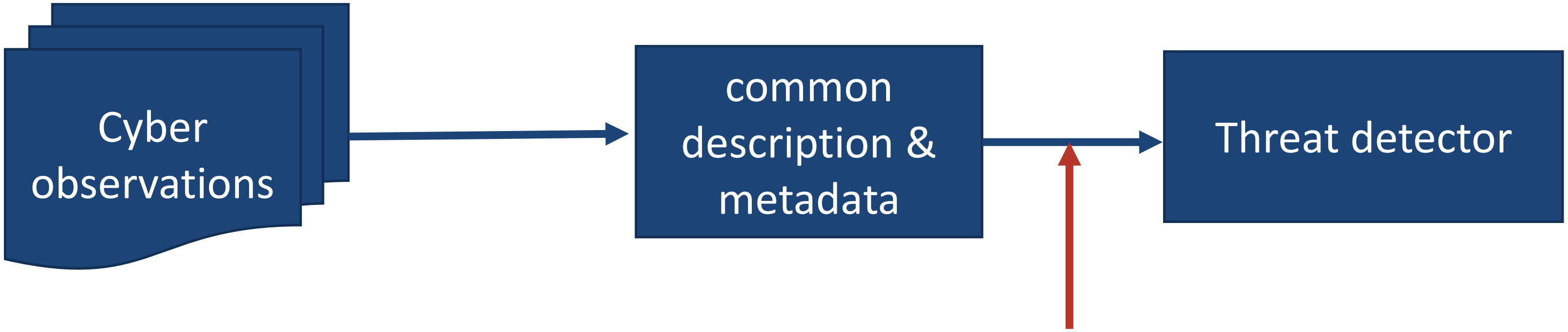
Malware events have happened for decades & will persist

Malware families change over time

Hashes can be trivially changed to evade detection

High Value for Sharing == Risky to Share

- Recalling cybersecurity is adversarial
- The most persistently useful cyber threat intelligence is
  - The hardest to generate
  - Requires most human intervention
  - Highest consequence if adversary learns about it

