Toward Graph-Based Network Traffic Analysis and Incident Investigation

DFRWS-EU 2022 – Short Presentations

Milan Cermak

Masaryk University, Brno, Czech Republic

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
- Graph data visualization allows us to get a broader context of the 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

Requirements for Network Traffic Analysis and Incident Investigation

Evaluation of Common Tools

When some serious incident happened in the 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?

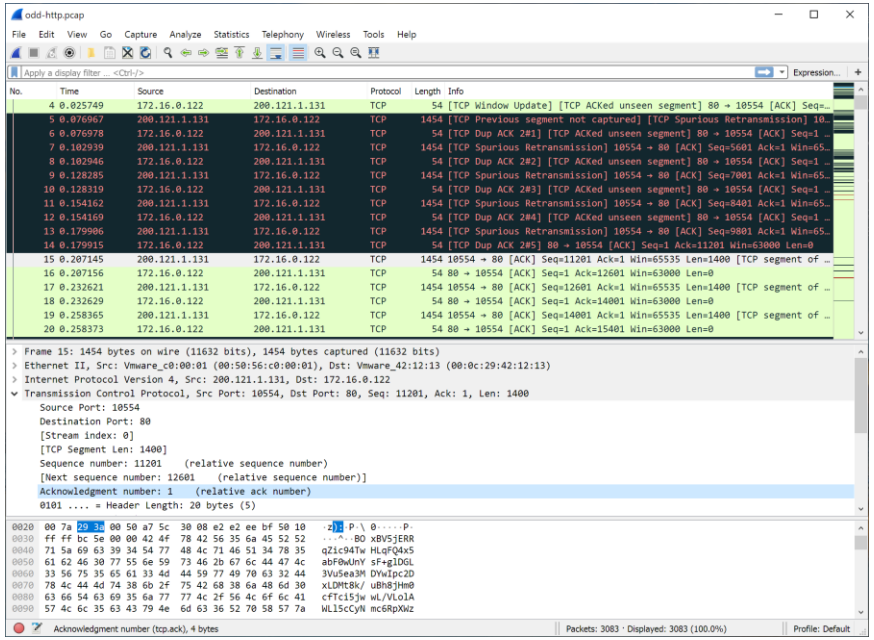
To understand the capabilities of today's tools, we have utilized CyberCzech dataset, selected an initial alert, and investigated it using **Wireshark**, **Arkime**, and **Brim**

