Open Source Threat Intelligence

Kyle R Maxwell (@kylemaxwell)
Senior Researcher, Verizon RISK Team
Before we begin…

All trademarks belong to their respective owners. No association with any other organizations, sites, or projects is implied. And all opinions are my own.
What are we talking about?
Breaking it down

Open Source
• Publicly available data from overt sources
• Distinct from open-source software
• But all software discussed today is FLOSS

Threat
• Non-asset, non-vulnerability
• In VERIS A^4 terms: actor and action
• Not investigation-focused but can support it

Intelligence
• True intel is product of data and analysis
• Generalizing slightly here to include raw-ish data
• Focus on broadly gathering data, tools for analysis

CISPA and other political or legislative issues are out-of-scope for this talk
Breaches happen quickly but get discovered slowly

![Bar chart showing phases and durations of a breach](chart)

- **Compromise** (t=180):
  - Overall: 60%
  - Phase:
    - Financial: 11%
    - Espionage: 13%
    - Other: 13%
  - Duration:
    - Seconds: 2%
    - Minutes: 2%
    - Hours: 18%
    - Days: 41%
    - Weeks: 14%
    - Months: 22%
    - Years: 4%

- **Exfiltration** (t=39):
  - Overall: 36%
  - Phase:
    - Financial: 15%
    - Espionage: 18%
    - Other: 10%
  - Duration:
    - Seconds: 1%
    - Minutes: 9%
    - Hours: 11%
    - Days: 12%
    - Weeks: 62%
    - Months: 4%
    - Years: 4%

- **Discovery** (t=221):
  - Overall: 3%
  - Phase:
    - Financial: 3%
    - Espionage: 10%
    - Other: 18%
  - Duration:
    - Seconds: 1%
    - Minutes: 3%
    - Hours: 10%
    - Days: 10%
    - Weeks: 10%
    - Months: 10%
    - Years: 10%

- **Containment** (t=49):
  - Overall: 2%
  - Phase:
    - Financial: 2%
    - Espionage: 2%
    - Other: 2%
  - Duration:
    - Seconds: 2%
    - Minutes: 2%
    - Hours: 2%
    - Days: 2%
    - Weeks: 2%
    - Months: 2%
    - Years: 2%
Organizations don’t discover breaches internally

Figure 44: Discovery methods

- **Overall**
  - Unrelated party (Ext): 34%
  - Fraud detection (Ext): 24%
  - Customer (Ext): 9%
  - Law enforcement (Ext): 8%
  - Actor disclosure (Ext): 7%
  - Unknown: 5%
  - Reported by user (Int): 4%
  - Financial audit (Int): 3%
  - NIDS (Int): 1%
  - Log review (Int): 1%
  - Fraud detection (Int): 1%
  - HIDS (Int): 1%
  - Incident response (Int): 1%
  - IT audit (Int): 1%
  - Monitoring service (Ext): <1%

- **Small**
  - Unrelated party (Ext): 23%
  - Fraud detection (Ext): 14%
  - Customer (Ext): 10%
  - Law enforcement (Ext): 6%
  - Actor disclosure (Ext): 6%
  - Unknown: 4%
  - Reported by user (Int): 2%
  - Financial audit (Int): 2%
  - NIDS (Int): 1%
  - Log review (Int): 1%
  - Fraud detection (Int): 1%
  - HIDS (Int): 1%
  - Incident response (Int): 1%
  - IT audit (Int): 1%
  - Monitoring service (Ext): <1%

- **Large**
  - Unrelated party (Ext): 52%
  - Fraud detection (Ext): 35%
  - Customer (Ext): 7%
  - Law enforcement (Ext): 1%
  - Actor disclosure (Ext): 5%
  - Unknown: 7%
  - Reported by user (Int): 9%
  - Financial audit (Int): 1%
  - NIDS (Int): 4%
  - Log review (Int): 4%
  - Fraud detection (Int): 2%
  - HIDS (Int): 4%
  - Incident response (Int): 1%
  - IT audit (Int): 1%
  - Monitoring service (Ext): 1%
Threat Data Sources
Collective Intelligence Framework

collectiveintel.net

- REN-ISAC project
- Sucks in feeds of IOCs from public and private sources
- Focuses on lower end of “pyramid of pain”
- Exports data to infrastructure or supports lookup during response

David J. Bianco
detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html
CIF query types

Searches

cif -q 129.110.10.1

cif -q ns1.utdallas.edu

Feeds

cif -q infrastructure/malware -c 50

CLI and RESTful API
OSINT IOCs

- Abuse.ch
- AlienVault
- Blocklist.de
- CleanMX
- Emerging Threats
- Forensic Artifacts
- Nothink
- Shadowserver
- Spamhaus

Among others…
Passive DNS

Historical records of actual DNS responses

• ISC DNSDB
• BFK edv-consulting
• Virustotal

;; bailiwick: butlesuh.ru.
;; count: 2
;; last seen: 2013-04-04 19:55:24 -0000
butlesuh.ru. IN A 1.174.2.127

