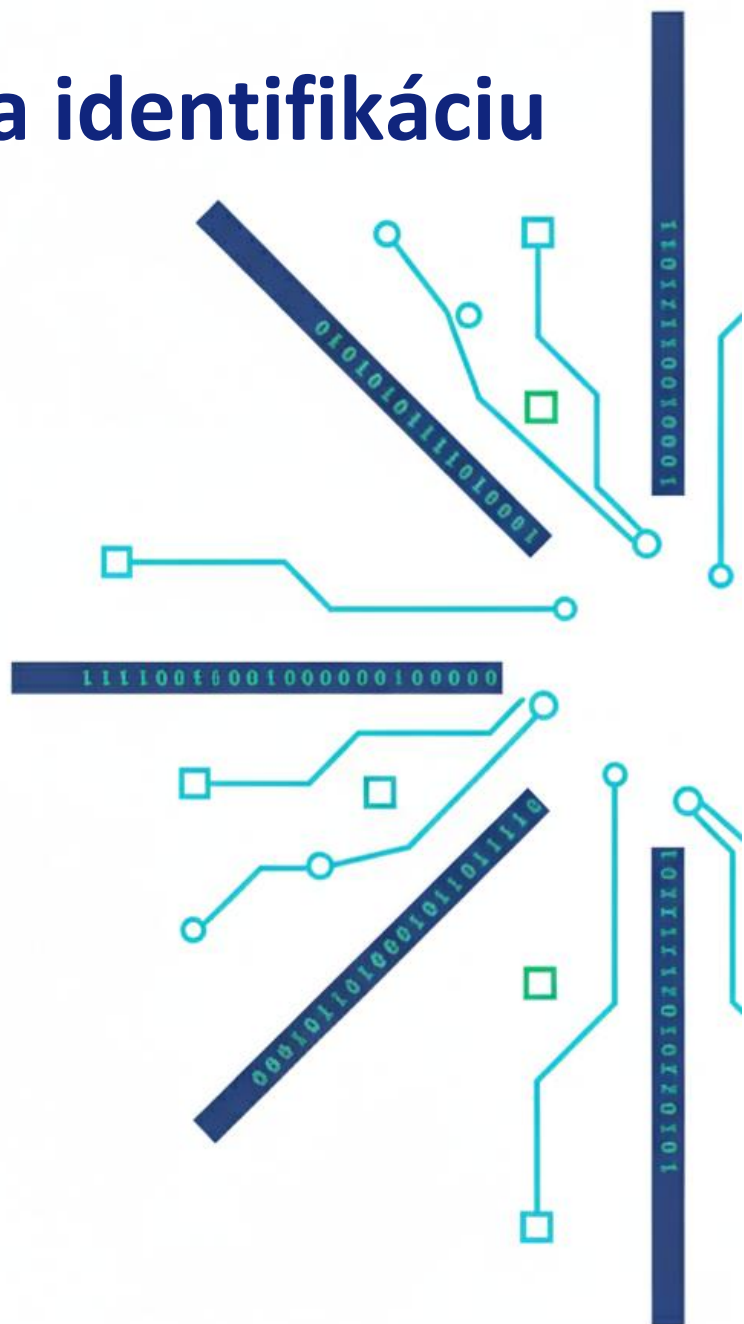


D19 – Metóda na identifikáciu digitálnych stôp



Projekt Automatizácia digitálnej forenznej analýzy a odpovede na incident (ADFIR) financovaný Európskou úniou – Next GenerationEU prostredníctvom Plánu obnovy a odolnosti Slovenskej republiky pod číslom projektu č. 09I05-03-V02-00079.

OBSAH

Popis projektu	3
Úvod	4
1 Úvod do zhlukovania	5
1.1 Zhlukovacia analýza v digitálnej forenznej vede.....	5
1.2 Hierarchický clustering a metodika vyhodnocovania	6
1.2.1 Forenzne metriky zhlukov	6
1.2.2 Validačné metriky pre určenie počtu zhlukov	7
1.2.3 Hľadanie anomálií pomocou zhlukovania	8
2 Aplikácia vybraných metód	9
2.1 Identifikácia digitálnych stôp pomocou clusteringu.....	9
2.1.1 Dátové súbory a ich predspracovanie	9
2.1.2 Hierarchické zhlukovanie	11
2.1.3 Vysvetliteľnosť výsledkov.....	12
2.1.4 Experimentálne výsledky	12
2.2 Identifikácia digitálnych stôp pomocou veľkých jazykových modelov	15
2.2.1 Dátové súbory a ich predspracovanie	15
2.2.2 Analýza pomocou veľkého jazykového modelu	16
2.2.3 Vyhodnotenie úspešnosti	17
2.2.4 Interpretovateľnosť výsledkov	17
Záver	18
Bibliografia.....	20

