Lifecycle of Automated Digital Forensic Analysis Using Machine Learning

Krišáková, S. P., Sokol, P., Bača, M., Pavol Jozef Šafárik University, Faculty of Science, Košice, Slovakia
Slíž, R., Mikuláš, M., IstroSec, s.r.o, Bratislava, Slovakia

Motivation & Research Gap

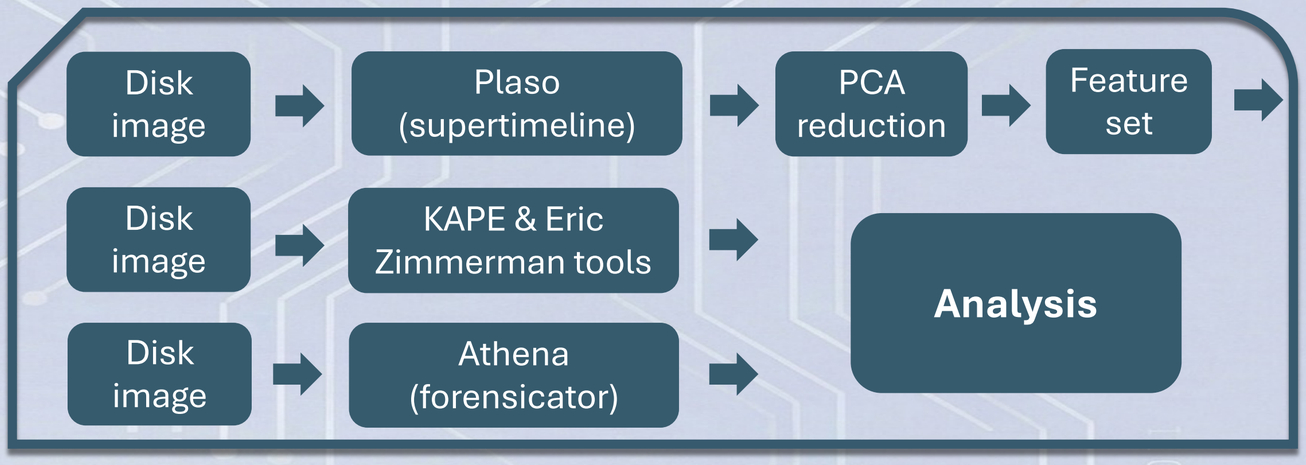
- Growing volume and heterogeneity of forensic artifacts make manual analysis impractical.
- Existing tools produce large supertimelines but lack integrated ML analytics.
- Current research applies isolated techniques without a unified lifecycle.

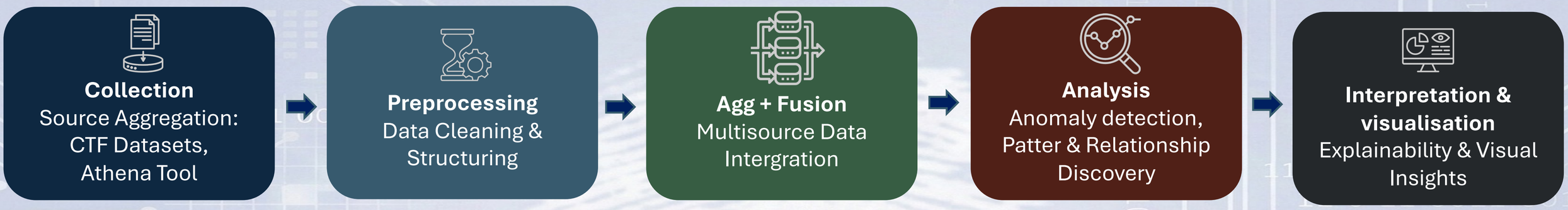
Our Contribution

- End-to-end ML-driven forensic lifecycle.
- Structured feature engineering from supertimeline data.
- Integration of anomaly detection, FCA, sequence mining, & time-series analysis.
- Validation on CTF datasets and a simulated APT scenario.

Dataset sources

- **CTF data (digital forensics)**
  - Stolen Szechuan Sauce DC and Desktop, Magnet CTF 2019 / CTF 2022 Windows Desktop, NIST Data Leakage Case
- **Realistically simulated APT-like attack**
  - attacker post-exploitation activities / persistence mechanisms / privilege escalation / ...





