SHARING AND AUTOMATION FOR PRIVACY PRESERVING ATTACK NEUTRALIZATION

Graph-based Network Traffic Analysis for Incident Investigation

Milan Cermak

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Network Traffic Analysis for Incident Investigation

When some serious incident happened in the maintained network, we need to investigate its type, origin, impact, and spread to prevent further damage.

- How did the malware get on the machine?
- Did the attacker exploit any vulnerability?
- Did the machine communicate to a malware C&C or another suspicious IP address?
- Did the machine communicate with other devices in our network? How?
- Did any device from our network communicate with the same destinations as the compromised one?

Incident investigators utilize various tools to answer these questions; this presentation focuses only on network traffic analysis and specifically packet trace analysis (the same approaches are relevant for IP flow analysis and other sources of network traffic data).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Network Traffic Analysis Tools – Wireshark

- A widely-used network protocol analyzer providing insights into network activity at a microscopic level.
- De facto standard for packet trace analysis.

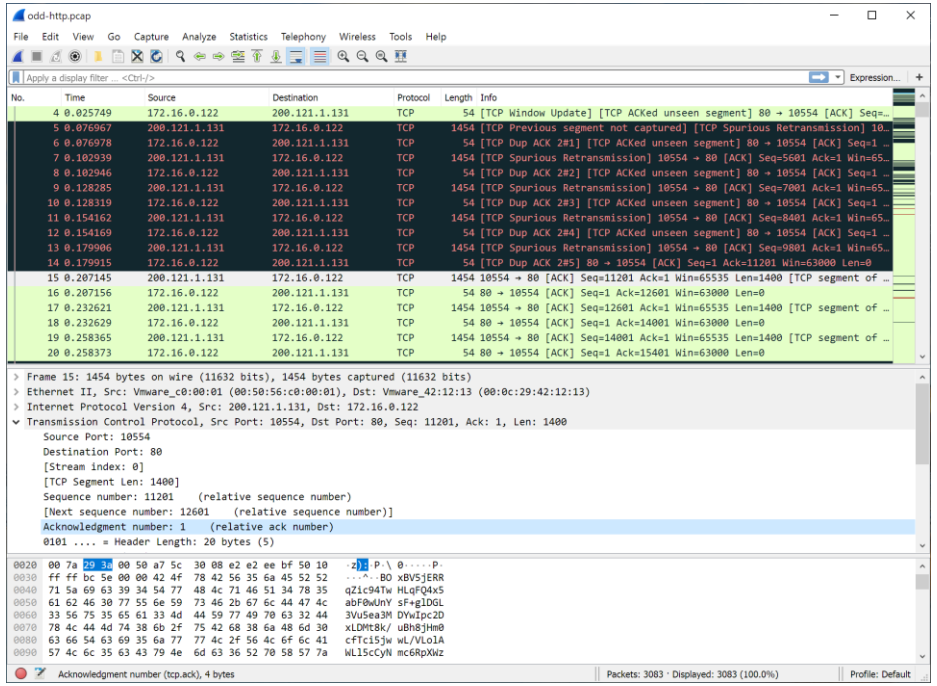
\+ Rich and detailed support of many different protocols.

\+ Ability to analyze all network traffic metadata.

\- Performance issues in analyzing large packet traces.

\- Limited overview of the whole packet trace.

\- Missing connection to other information sources.