Popis projektu

Projekt **Automatizácia digitálnej forenznej analýzy a odpovede na incident** (ďalej len „ADFIR“) je financovaný Európskou úniou – Next GenerationEU prostredníctvom Plánu obnovy a odolnosti Slovenskej republiky pod číslom projektu č. 09I05-03-V02-00079. Tento projekt sa zaoberá jednou z kľúčových výziev v oblasti kybernetickej bezpečnosti a informačnej bezpečnosti – ako spracovať obrovské množstvo digitálnych dôkazov, ktoré vznikajú počas incidentov kybernetickej bezpečnosti alebo forezných vyšetrovaní. V súčasnosti je tento proces veľmi náročný z hľadiska ľudských zdrojov a času. Automatizácia pomocou metód strojového učenia môže preto výrazne **zlepšiť kvalitu digitálnej forenznej analýzy** a skrátiť čas potrebný na jej vykonanie. Celkovo to umožňuje bezpečnostným tímom efektívnejšie reagovať na kybernetické hrozby. Hlavné prínosy tohto projektu sú:

- **Rýchlejšie riešenie incidentov v oblasti kybernetickej bezpečnosti.** Projekt ADFIR zavádza automatizované prístupy k zberu, spracovaniu a analýze digitálnych stôp. Vďaka tomu môžu bezpečnostné tímy rýchlejšie identifikovať príčiny incidentov a prijať účinné opatrenia na ich riešenie.
- **Zníženie pracovnej záťaže forezných analytikov.** Rutinné a časovo náročné úlohy spojené so spracovaním digitálnych stôp budú nahradené automatizovanými metódami. To umožní analytikom sústrediť sa na zložitejšie prípady a strategické rozhodovanie.
- **Vyššia kvalita a konzistentnosť výstupov.** Použitie jednotných metodík a nástrojov zaručuje, že spracované digitálne stopy budú presnejšie, konzistentnejšie a ľahšie overiteľné. To výrazne znižuje riziko chýb spôsobených ľudskými faktormi.
- **Možné využitie v trestnom konaní.** Výstupy projektu budú vyvinuté v súlade s právnymi požiadavkami a normami, čo umožní, aby digitálne stopy boli akceptované ako relevantné dôkazy pre vyšetrovanie a súdne konania.

Úvod

V oblasti digitálnej forenznej analýzy predstavuje efektívne spracovanie veľkého množstva heterogénnych dát zásadnú výzvu, najmä pri rekonštrukcii udalostí a identifikácii kauzálnych súvislostí medzi jednotlivými artefaktmi. Moderné prístupy preto čoraz viac využívajú metódy strojového učenia a umelej inteligencie, ktoré umožňujú transformovať izolované forenzné stopy do analytikmi využiteľného modelu.

Tento dokument sa zameriava na návrh a implementáciu metódy identifikácie forenzných stôp v rámci projektu ADFIR (Úloha KPB4.1 – Identifikácia vhodných metód na identifikáciu digitálnych dôkazov súvisiacich s analyzovaným prípadom). Cieľom je vytvoriť model, ktorý umožní systematické spracovanie artefaktov, redukciu množstva záznamov, ich koreláciu naprieč rôznymi zdrojmi a následnú interpretáciu v kontexte identifikovaných scenárov správania systému alebo používateľa.

Kľúčovou časťou práce je spracovanie dát a výber vhodných modelov strojového učenia na identifikáciu digitálnych stôp v digitálnej forenznej analýze. Tento prístup umožňuje vyhľadávať digitálne stopy pomocou artefaktov a zároveň znížiť množstvo jednotlivých záznamov, ktoré musí analytik spracovať.

Navrhovaný model tak poskytuje systematický spôsob, ako prejsť od fragmentovaných dát k integrovanému pohľadu na digitálne stopy, pričom podporuje nielen rekonštrukciu udalostí, ale aj identifikáciu vzorcov správania relevantných pre forenznú analýzu.

1 Úvod do zhlukovania

V kontexte digitálnej forenznej analýzy predstavuje identifikácia digitálnych stôp komplexný proces, ktorého cieľom je odhalenie relevantných vzorov, vzťahov a odchýlok v rozsiahlych dátových množinách. Keďže charakter a štruktúra digitálnych stôp sa môžu výrazne líšiť v závislosti od zdroja dát a analyzovaného prostredia, neexistuje univerzálna metóda vhodná pre všetky typy analýz. Z tohto dôvodu sú v rámci tohto dokumentu predstavené viaceré metódy vhodné na identifikáciu digitálnych stôp z rôznych perspektív.

