Automated DFIR in Windows operating system

Marková, E., Krišáková, S. P., Sokol, P.
Pavol Jozef Šafárik University, Faculty of Science, Košice, Slovakia

Introduction

An important aspect of digital forensics data research involves creating datasets that meet specific expectations and requirements. Generally, there is no single dataset suitable for all research purposes in the field of digital forensics [1, 2]. Researchers encounter various challenges when using, creating, and sharing datasets. For our research, we require datasets that depict real-world scenarios encountered in security incidents. The main aim is to develop a suitable dataset for comparing methods in digital evidence analysis, which can be applied to investigate different issues.

Aims of the automated DFIR

- find **relevant digital evidence** (outlier detection methods),
- find **relationships between digital evidence** (graph theory, formal concept analysis, clustering),
- find **relationships between evidence's attributes**,
- and others.

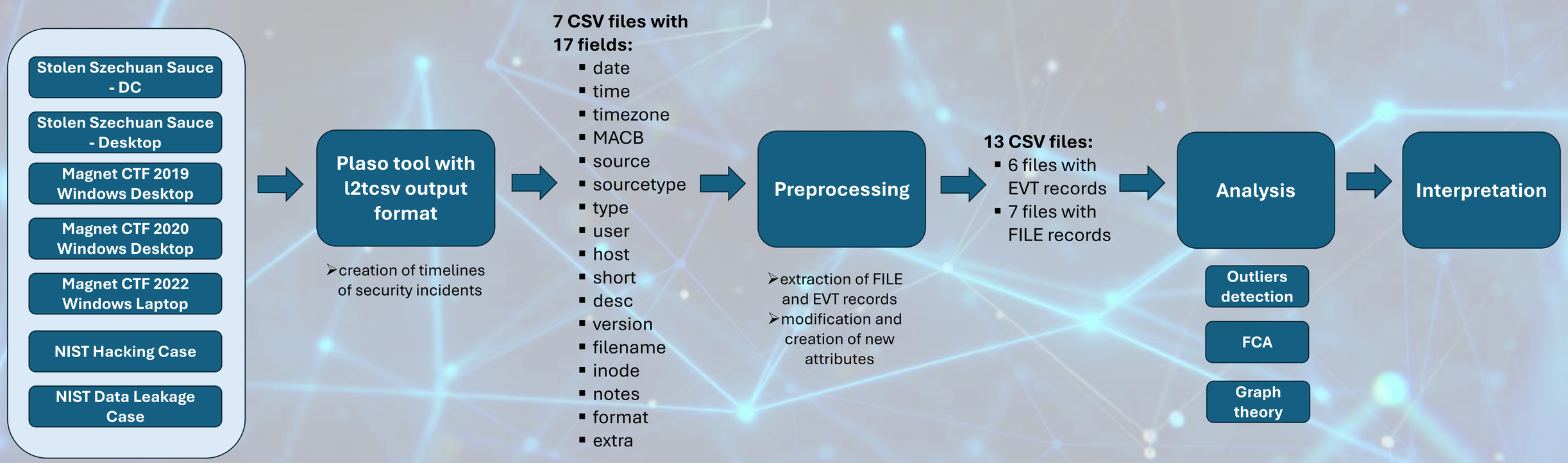
Datasets

The created datasets contain records from the NTFS file system and event logs. We utilized the following datasets:

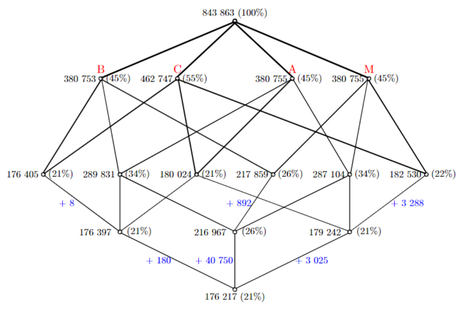
- Stolen Szechuan Sauce DC[1],
- Stolen Szechuan Sauce Desktop[2],
- Magnet CTF 2019 Windows Desktop[3],
- Magnet CTF 2020 Windows Desktop[4],
- Magnet CTF 2022 Windows Laptop[5],
- NIST Data Leakage Case[6], and
- NIST Hacking Case[7].

1. https://dfirmadness.com/case001/DC01-E01.zip
2. https://dfirmadness.com/case001/DESKTOP-E01.zip
3. https://digitalcorpora.s3.amazonaws.com/corpora/scenarios/magnet/2019%20CTF%20-%20Windows-Desktop.zip
4. https://digitalcorpora.s3.amazonaws.com/corpora/scenarios/magnet/2020%20CTF%20-%20Windows.zip
5. https://digitalcorpora.s3.amazonaws.com/corpora/scenarios/magnet/2022%20CTF%20-%20Windows.zip
6. https://cfreds-archive.nist.gov/data_leakage_case/data-leakage-case.html
7. https://cfreds.nist.gov/all/NIST/HackingCase