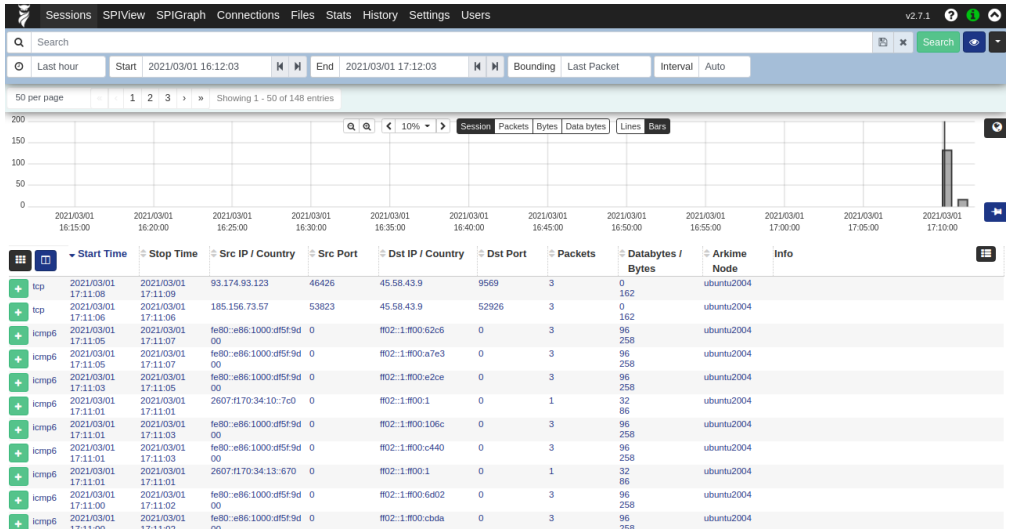
Wireshark: https://www.wireshark.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Network Traffic Analysis Tools – Arkime (Moloch)

- A large-scale, open-source, indexed packet capture and search tool with a web interface.

+ Indexed data storage for fast data analysis.
+ Extraction of various information from network sessions and other metadata.
+ Basic statistics of extracted data.
+ Export of selected connections as packet traces.
- No alerts correlation.
- Missing connection to other information sources.



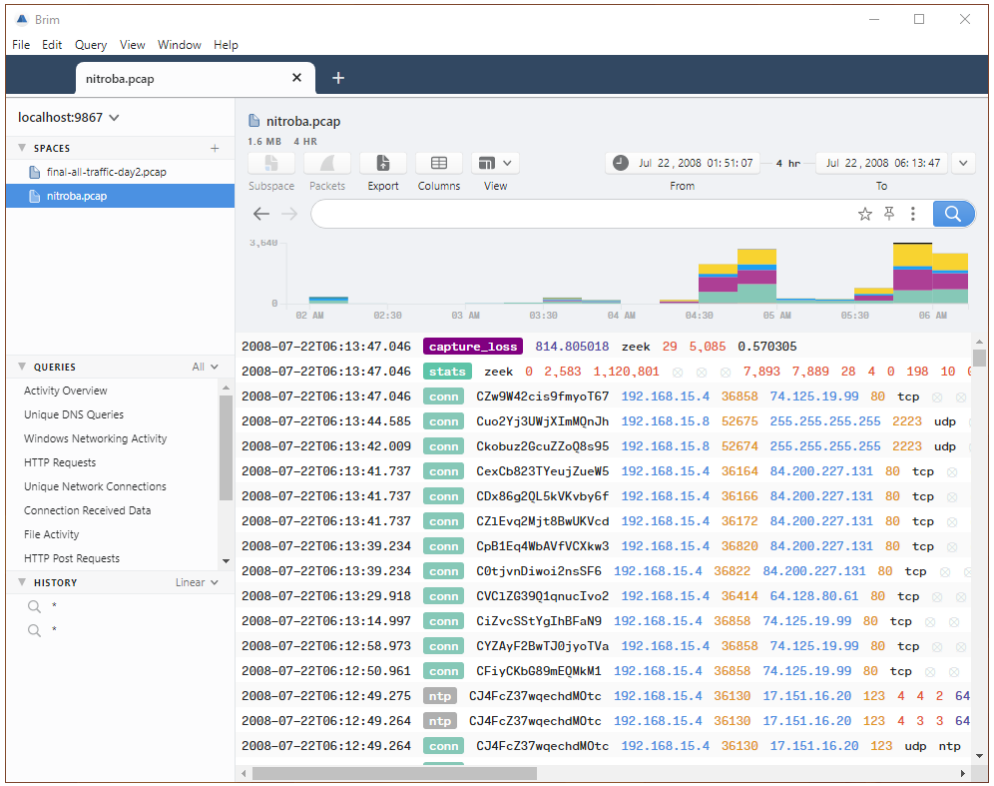
Arkime: https://arkime.com/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Network Traffic Analysis Tools – Brim

- An open-source desktop application combining Wireshark and Zeek (https://zeek.org/) network security monitor.