Prvou využitou metódou bolo hierarchické zhlukovanie, ktoré umožňuje identifikovať prirodzene sa vyskytujúce skupiny v dátach bez nutnosti vopred definovať presný počet výsledných klastrov, resp. zhlukov. Táto vlastnosť je v oblasti digitálnej forenznej analýzy významná, keďže počet a charakter podobných skupín digitálnych stôp nie je možné vopred spoľahlivo určiť. Metóda preto umožňuje odhaliť skryté štruktúry v dátach a identifikovať potenciálne súvisiace aktivity alebo udalosti.

Druhým skúmaným prístupom bola identifikácia digitálnych stôp pomocou veľkých jazykových modelov (Large Language Models – LLM). Tento prístup sa od tradičných metód strojového učenia odlišuje najmä schopnosťou pracovať s komplexnými textovými reprezentáciami digitálnych stôp a využívať znalosti získané počas predtrénovania modelu. V oblasti digitálnej forenznej analýzy umožňujú veľké jazykové modely analyzovať rozsiahle a heterogénne časové osi udalostí, identifikovať potenciálne bezpečnostne relevantné aktivity a zároveň poskytnúť ich kontextovú interpretáciu. Významnou výhodou tohto prístupu je možnosť pracovať aj s obmedzeným množstvom označených dát, čo je v prostredí digitálnej forenznej analýzy dôležité vzhľadom na náročnosť manuálneho označovania forenznych záznamov.

Významnou súčasťou využitia metód strojového učenia v oblasti digitálnej forenznej analýzy nie je iba samotná schopnosť identifikovať vzory, skupiny alebo odchýlky v dátach, ale aj možnosť interpretovať získané výsledky. V prostredí digitálnej forenznej analýzy predstavuje vysvetliteľnosť (explainability) jeden z kľúčových aspektov, keďže výsledky automatizovaných metód musia byť zrozumiteľné a overiteľné forezným analytikom. Model alebo metóda, ktorá identifikuje podozrivé správanie bez poskytnutia informácie o dôvodoch svojho rozhodnutia, môže mať v procese vyšetrovania obmedzenú využiteľnosť. Z tohto dôvodu sa v rámci navrhnutého prístupu venujeme aj identifikácii atribútov a charakteristík, ktoré najvýznamnejšie prispievajú k výslednej klasifikácii alebo vytvoreným skupinám digitálnych stôp.

1.1 Zhlukovacia analýza v digitálnej forenznej vede

V digitálnej forenznej analýze slúži zhlukovacia analýza (clustering) na automatickú organizáciu artefaktov, logov alebo udalostí do skupín na základe ich podobnosti. Tento prístup radikálne znižuje komplexnosť dát a umožňuje efektívnejšiu analýzu. Vzhľadom na to, že moderné systémy generujú pri forenznej analýze časovej osi milióny nízkoúrovňových

udalostí [1], clustering pomáha vyšetrovateľom redukovať objem dát na zvládnuiteľnú úroveň a zamerať sa na reprezentatívne skupiny namiesto analýzy jednotlivých záznamov [2].

Hlavnou výzvou digitálneho forenzného vyšetrovania je identifikácia relevantných dôkazov v obrovskom množstve irelevantných dát. Výskumy ukazujú, že kategorizácia na základe strojového učenia a zhukovania dokáže efektívne grupovať súbory, prioritizovať dôkazy (napr. v e-mailovej komunikácii) a významne skrátiť čas vyšetrovania [3][4].

Clustering je úzko prepojený s **detekciou anomálií** a **rekonštrukciou udalostí**:

- **Detekcia anomálií:** Udalosti (napr. autentifikačné logy) sú zhukované (napr. algoritmom MajorClust) a anomálie sa identifikujú na základe špecifických vlastností zhukov, ako sú ich veľkosť, frekvencia či časové charakteristiky [5][6][7].
- **Rekonštrukcia udalostí:** Tento proces býva často fragmentovaný. Zhukovanie pomáha organizovať udalosti do zmysluplných sekvencií a odhaľovať vzťahy medzi nimi, čo v kombinácii s procesným dolovaním (process mining) umožňuje extrahovať reálne vzorce správania aj z neúplných alebo manipulovaných dát [8][9][10].

V moderných systémoch sa clustering integruje s ďalšími technikami, ako je **analýza formálnych konceptov (FCA)** a dolovanie asociačných pravidiel na hlbšie pochopenie vzťahov medzi digitálnymi dôkazmi [11]. Pokročilé rámce využívajú aj veľké jazykové modely (LLM) pre interpretáciu časových osí [12] alebo kombinujú redukciiu dimenzionality s clusteringom pre detekciu vnútorných hrozieb v reálnom čase [13].