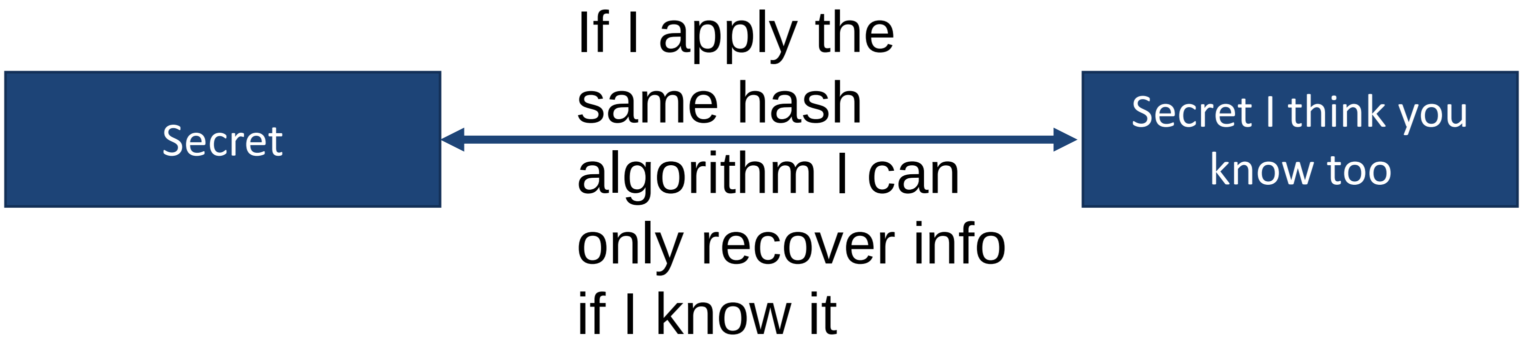


Adversary learns our detection criteria

Shareable Cyber Threat Intelligence

Tools to Enable Sharing

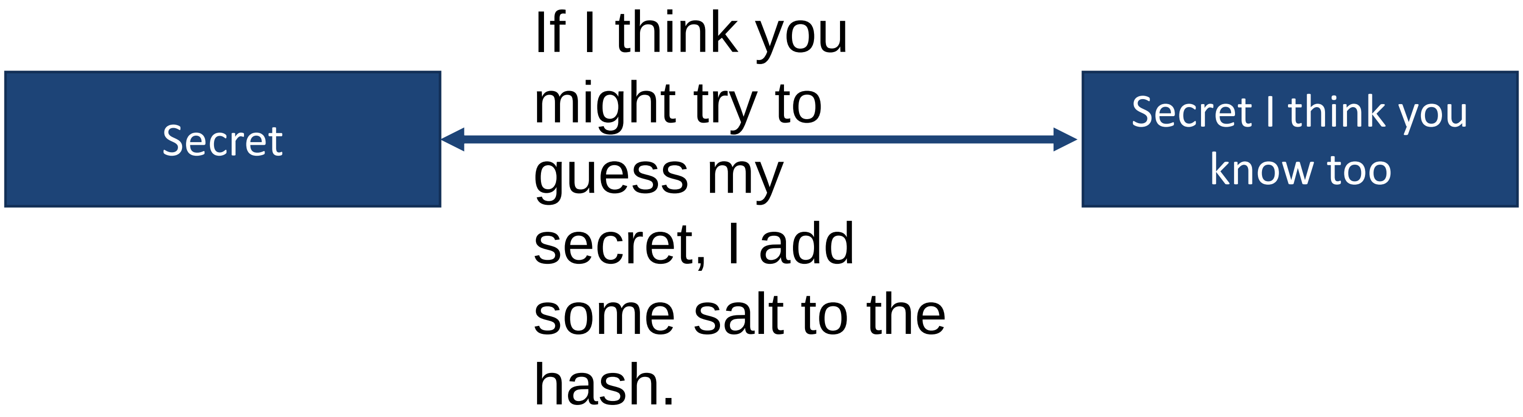
Hashing



Secret

If I apply the same hash algorithm I can only recover info if I know it

Secret I think you know too



Secret

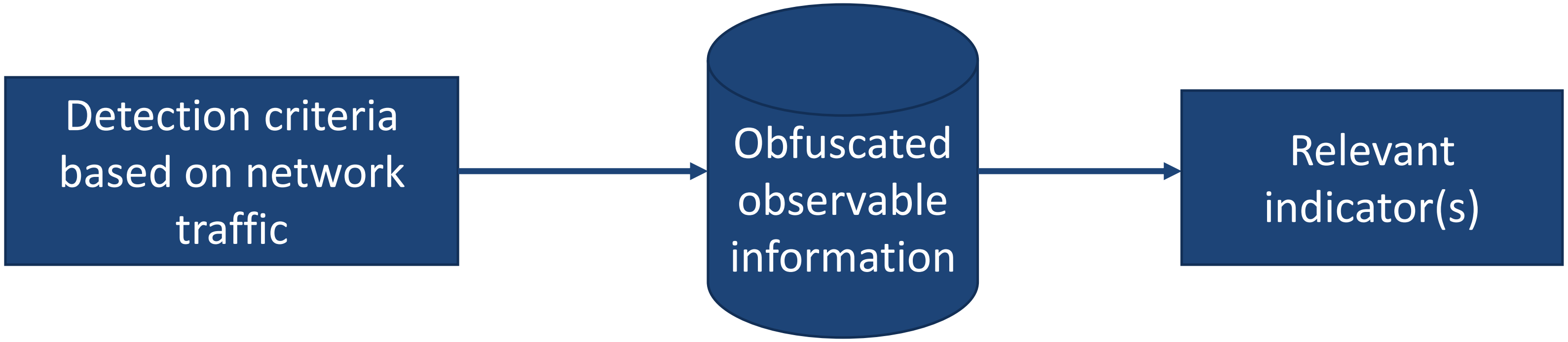
If I think you might try to guess my secret, I add some salt to the hash.