+ Utilization of a Zeek to extract relevant information.
+ Indexed data storage for fast data analysis.
+ Alerts correlation (Suricata or external source).
+ Basic statistics of extracted data.
+ Export of selected connections as packet traces.
- Custom query language.
- Missing connection to other information sources.



Brim: https://www.brimsecurity.com/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Data Analysis and Human Brain

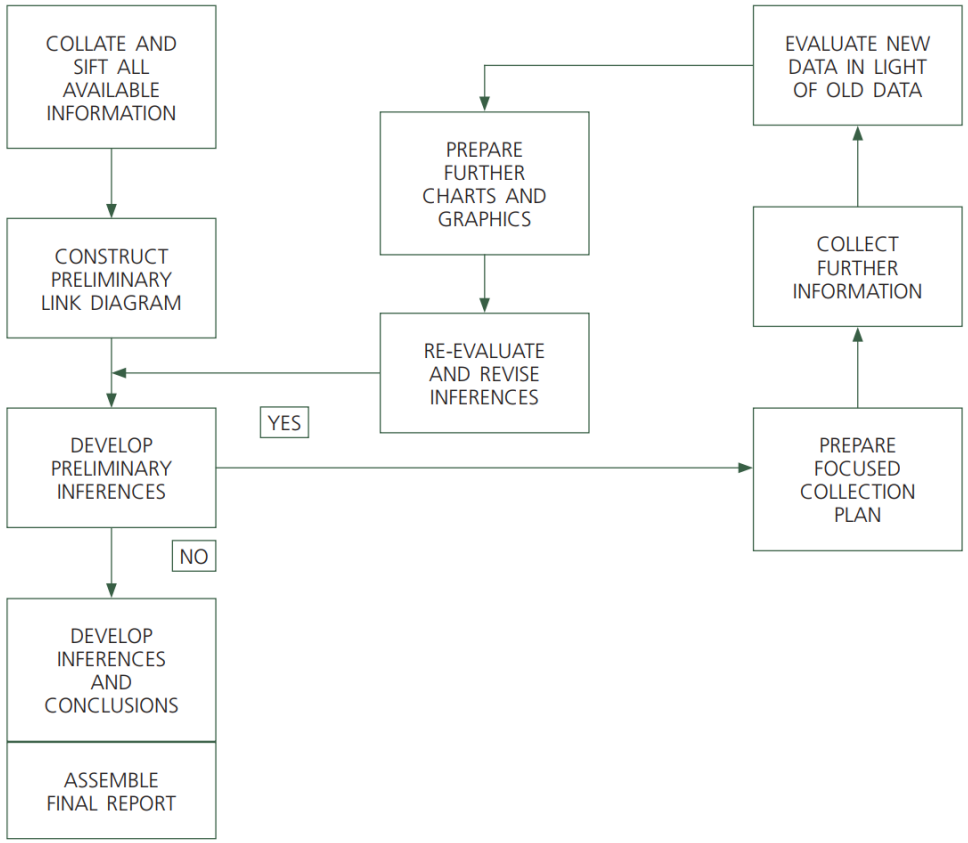
The human brain is used to perceiving the surrounding world and data in associations

- We use associations every day, so why not use them during network traffic analysis and incident investigation?
- Traditional analysis tools provide association-based analysis only in limited form or not at all.
- Relationship diagram allows us to get a broader context to analyzed data thanks to the visual aspect.
- It is a commonly used technique in a criminal investigation.