1.2 Hierarchický clustering a metodika vyhodnocovania

V predkladanom prístupe aplikujeme **aglomeratívny hierarchický clustering** [14][15] na definované časové okná. Ide o prístup zdola-nahor (bottom-up), kde na začiatku každé časové okno tvorí samostatný zhuk. V každom ďalšom kroku sa zlúčia dva najbližšie zhuky, až kým nevznikne jeden finálny zhuk zahŕňajúci celý dataset.

Výsledkom je hierarchia zhukov, pre ktorú platí, že pre ľubovoľné dva zhuky A a B z rôznych úrovní platí práve jeden zo vzťahov:

$$A \cap B = \emptyset, A \subset B, \text{ alebo } B \subset A.$$

Jednotlivé aglomeratívne metódy sa líšia najmä spôsobom, akým po každom zlúčení aktualizujú maticu vzdialeností (dissimilarity measure) medzi zhukmi. V našom výskume nepredpokladáme, že zhuky musia tvoriť striktne po sebe idúce časové okná – zhuky sa môžu prekryvať a reprezentovať podobné správanie v rôznych častiach časovej osi.

1.2.1 Forenzné metriky zhukov

Ak sa podobné časové okná nachádzajú pred incidentom aj po ňom, zhuk reprezentuje bežné správanie systému. Ak však v zhuku dominujú okná z obdobia po prepuknutí

kybernetického incidentu, zhhluk považujeme za podozrivý. Na kvantifikáciu tejto relevancie zavádzame nasledovné metriky:

- **Attack Rate (miera útoku):** Podiel záznamov o útoku v rámci daného zhľuku.

$$attack\ rate = \frac{attack\ count}{cluster\ size}.$$

- **Attack Lift (nárast útoku):** Porovnanie miery útoku v zhľuku voči globálnej miere útoku v celom datasete.

$$attack\ lift = \frac{attack\ rate}{attack\ rate_{global}}.$$

- **Cluster Size Proportion (CSP - Podiel veľkosti zhľuku):** Vyjadruje relatívnu veľkosť i-teho zhľuku vzhľadom na celkový počet časových rámcov (N) v datasete.

$$CSP_i = \frac{cluster\ size_i}{N}.$$

1.2.2 Validačné metriky pre určenie počtu zhľukov

Pre určenie optimálneho rezu hierarchickým stromom (a teda optimálneho počtu zhľukov) využívame tri interné validačné metriky počítané paralelne, ktoré interpretujeme nezávisle:

1. **Silhouette score** [20]: Meria rovnováhu medzi kohéziou vo vnútri zhľuku a separáciou od ostatných zhľukov. Vyššia hodnota indikuje lepšie zhľukovanie.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

kde $a(i)$ je priemerná vnútrozhľuková vzdialenosť a $b(i)$ je minimálna priemerná vzdialenosť k objektom v iných zhľukoch.

2. **Davies–Bouldin score** [21]: Hodnotí pomer rozptylu vo vnútri zhľukov k vzdialenosti medzi ich centroidmi. Nižšia hodnota indikuje lepšie rozdelenie.

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{d(c_i, c_j)} \frac{\sigma_i + \sigma_j}{d(c_i, c_j)},$$

kde σ_i je rozptyl zhľuku a $d(c_i, c_j)$ je vzdialenosť centroidov.

3. **Calinski–Harabasz score** [22]: Kvantifikuje pomer disperzie medzi zhľukmi a disperzie vo vnútri zhľukov. Vyššia hodnota znamená lepšie definované zhľuky.

$$CH = \frac{Tr(B_k)/(k-1)}{Tr(W_k)/(n-k)},$$

kde B_k a W_k označujú matice disperzie medzi zhlukmi a vo vnútri zhlukov.

1.2.3 Hľadanie anomálií pomocou zhlučovania

Výsledky hierarchického zhlučovania je možné využiť nielen na identifikáciu skupín podobných digitálnych stôp, ale aj ako základ pre následnú detekciu anomálií. Keďže jednotlivé udalosti sú najskôr agregované do časových okien a následne zoskupené podľa podobnosti, anomálne správanie možno identifikovať analýzou vlastností vzniknutých zhlukov a ich vývoja v čase.

Prvým prístupom je hodnotenie vlastností jednotlivých zhlukov. Osobitná pozornosť je venovaná metrikám charakterizujúcim veľkosť a kompaktnosť zhlukov, napríklad ukazovateľu **Cluster Size Percentage** (CSP). Malé alebo izolované zhluky, ktoré sa výrazne odlišujú od dominantných skupín správania, môžu predstavovať kandidátov na bodové alebo kolektívne anomálie a indikovať výskyt neštandardných udalostí v analyzovanom systéme.