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



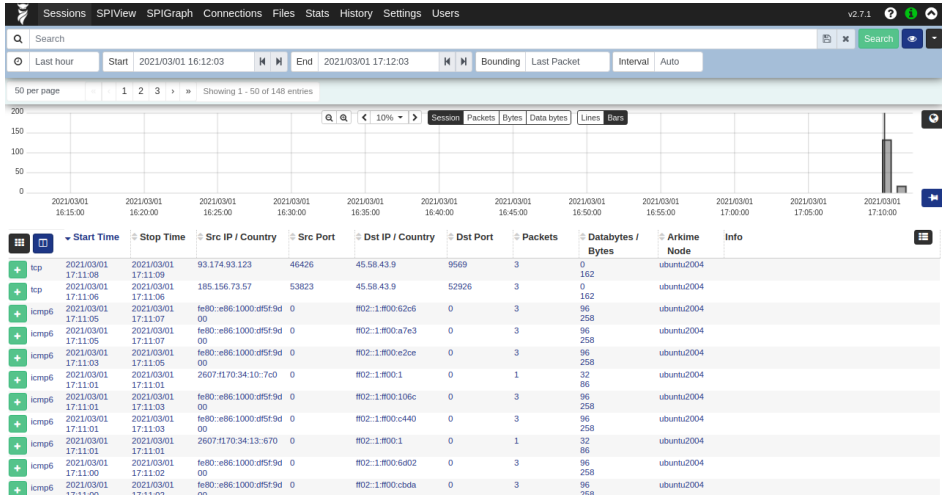
Wireshark: https://www.wireshark.org/

Arkime (formerly 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



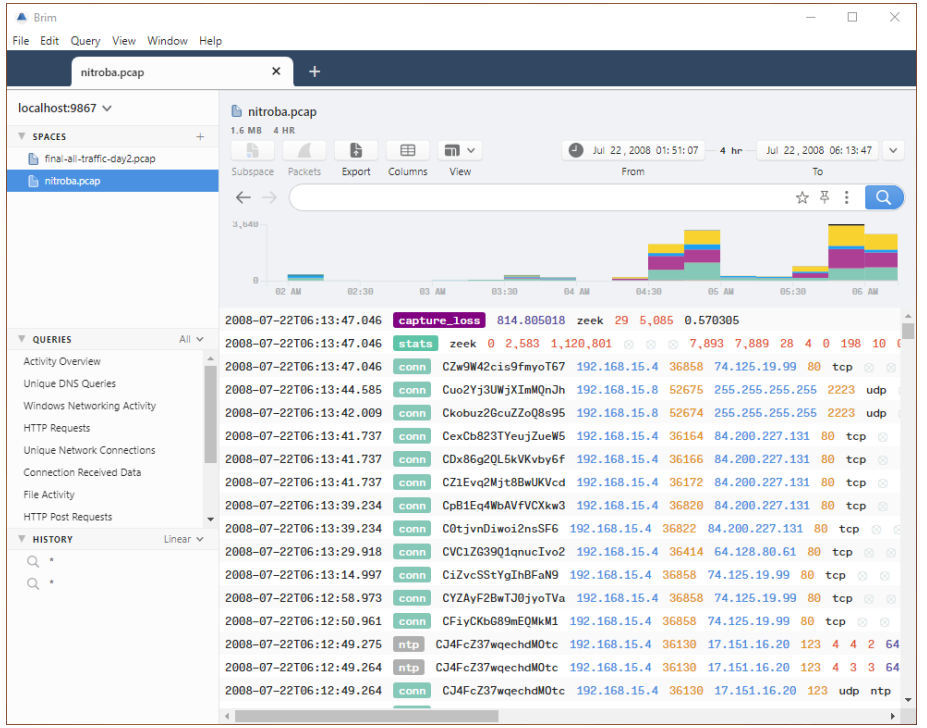
Arkime: https://arkime.com/

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

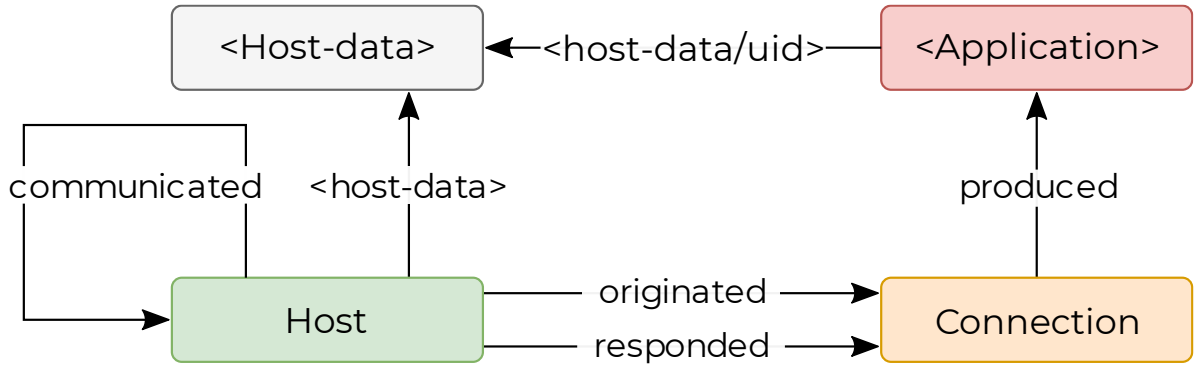
- Custom query language
- Limited visualizations
- Limited connection to other information sources



