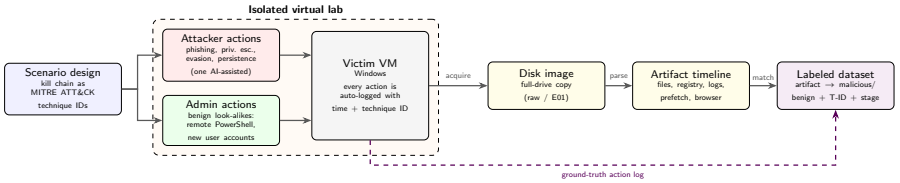
ForenDisk: A MITRE ATT&CK-Labeled DFIR Dataset with AI-Assisted Attacks



When a computer is hacked, digital forensics incident response (DFIR) investigators may copy its entire hard drive (a disk image) and search it for traces of the attacker, such as dropped files, registry changes and log entries. To train machine learning algorithms, we need disk images where we know which traces belong to an attack. Such ground truth is rare: real incident images are confidential, and public datasets mostly contain logs or network traffic. For example, DARPA's *Operationally Transparent Cyber* (OpTC) data release [1] records endpoint telemetry from hundreds of Windows hosts during a red-team campaign mixed with benign activity, and Unraveled [2] emulates a multi-week advanced persistent threat (APT) campaign alongside simulated users behaviour. CasinoLimit [3] further shows how such attack data can be labeled with MITRE ATT&CK techniques. None of them, however, provides labeled artifacts on disk or benign activity that deliberately looks malicious (like admin actions). As a first step, the student will extend these works with a literature review of attack datasets, labeling methods, and AI-assisted attacks, to position the thesis and refine its scenarios.

The goal of this thesis is to build a small, reproducible dataset of labeled disk images containing realistic attacks. In an isolated virtual lab with a single Windows victim, the student will run 3 or more attack scenarios, each written as a chain of a MITRE ATT&CK techniques, reusing open tools. One of the chain will include an AI-assisted step, such as an LLM-written phishing lure. In addition, the student will script legitimate administrator sessions (e.g. remote PowerShell) that resemble attacks, as realistic false positives.

Every action is logged automatically with its timestamp, technique ID, and its corresponding events in the timeline. After each scenario, the student acquires a disk image, extracts a timeline with an open-source tool (e.g. Plaso), and matches artifacts to the action log to label them as malicious or benign with their technique. Deliverables are the disk images, labeled timeline and regeneration scripts.

[1] R. Arantes, C. Weir, H. Hannon, M. Kulseng, "Operationally Transparent Cyber (OpTC)," DARPA data release, IEEE DataPort, 2021, doi:10.21227/edq8-nk52.

[2] S. Myneni et al., "Unraveled – A Semi-Synthetic Dataset for Advanced Persistent Threats," *Computer Networks*, vol. 227, 2023.

[3] S. Kilian et al., "CasinoLimit: An Offensive Dataset Labeled with MITRE ATT&CK Techniques," RAID 2025.

Practicalities

Promotor:	Bart Preneel		
Daily supervisor:	Félix INCERTI	fincerti@esat.kuleuven.be	office B01.20
	Alireza Aghabagherloo	alireza.aghabagherloo@esat.kuleuven.be	office B01.27
	Svetla Petkova-Nikova	svetla.nikova@kuleuven.be	office B01.14

Nature of the work:	25% literature, 10% theoretical work, 65% software and experiments
Number of students:	1 or 2