Druhým prístupom je sledovanie časového vývoja jednotlivých zhlukov. Ak sa určitý typ správania objaví až po prepuknutí bezpečnostného incidentu alebo sa jeho výskyt výrazne zvýši počas kompromitácie systému, môže ísť o anomálne správanie vzhľadom na historický normál. Na tento účel boli využité metriky **Attack Rate** a **Attack Lift**, ktoré umožňujú kvantifikovať vzťah medzi výskytom konkrétneho zhuku a analyzovaným incidentom. Zhluky s vysokými hodnotami týchto metrík predstavujú vhodných kandidátov na podrobnejšiu forenznú analýzu.

2 Aplikácia vybraných metód

V tejto kapitole sú predstavené praktické kroky aplikácie vybraných analytických metód na pripravené forenzné dátové sady. Cieľom experimentálnej časti je overiť využiteľnosť metód popísaných v predchádzajúcich kapitolách pri identifikácii významných vzorov v digitálnych stopách, detekcii odchýlok a následnej interpretácii získaných výsledkov. Jednotlivé metódy sú aplikované v nadväzujúcich krokoch – zoskupovanie podobných záznamov pomocou clusteringu a identifikácia digitálnych stôp pomocou veľkých jazykových modelov.

2.1 Identifikácia digitálnych stôp pomocou clusteringu

V rámci tejto časti bola skúmaná možnosť využitia metód zhľukovania (clusteringu) na identifikáciu podobných skupín digitálnych stôp v rozsiahlych forenzných datasetoch. Keďže v reálnych vyšetrovaniach často nie je vopred známe, aké typy aktivít alebo udalostí sa v analyzovaných dátach nachádzajú, boli využité metódy učenia bez učiteľa (unsupervised learning), ktoré umožňujú odhaliť prirodzene sa vyskytujúce štruktúry bez potreby existencie vopred definovaných tried.

Cieľom tejto analýzy nebolo nahradiť rozhodovanie forenzného analytika, ale znížiť množstvo manuálne kontrolovaných dát a podporiť identifikáciu potenciálne súvisiacich udalostí. Predpokladom je, že digitálne stopy patriace k podobným aktivitám alebo rovnakému typu správania budú v reprezentovanom dátovom priestore vykazovať podobné charakteristiky a vytvoria zoskupenia s vyššou vnútornou podobnosťou.

V rámci experimentov bolo využité hierarchické zhľukovanie, ktoré umožňuje vytvárať hierarchickú štruktúru vzťahov medzi objektmi bez nutnosti vopred stanoviť počet výsledných zhľukov. Tento prístup je vhodný najmä pre forenzné dáta, kde počet a charakter jednotlivých skupín udalostí nie je možné spoľahlivo určiť ešte pred samotnou analýzou.

2.1.1 Dátové súbory a ich spracovanie

Pre experimentálnu časť boli použité upravené datasety pochádzajúce zo súťaží typu Capture The Flag (CTF), ktoré sú zamerané na forenznú analýzu obrazov diskov operačného systému Windows. Tieto datasety obsahujú realistické scenáre používateľskej aktivity, systémových udalostí a simulovaných bezpečnostných incidentov.

Pôvodné obrazy diskov boli spracované s cieľom extrahovať relevantné forenzné artefakty a transformovať ich do štruktúrovanej reprezentácie vhodnej na následnú analýzu. Získané udalosti obsahujúce časovú informáciu boli následne organizované do časových okien.

V rámci tohto prístupu boli použité datasety:

- NIST Data Leakage Case [18],
- Magnet CTF 2019 [16],
- Magnet CTF 2022 [17].

Tieto datasety reprezentujú rôzne typy správania operačného systému vrátane bežnej používateľskej aktivity, administratívnych operácií a simulovaných bezpečnostných incidentov, ako napríklad neoprávnený prístup alebo únik dát. Použitie viacerých nezávislých datasetov umožňuje overiť robustnosť navrhnutého prístupu a znižuje riziko prispôsobenia metódy iba jednému špecifickému prostrediu.

Jednotlivé udalosti operačného systému boli rozdelené do časových okien s pevne definovanou dĺžkou 60 sekúnd. Každé časové okno reprezentuje stav systému v danom intervale a obsahuje agregované informácie zo všetkých digitálnych stôp zaznamenaných počas tohto obdobia.

Voľba 60-sekundového intervalu predstavuje kompromis medzi detailnosťou analýzy a množstvom výsledných časových okien. Kratší interval by síce poskytoval detailnejšiu reprezentáciu správania systému, avšak výrazne by zvýšil počet analyzovaných jednotiek. Naopak, dlhšie časové okná by mohli viesť k strate dôležitých detailov.