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

Graph-Based Analysis of Network Traffic Data

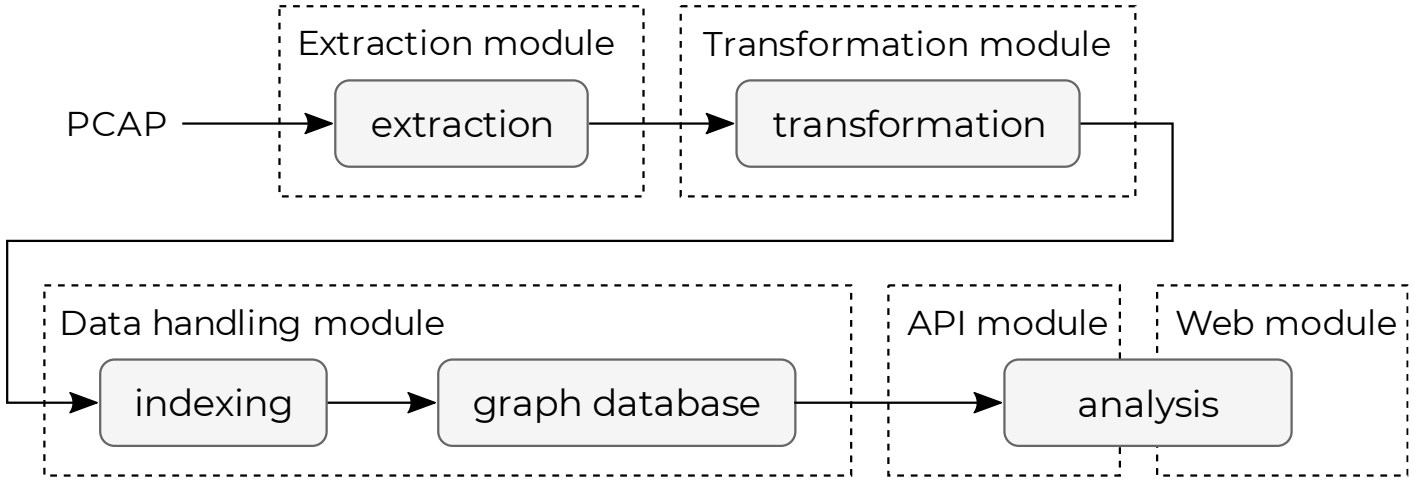
Representation of Network Traffic Data



- Initial version was proposed by Niese and further developed by Leichtnam et al.
- We have further developed these proposals and simplified them to ease data understanding

- 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 could be possible

Granef Toolkit

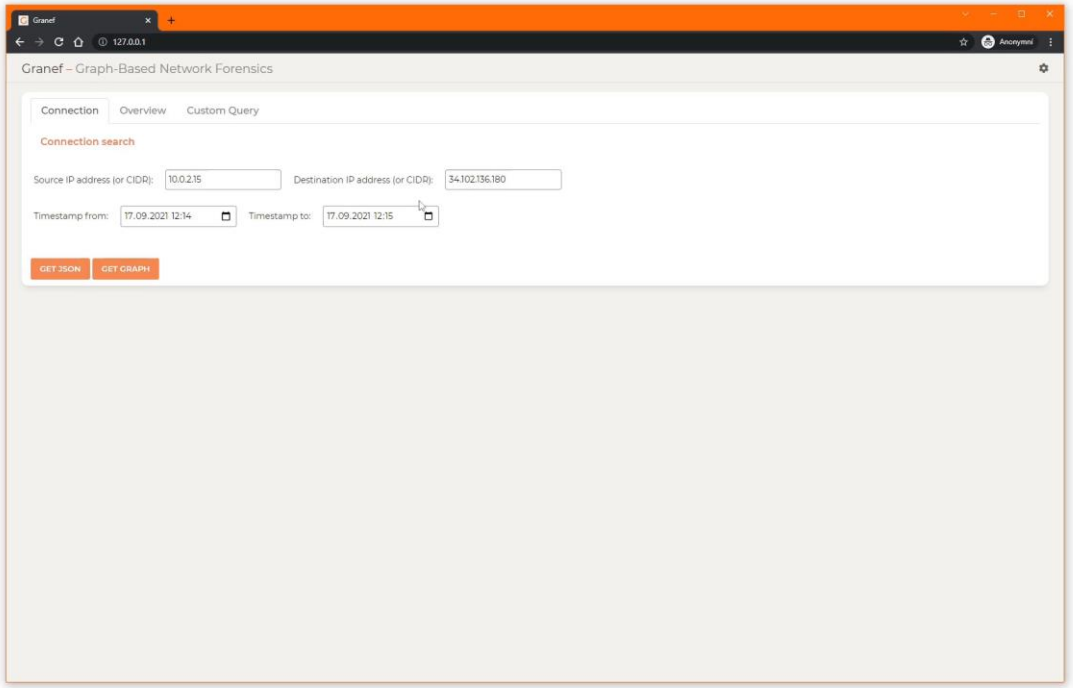


- Demonstration of the new approach to exploratory network traffic data analysis based on associations stored in a graph database
- The toolkit's core consists of a scalable graph database Dgraph that stores transformed information from network traffic extracted by Zeek network security monitor
- Modules 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

Interactive Data Exploration

- The analyst can use predefined queries or custom DQL queries (Dgraph Query Language)
- 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

```
{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 }
  }
}}
```



Towards Unified Analysis of Data Related to the Incident Investigation

Unified Data Analysis

General incident information and OSINT data can be simply represented as graph data (as it is possible, for example, in the Maltego tool)

- Modern GraphQL API allows us to obtain various data from external sources directly in a format suitable for connection to a graph database
- It is also possible to link other primary data sources, such as alerts, host data, threat intelligence, etc. (if they are represented as a graph)
- The proposed approach enables us to define queries and interactively browse various data in a single graphical environment:

  check the process tree that originated analyzed connection; search for IOCs and linked data; investigate network communications, disk records, account transactions, and phone calls in a single database query

This represents the main goal of our research!

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 an open-source Granef toolkit to demonstrate exploratory network traffic analysis based on associations stored in a graph database: https://granef.csirt.muni.cz
- The modern GraphQL API offers great potential to connect various data related to an incident investigation

Check granef.csirt.muni.cz to get more information about Granef and our research!

Feel free to contact me also at cermak@ics.muni.cz

Sharing and Automation for Privacy Preserving Attack Neutralization

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