Paper vector created by macrovector - www.freepik.com

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

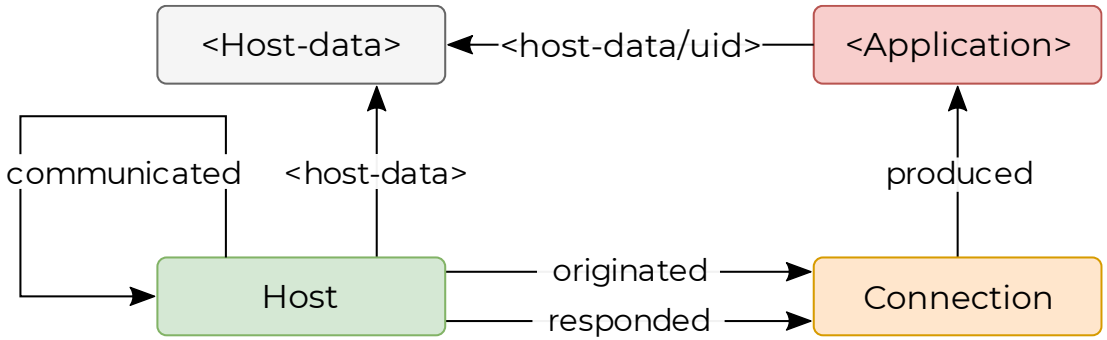
Criminal Investigation and Link Analysis



United Nations Office on Drugs and Crime (UNODC) – Criminal Intelligence: Manual for Analysts

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Association-based Network Traffic Representation



- Initial version of this representation was proposed by Niese and further developed by Leichtnam et al.
- Our model further develops these proposals and simplifies them to ease understanding by the analyst.

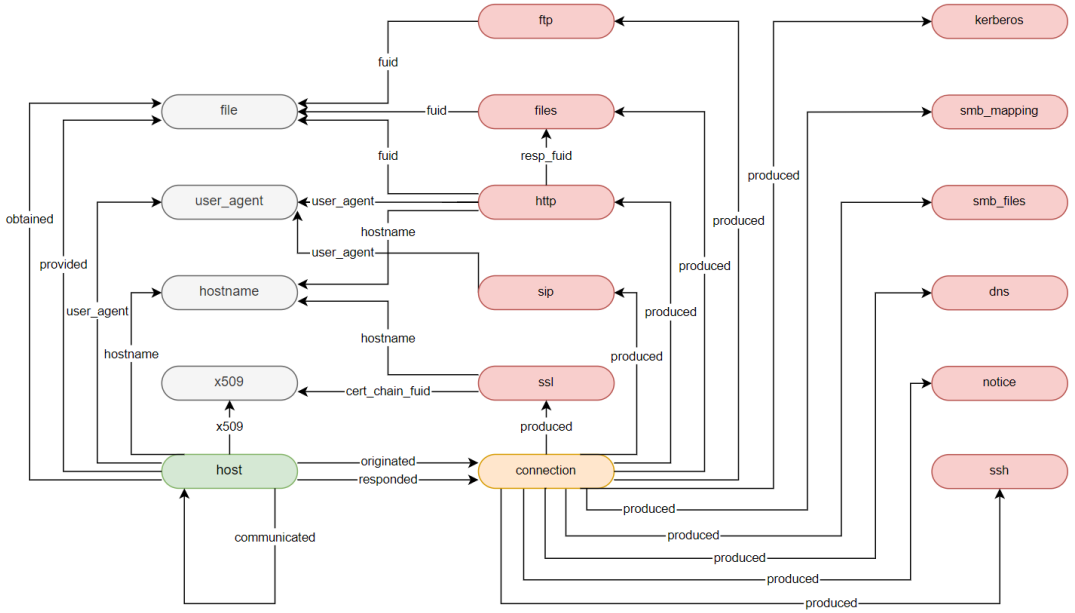
- Host – a device with IP address observed in the network traffic capture.
- Host-data – data related to the host extracted from network traffic (hostname, certificate, …).
- Connection – information about individual network connections (statistics, flags, …).
- Application – application data extracted from the connection (DNS, HTTP, TLS, …).
- All edges should be directional to ease analysis, but reverse processing should be allowed too.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Graph-based Data Storage