Počet vytvorených časových okien pre jednotlivé datasety bol nasledovný:

- NIST Data Leakage Case – 1 219 časových okien,
- Magnet CTF 2019 – 4 745 časových okien,
- Magnet CTF 2022 – 18 170 časových okien.

Jednotlivé nízkoúrovňové udalosti boli transformované na vyššiu úroveň reprezentácie popisujúcu správanie systému v čase. Na extrakciu časovo označených forenzných udalostí bol použitý framework Plaso, ktorý umožňuje získavať informácie z rôznych systémových zdrojov, najmä z metadátového systému NTFS. Bližšie popísané vo výstupe (ADFIR-D13-KPB2-02 Metóda na tvorbu datasetu).

Na základe extrahovaných artefaktov bolo vytvorených približne 200 atribútov reprezentujúcich rôzne typy systémovej aktivity, napríklad operácie so súbormi, zmeny metadát, prístupy k objektom alebo systémove udalosti.

Atribúty boli reprezentované diskretnými hodnotami:

- hodnota 1 označuje výskyt konkrétneho typu udalosti,
- hodnota 0 označuje absenciu udalosti,
- hodnota NaN označuje nedostupnú alebo chýbajúcu informáciu.

Pre agregáciu udalostí v rámci časových okien boli použité dve agregačné funkcie:

- sum – reprezentuje počet výskytov konkrétnej udalosti v danom časovom okne,
- max – reprezentuje iba informáciu, či sa daná udalosť v časovom okne vyskytla.

Tieto dve funkcie boli zvolené z dôvodu vhodnosti pre forenzné dáta obsahujúce chýbajúce hodnoty (NaN). Na rozdiel od štatistických metód založených napríklad na priemere alebo mediáne nie sú ovplyvnené absenciou niektorých typov udalostí.

Výsledkom predspracovania je štruktúrovaná časová reprezentácia systému, ktorá umožňuje následnú analýzu pomocou metód zhlukovania a detekcie anomálneho správania.

2.1.2 Hierarchické zhlukovanie

Na identifikáciu podobných vzorov správania systému bola použitá metóda hierarchického zhlukovania, konkrétne aglomeratívne zhlukovanie (*agglomerative clustering*).

Táto metóda využíva postup zdola nahor, kde na začiatku každý objekt predstavuje samostatný klastor. V jednotlivých krokoch sa následne spájajú najpodobnejšie dvojice klastrov, až kým nevznikne jeden výsledný klastor obsahujúci všetky objekty. Výsledkom je hierarchická štruktúra klastrov, ktorá umožňuje analyzovať vzťahy medzi skupinami na rôznych úrovniach podobnosti.

Výhodou tejto metódy v oblasti digitálnej forenznej analýzy je, že nevyžaduje vopred definovaný počet klastrov. Počet a charakter jednotlivých skupín digitálnych stôp totiž nie je možné pred samotnou analýzou spoľahlivo určiť.

V rámci analýzy sa nepredpokladá, že podobné časové okná musia nasledovať priamo za sebou v časovej osi. Rovnaký typ správania sa môže objaviť v rôznych častiach analyzovaného obdobia. Klastre, ktoré obsahujú časové okná pred aj po začiatku incidentu, pravdepodobne reprezentujú bežné systémové správanie. Naopak, klastre pozostávajúce prevažne z udalostí po začiatku incidentu môžu predstavovať potenciálne podozrivé správanie a predstavujú kandidátov na ďalšiu forenznú analýzu.

Pre vyhodnotenie významnosti jednotlivých klastrov sme použili metriky definované v časti **Forenzné metriky zhlukov** :

- attack rate,
- attack lift,
- cluster size proportion (CSP).

Pre určenie vhodného počtu klastrov boli použité tri validačné metriky spomenuté v kapitole **Validačné metriky pre určenie počtu zhlukov:**

- **Silhouette score**
- **Davies-Bouldin index**
- **Calinski-Harabasz index**

Jednotlivé riešenia zhlučovania boli hodnotené pomocou všetkých troch metrík samostatne, pričom nebol vytvorený jeden kombinovaný skórovací ukazovateľ.

Druhý krok analýzy využíva hierarchickú štruktúru klastrov na identifikáciu podobných nadradených skupín a hľadanie vzťahov medzi nimi. Tento postup poskytuje ďalšie možnosti pre forenzného analytika pri skúmaní digitálnych stôp.

2.1.3 Vysvetliteľnosť výsledkov

Poslednou časťou je zabezpečenie interpretovateľnosti výsledkov pre digitálneho forenzného analytika. Samotné vytvorenie klastrov neposkytuje dostatočnú informáciu o ich charaktere, preto je potrebné identifikovať atribúty, ktoré najvýraznejšie opisujú jednotlivé skupiny časových okien.