Aggregation

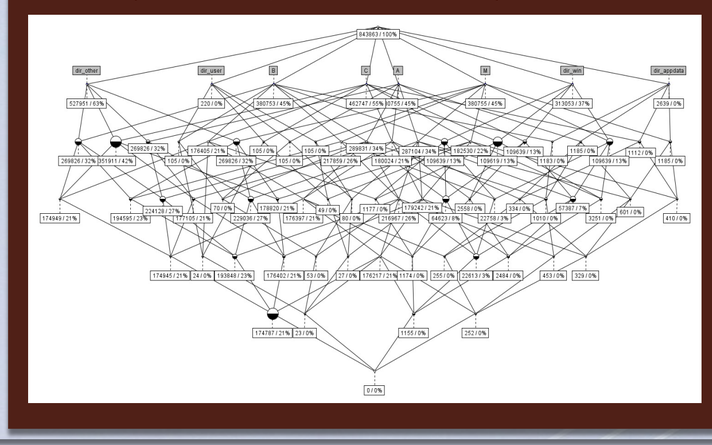
- by time frames, users, processes, inodes, ...
- max, min, sum, mean, quantiles, z-score, ...
- N/A values problem

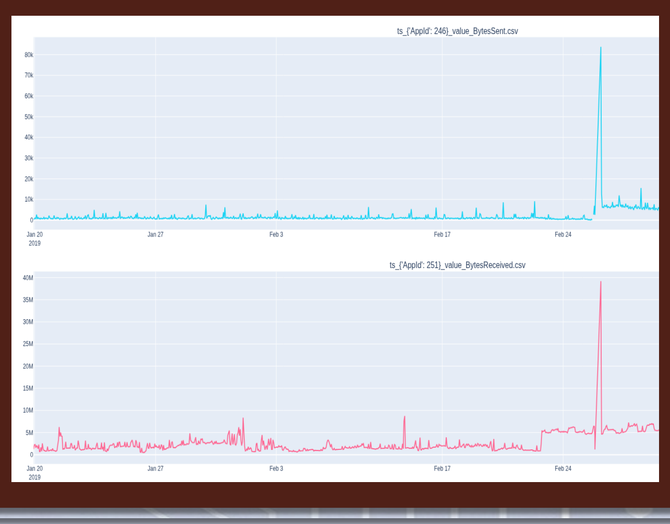
Fusion - Use Cases

- Local login / RDP login/logout
- File access/delete/download/timestomping
- Executable/Archive creation/ ...

Analysis approaches

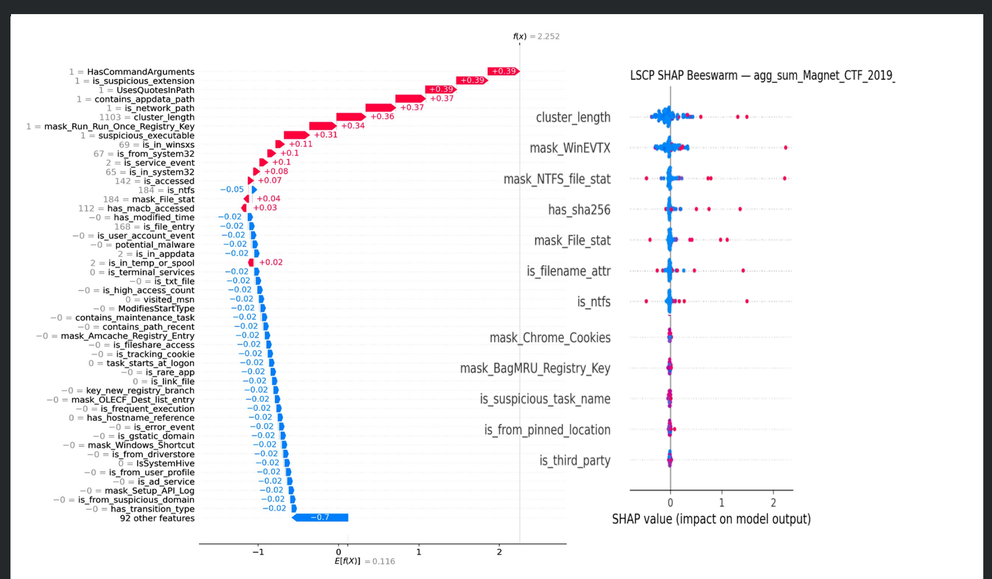
- Anomaly detection – ML methods
- Relationships between digital evidence – formal concept analysis
- Behavioral patterns – pattern sequence mining
- Time frames – time series analysis (statistical, neural networks)





Intepretation and visualisation

- explainability for anomaly detection
- features engineering



Conclusion

- ML-driven lifecycle improves scalability and consistency of digital investigations.
- Dimensionality reduction + structured feature engineering enables effective anomaly discovery.
- Multi-perspective analysis provides richer behavioral insight.

Future Work

- Evaluation on large-scale real-world enterprise datasets.
- Integration with SIEM/IR platforms.
- Benchmarking against state-of-the-art forensic ML approaches.
- Automated explainability reports for investigators.

This research was funded by the EU NextGenerationEU through Recovery and Resilience Plan for Slovakia under project No. 09-I05-03-V02-00079 (Automatization of digital forensics and incident response - ADFIR).

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU
VAIA RESEARCH AND INNOVATION AUTHORITY
VÝSKUMNÁ AGENTÚRA