;; bailiwick: butlesuh.ru.
;; count: 2
;; first seen: 2013-04-05 01:59:40 -0000
;; last seen: 2013-04-05 01:59:40 -0000
butlesuh.ru. IN A 2.60.67.146
Malware data

VirusTotal
- *Sine qua non* for existing public data
- Search by hash, URL, domain, or other indicators
- Includes passive DNS related to malware callouts

Malwr.com
- Additional data including feeds of recent samples and indicators
- Part of Shadowserver Foundation

VirusShare.com
- Large repository of malware samples of all types
- 3 TB of data, indexed and searchable
- Distributed via BitTorrent
Threat Actor Tracking
What’s a threat actor?

From VERIS:

Entities that cause or contribute to an incident are referred to as “threat actors”. There can be more than one actor involved in any particular incident, and their actions can be malicious or non-malicious, intentional or unintentional, causal or contributory. VERIS recognizes three primary categories of threat actors – External, Internal, and Partner.

www.veriscommunity.net/doku.php?id=actors

Not THAT kind of threat actor!
(Gary Oldman, public domain image)
Threat actor sources

Social Media

- Twitter (particularly via the API or RSS)
- Pastebin (e.g. @pastebindorks)
- Google Alerts are particularly useful for monitoring specific actors

Defacements and incidents

- Zone-h.org
- Mirror-ma.com
### Storing raw data

**BYODB**
- Use APIs and scripting languages (Python)
- Store in document database (MongoDB)
- Highly flexible but requires a bit more effort

**Web tools**
- Evernote
- Feedly
- Ifttt
- Delicious

Impossible to do properly without automation
Analysis
Maltego

Write local transforms to assist in enriching your data
Canari platform simplifies the process of development and deployment

canariproject.com
Malformity

Written principally by Keith Gilbert (VZ RISK)
MALware transFORMs and ent[ITY]ities

github.com/digital4rensics/Malformity/
Malformity

Simplifies basic analysis and research
Local repositories and analysis

malwarehouse

Basic database for storing samples from the command line. Think of this as your “working set”.

sroberts.github.io/malwarehouse/

VxCage

Larger, more complete database with a RESTful API interface. Think of this as your complete historical repository.

github.com/cuckoobox/vxcage

Cuckoo Sandbox

Dynamic malware analysis using Virtualbox. Takes screenshots, integrates with Virustotal, exposes an API, and is written in Python.

www.cuckoosandbox.org
Threat intel standards

OpenIOC and CybOX

- OpenIOC originally produced by Mandiant under Apache 2 license (openioc.org)
- Similar to CybOX from MITRE (cybox.mitre.org)
- Capture stateful properties (file hashes, IPs, HTTP GET, registry keys and values)

STIX

- Give context to indicators (CybOX) and other data (stix.mitre.org)
- TTPs
- Exploitation targets
- Campaigns
- Courses of Action [COA]
General threat analysis

**Threat intelligence and actors**

Use a wiki with defined templates like those from Scott Roberts for keeping profile data on specific threat actors. Link back to your document repository (e.g. in MongoDB).

- Artifacts
- Exploits
- Intrusion sets
- Third-party intelligence
- Threat actors

[github.com/sroberts/threat-intel-templates](https://github.com/sroberts/threat-intel-templates)

**Indicators of Compromise**

Pull feeds from CIF or similar tools into your SIEM. Organizations without an existing deployment may want to look into OSSIM to get started.

[communities.alienvault.com](https://communities.alienvault.com)

Not a lot of open-source tools for sweeping hosts broadly. pyioc is one example:

[github.com/jeffbryner/pyioc](https://github.com/jeffbryner/pyioc)

This is where a lot of the heavy lifting occurs.
How can you collaborate?

Use standards
- OpenIOC / CybOX
- STIX (builds on CybOX)

Trust groups
- Not “open source”, strictly speaking
- But do good work and keep some of it in the public
- Can be significant and targeted boost

Software development
- FLOSS projects depend on the community
- Github is a great place to get started
- Not just developers: use case feedback, docs, etc!

Threat actors talk to each other. We have to do the same.
Thanks to great people doing great work

David J Bianco (@davidjbianco)
Jeff Bryner (@p0wnlabs)
Keith Gilbert (@digital4rensics)
Claudio Guarnieri (@botherder)
Andrew Macpherson (@andrewmohawk)

J-Michael Roberts (@forensication)
Scott Roberts (@sroberts)
Alessandro Tanasi (@jekil)
Wes Young (@barely3am)
Future Directions

• Threat actor tracking in particular is relatively nascent in the public domain
• Lots of attention on getting better at sharing low-end IOCs
• Determine and detect TTPs (machine learning?)

Want to talk more?
@kylemaxwell
kyle.maxwell@verizon.com