Interpretácia bola vykonaná v pôvodnom priestore atribútov, nie priamo v transformovanom priestore použitom počas zhlučovania. Zhlučovanie teda prebieha nad upravenou reprezentáciou dát, zatiaľ čo význam jednotlivých klastrov je určovaný spätnou analýzou pôvodných forenzných atribútov.

Pri hodnotení významnosti atribútov boli odstránené vlastnosti, ktoré nevykazovali žiadnu aktivitu ani v danom klastri, ani v celom datasete. Následne boli atribúty zoradené podľa absolútnej hodnoty štandardizovaného skóre (score std), ktoré umožňuje porovnávať význam atribútov s rôznymi rozsahmi hodnôt.

Tento prístup umožňuje identifikovať atribúty so zvýšeným výskytom (elevated) alebo naopak zníženým výskytom (suppressed) v rámci konkrétneho klastra. Výsledkom je interpretovateľný popis správania systému, ktorý môže slúžiť ako podpora pri následnom foreznom vyšetrowaní.

2.1.4 Experimentálne výsledky

Navrhnutý prístup bol experimentálne overený na datasetoch NIST Data Leakage Case, Magnet CTF 2019 a Magnet CTF 2022. Cieľom experimentov bolo overiť schopnosť hierarchického zhlučovania identifikovať skupiny časových okien reprezentujúce potenciálne podozrivé správanie a zároveň zachovať interpretovateľnosť výsledkov pre digitálneho forenzného analytika.

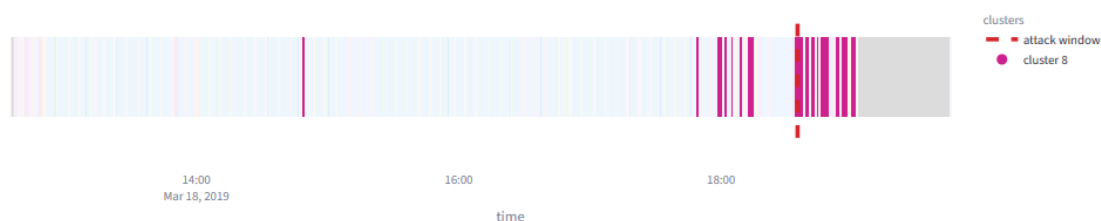
Počas experimentov boli analyzované rôzne rezy hierarchického stromu predstavujúce 10 až 50 klastrov. Ako najvhodnejšie riešenie sa ukázalo rozdelenie na 50 klastrov, ktoré poskytovalo vhodný kompromis medzi úrovňou detailu a interpretovateľnosťou výsledkov. Vyšší počet klastrov umožnil lepšie oddeliť špecifické formy správania systému a identifikovať menšie skupiny udalostí súvisiacich s bezpečnostným incidentom.

Na základe analýzy boli identifikované štyri základné typy klastrov:

- dominantné klastre reprezentujúce bežnú prevádzku systému,
- klastre obsahujúce výlučne incidentné správanie,
- klastre reprezentujúce obdobie pred incidentom,
- podozrivé klastre obsahujúce kombináciu bežného a incidentného správania.

Dominantné klastre tvorili významnú časť analyzovaných dát. Napríklad pri datasete Magnet CTF 2022 pokrýval najväčší klaster približne 72,6 % časových okien pri agregácii pomocou funkcie *sum* a približne 49,9 % časových okien pri agregácii pomocou funkcie *max*. Tieto klastre boli charakteristické atribútmi súvisiacimi s bežnou systémovou aktivitou operačného systému Windows.

Osobitnú pozornosť si zaslúžili tzv. incidentné klastre, pre ktoré bola hodnota metriky *Attack Rate* rovná 1. Takéto klastre obsahovali výhradne časové okná vzniknuté po začiatku bezpečnostného incidentu. V datasete Magnet CTF 2022 boli identifikované klastre obsahujúce udalosti súvisiace s vytvorením používateľského účtu alebo spustením podozrivého súboru využívajúceho mechanizmy perzistencie systému Windows. Analýza atribútov v rámci týchto klastrov umožnila rýchlo identifikovať digitálne stopy relevantné pre vyšetrovanie incidentu.