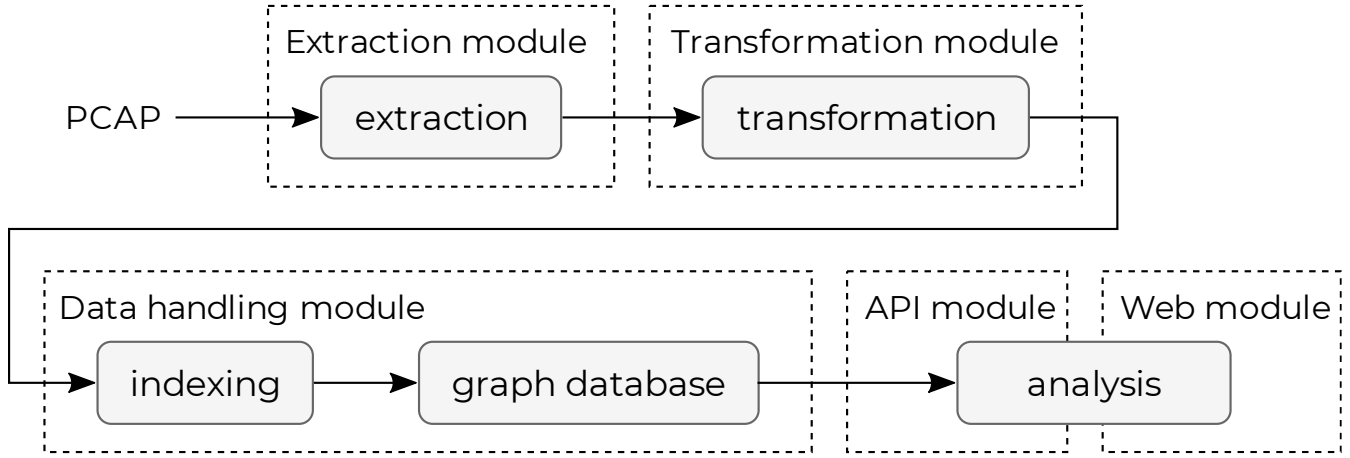
- Nowadays, we can observe the rapid development of various types of databases, including graph databases that allow us to store and analyze data in the form of associations efficiently.
- Graph database examples: Neo4j (https://neo4j.com/), Dgraph (https://dgraph.io/), …
- The graph-based approach is also used in GraphQL, an increasingly popular API query language.

- Utilization of a scalable database is necessary to store and analyze large-volume of network traffic data.
- For example, the dataset from the CyberCzech exercise with 330,564 connections results in 718,475 nodes and 397,632 edges.
- Current databases are better on ex-post analysis rather than continuous data storage.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Granef Toolkit



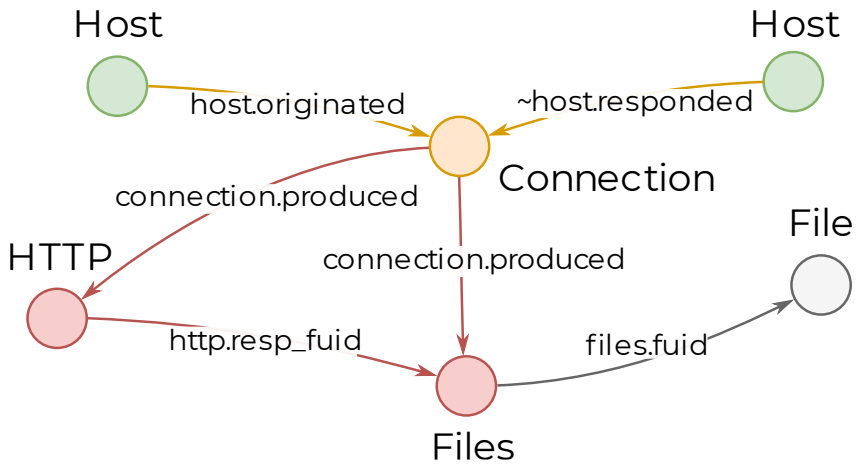
- The Granef toolkit demonstrates the presented approach to exploratory network traffic data analysis based on associations stored in a graph database.
- The toolkit's core consists of a scalable graph database Dgraph that stores transformed information from network traffic captures extracted by Zeek network security monitor.
- Modules are implemented as Docker containers.
- Custom Python scripts control all modules to ease toolkit setup and usage.
- Web interface visualizes data as an interactive relationship diagram.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Graph-based Analysis of Network Traffic Data