Secret I think you know too

Proxy Information

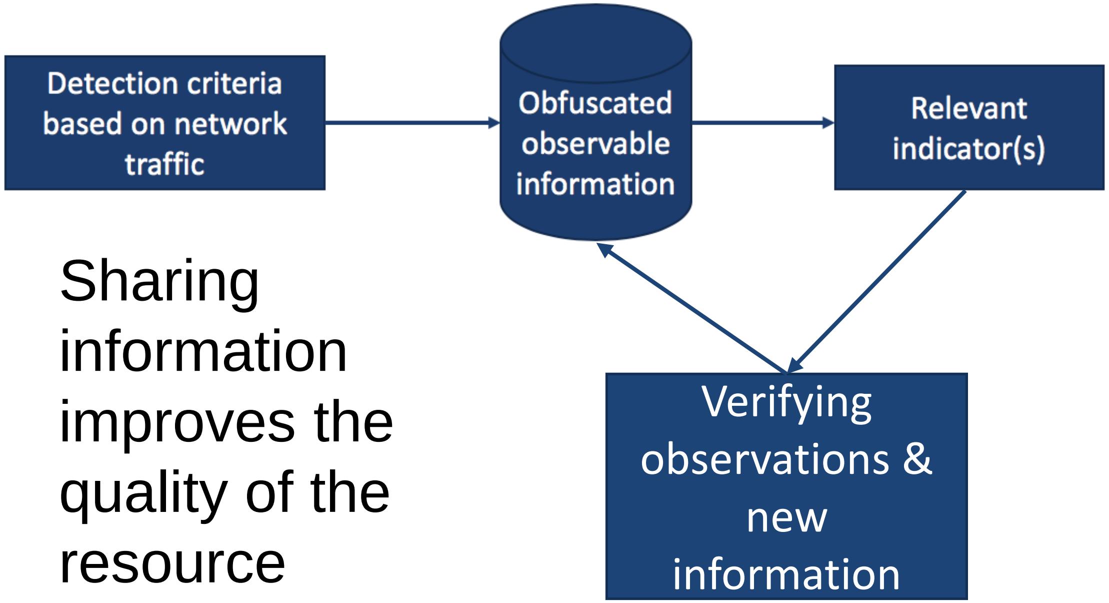
I can tell you something about an activity that is suspicious but I exclude some information that might tip off an adversary

If you see the behavior I told you about, I will share relevant information only

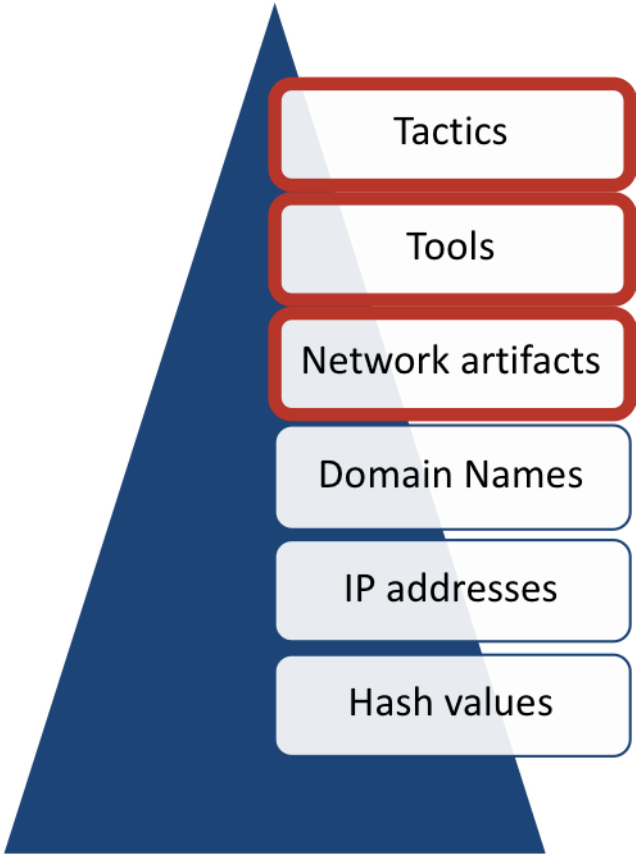

Detection criteria based on network traffic
Obfuscated observable information
Relevant indicator(s)

Sharing Is Good for Reuse

End goal is to construct more information at the higher levels of the pyramid


Detection criteria based on network traffic
Obfuscated observable information
Relevant indicator(s)
Verifying observations & new information

Sharing information improves the quality of the resource



Shareable Cyber Threat Intelligence

Conclusion

- Cyber threat intelligence is sensitive information
- Risks from sharing exist
- In adversarial space we have to go faster
- Need ML techniques to get there
- ML needs data
- Rewards from sharing exist
- We can provide mechanisms to share information

Lena Pons

Machine Learning Research Scientist

Software Engineering Institute

lepons@cert.org