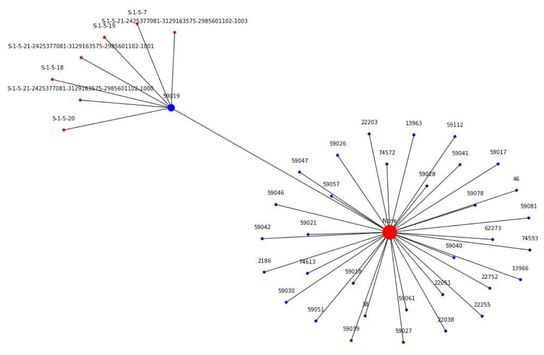
Scheme



Approaches to DFIR



Association rule	Confidence	Support	Behaviour
{M, C, B} => {A}	100%	20.88%	Expected
{C, B} => {A}	99.99%	20.90%	Suspicious
{A, C, B} => {M}	99.90%	20.90%	Suspicious
...	...	...	...
{M} => {A}	75.40%	45.12%	
{} => {C}	54.84%	100%	



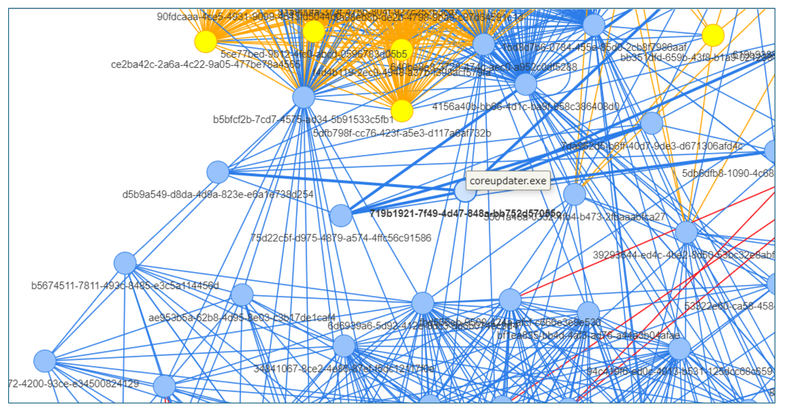
In our research [3, 4] we have used the formal concept analysis on datasets to find relationships between digital evidence and their attributes. On the left figure is the MACB concept lattice shown with 15 vertices representing formal concepts and 28 edges.

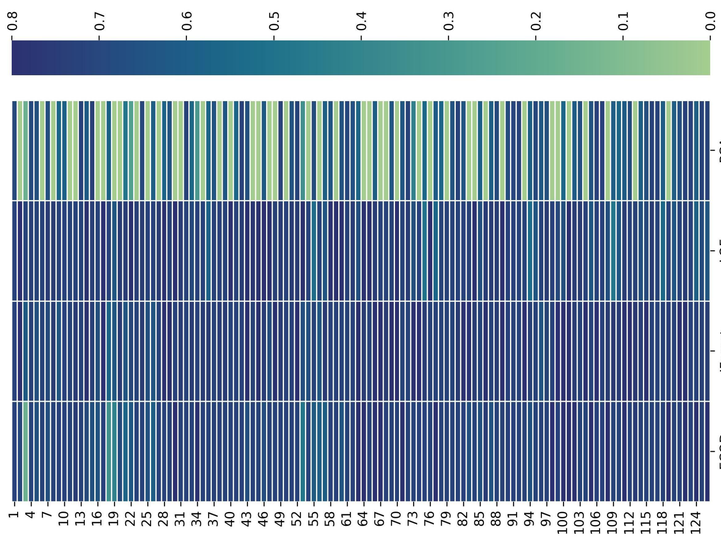
Association rules take statistical relevance into account. In left table, we present the rules for MACB attributes. The expected behavior group emphasizes operating system standards. Records with 90 - 100% confidence are interesting for digital forensics.

The figure below shows a graph from the NIST Data Leakage Case - EVT dataset. The red nodes represent user_sid and the blue nodes represent inode.

Figure on the right shows the visualization of the FILE records from The Stolen Szechuan Sauce – DC dataset. Edges and vertices originating from the relation of formal concept analysis are marked in yellow, edges and vertices belonging to relations from inodes are shown in red, and vertices and edges originating from relations based on the name attribute are shown in blue.

In the figure on the right we display results of outlier detection methods (ECOD, Isolation forest, Local Outlier Factor, Principal Component Analysis) for different combinations of attributes. We can see a heatmap of the maximum F1 Score for detection methods for file inodes. As we can see, the results for PCA are generally unsatisfactory.





Literature

[1] Grajeda, Cinthya, Frank Breitinger, and Ibrahim Baggili. "Availability of datasets for digital forensics–and what is missing." Digital Investigation 22 (2017): S94-S105.
[2] Luciano, Laoise, et al. "Digital forensics in the next five years." Proceedings of the 13th International Conference on Availability, Reliability and Security. 2018.
[3] Sokol, Pavol, Marková, Eva, et al. "The analysis of digital evidence by Formal concept analysis." The 16th International Conference on Concept Lattices and Their Applications (CLA 2022). 2022.
[4] Sokol, Pavol, Marková, Eva, et al. "Formal concept analysis approach to understand digital evidence relationships." International Journal of Approximate Reasoning 159 (2023): 108940.

Digital Security Progress. Protected.

The research is conducted and funded in cooperation with ESET company.

Cyber Security Lab