- Example of a DQL query (Dgraph Query Language), based on GraphQL, with a selection of TCP connections with a file transfer from a local network:

```
{getConn(func: allof(host.ip, cidr, "192.168.0.0/16"))
  name : host.ip
  host.originated @filter(eq(connection.proto, "tcp"))
    expand(Connection)
    connection.produced {
      expand(_all_)
      files.fuid { expand(File) }
    }
    ~host.responded { responded_ip : host.ip }
  }
}}
```

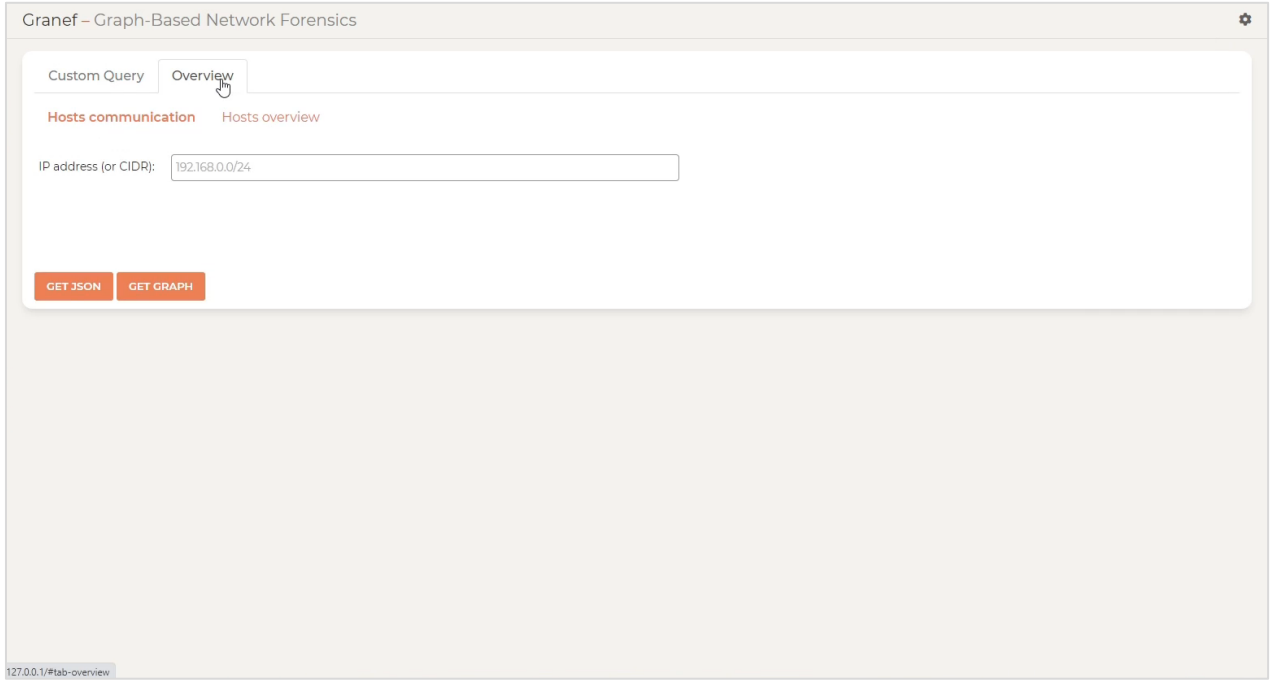


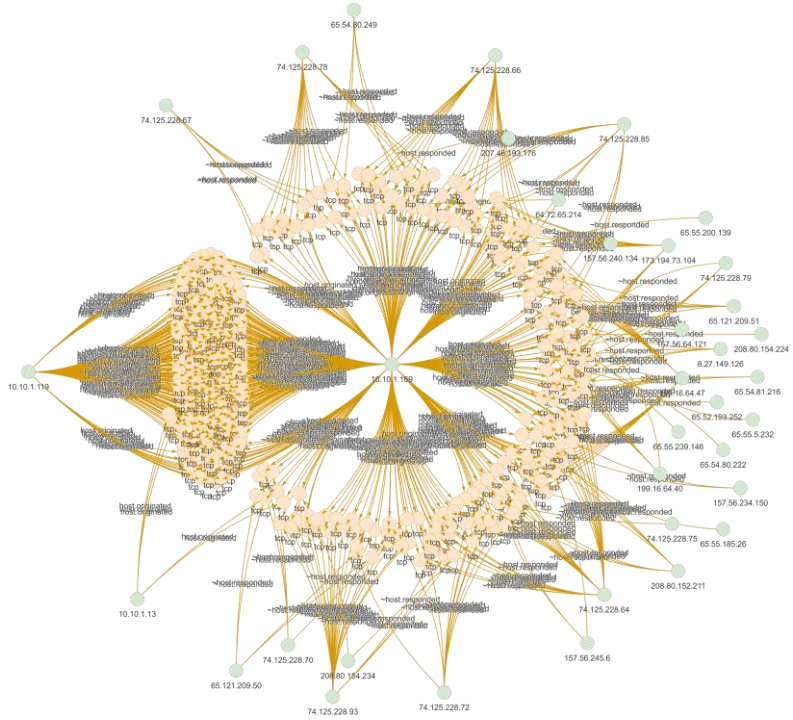
- The toolkit contains an abstract layer API with common analysis functions to ease data investigation (e.g., node neighbors' discovery, data filtering, connections overview).
- Results are provided as JSON or visualized in an interactive relationship diagram.
- Visualization uses a force-directed graph layout and allows nodes aggregation to show large relationship diagrams while preserving a simple overview of the data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Interactive Exploration

- The interactive relationship visualization allows the analyst to get details about any selected node, go into the graph's depth, and gain new observations.
- Various types of attacks and anomalies can be spotted at first glance based on visual patterns.





This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Connection of Other Data Sources

The proposed data representation allows us to easily connect other data sources and analyze them together with network traffic data within a unified environment.

- Alerts – Anomalies and attacks observed by Intrusion Detection Systems can be associated either to a relevant Host node or to a specific connection.
- OSINT – Data from OSINT sources can be linked to any graph node to provide a broader context.
- Host data (logs, EDR, …) – Similar graph-based representation can also be used for host-based data that can be further connected to network traffic data on a connection level.

GraphQL API allows us to obtain data from external sources directly in a format suitable for connection to a graph database. If we do not want to disturb the original data, attaching the external information only within the interactive visualization is also possible.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Conclusion

- Graph-based analysis follows the typical way of human thinking and perception of the characteristics of the surrounding world.
- The presented approach is not only the new method of network data storage and analysis, but it is also a shift of mindset that allows us to perceive network traffic in a new way.
- We have introduced the Granef toolkit to demonstrate exploratory network traffic analysis based on associations stored in a graph database.
- The same approach can be applied to other data types (IP flows, logs, EDRs, etc.).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418

Granef

We are currently finalizing the initial version of the graphical analysis environment, so the Granef toolkit is not publicly available now. But, if you are interested, just send me an email (cermak@ics.muni.cz) and I will be happy to provide it to you.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833418