Challenges in Incident Response

Cyber Threat Intelligence - an overview and practical approaches using open source security tools

CIRCL
Computer Incident
Response Center
Luxembourg

Alexandre Dulaunoy -
TLP:WHITE

CIRCL

CIRCL
Computer Incident
Response Center
Luxembourg

- The Computer Incident Response Center Luxembourg (CIRCL) is a government-driven initiative designed to provide a systematic response facility to computer security threats and incidents.
- CIRCL is the CERT for the private sector, communes and non-governmental entities in Luxembourg.

Figures at CIRCL

- **1.4GB** of compressed malware sample in a day.
- An average of **2-4TB** per evidence acquisition (disk, memory, ...) including analysis artefacts or duplicate analysis information.
- **1.2GB** of compressed network capture from the operational honeypot network (HoneyBot).
- **10-20 million** records added or updated in the Passive DNS in a day.
- **500 million** of X.509 certificates in the Passive SSL.

Do we have an issue with such volume of data?

- Storage price goes down and it will probably follow this trend.
- Storing huge amount of data is still practical and CSIRT can usually handle it.
- **Write-speed on disk** is still the main limitation (e.g. wire speed increased faster than the I/O).

Where are the real challenges in a day-to-day CSIRT operation?

- **12000 requests per second** to lookup records in the Passive DNS.
- Collections (network, disk, memory) by CSIRTs are often **unstructured**,
- sources of data are **uncontrolled and unstrusted**
- and **incomplete**.

Homogeneous data versus heterogeneous data

- 45TB of **normalized and homogeneous network capture** is fundamentaly different than 45TB of **black-hole network capture**.
- Discarding is easy in normalized traffic.
- In incident response, **protocol errors or incomplete packets are part of the potential attacks**.
- Parser errors and exceptions are more common on an untrusted and uncontrolled data sources.
- Data mining capabilities highly depend of the **data structuration** (e.g. exfiltration channels are rarely respecting the network layers).
- If the structuration is close to zero, more human pre-analysis is required.

What are the key factors in incident response?

- **Reduce workload** for the analysis (e.g. a full file-system forensic analysis of a standard system can take up to 10 days).
- Allow **fast lookup** in the data collected and processed.
  - Easier the access of correlation is, faster is the exclusion or inclusion of data.
  - Dynamic feedback on the data from the users (what are the most queried records?).
- Reduce false positive but **false negative reduction is more important** (e.g. can you miss an evidence in a critical case?).

How do we try to improve?

- Data-structure allowing **fast lookup** and fast update/counting.
  - Bitindex, Bloom filters, HyperLogLog...
  - Space efficient in-memory key/value store.
- **Parallel processing** of large datasets introduces challenges in checkpointing and updates (e.g. a crash of a parser is not uncommon from untrusted datasets).
  - Simple "parallel processing" frameworks versus complex frameworks (e.g. "limiting the cost of bootstrapping", memory usage and overhead of a framework).

Improving with the feedback loop

- The greatest benefit for data mining is to introduce **human feedback early**.
- Analysts discover outliers, errors or even missing data.
- Feedback can be used to improve algorithms, data structuration (e.g. 4th iteration of the CIRCL Passive SSL data structure) or query interfaces.

How to get analysts feedback?

- Integrate lookup services in the tools used by the analysts.
- Provide multiple UI to promote the reuse of the datasets.
- Support the classification of the results (e.g. a source of classified dataset).
- $\rightarrow$ MISP, malware information and threat sharing platform, is developed to support this.

Q&A

- Session of the M2 SSI:
  `https://www.foo.be/cours/dess-20172018`
- Joining us for an internship?
  `https://www.circl.lu/projects/internships`
- PGP key fingerprint: 3B12 DCC2 82FA 2931 2F5B 709A 09E2 CD49 44E6 CBCD