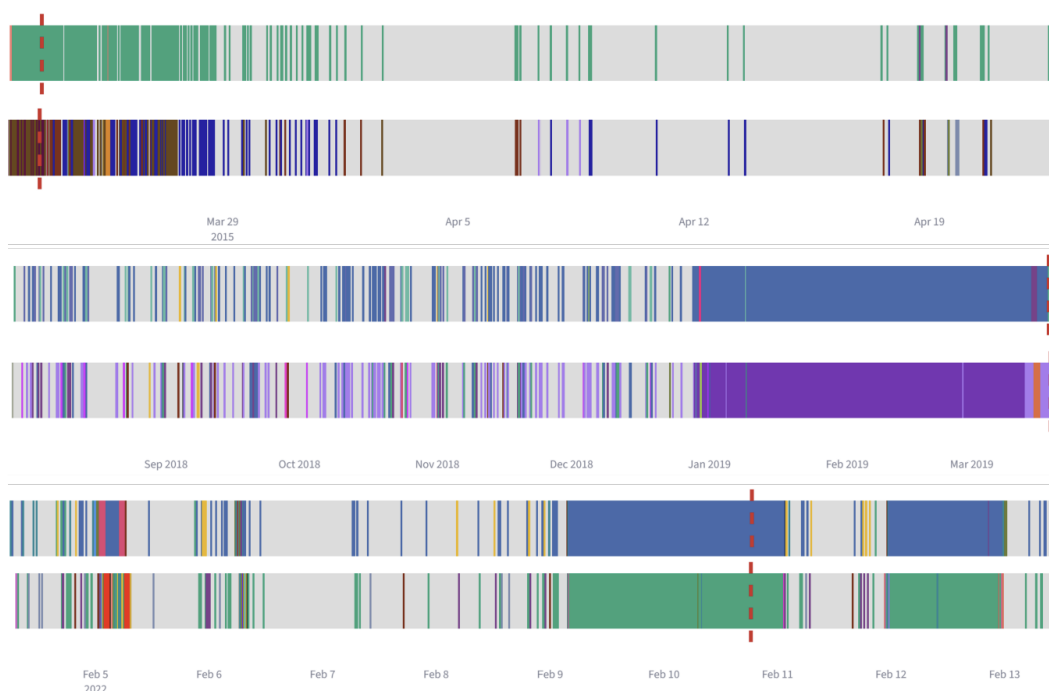
Obrázok 1: Klastre v Magnet 2019

Ako konkrétny príklad možno uviesť klaster 8 v datasete Magnet CTF 2019 pri použití agregácie pomocou funkcie *max*. Jeho časové rozloženie osciluje v okolí začiatku incidentu, čo naznačuje, že časové okná priradené k tomuto klastru sa vyskytujú pred aj po ručne definovanom začiatku incidentu. Takéto správanie môže poukazovať na udalosti súvisiace s počiatočnými fázami incidentu, prípadne na nepresnosť pri určovaní presného času jeho začiatku. Príklad tohto správania je znázornený na **Obrázok 1**.

Výsledky zároveň ukázali, že kombinácia metrík *Attack Rate*, *Attack Lift* a *Cluster Size Proportion* umožňuje efektívne odlíšiť bežné správanie systému od správania súvisiaceho s incidentom. Zároveň bolo možné identifikovať aj klastre nachádzajúce sa v blízkosti začiatku incidentu, ktoré môžu poukazovať na počiatočné fázy kompromitácie alebo na nepresné určenie času vzniku incidentu.

Výsledná štruktúra zoskupenia časových okien je vizualizovaná na **Obrázok 2**. Obrázok zobrazuje rozdelenie časových okien pre všetky tri analyzované datasety – NIST Data Leakage Case, Magnet CTF 2019 a Magnet CTF 2022 – pri použití agregácie pomocou funkcie *sum* a *max*. Jednotlivé farby predstavujú príslušnosť časového okna ku konkrétnemu klastru, pričom celkovo bolo vytvorených 50 klastrov. Červená prerušovaná čiara označuje ručne definovaný začiatok bezpečnostného incidentu a umožňuje porovnať časové rozloženie identifikovaných klastrov s priebehom incidentu. Z vizualizácie je možné pozorovať, že podozrivé správanie sa koncentruje v špecifických časových oblastiach, zatiaľ čo dominantné klastre pokrývajú väčšiu časť časových okien mimo incidentu.

Experimenty potvrdili, že navrhnutá metodika umožňuje významne znížiť množstvo dát, ktoré musí forenzný analytik manuálne preskúmať, a zároveň zachováva dostatočnú mieru vysvetliteľnosti potrebnú pre praktické využitie v digitálnej forenznej analýze.



Obrázok 2: Vizualizácia časových okien

2.2 Identifikácia digitálnych stôp pomocou veľkých jazykových modelov

Ďalším skúmaným prístupom bola identifikácia digitálnych stôp pomocou veľkých jazykových modelov (Large Language Models – LLM). Na rozdiel od klasických metód strojového učenia nebol v tomto prípade vytvorený model špecializovaný priamo na forenzné dáta, ale bol využitý existujúci predtrénovaný jazykový model, ktorého znalosti pochádzajú z rozsiahleho tréningu na všeobecných textových dátach. Cieľom bolo overiť schopnosť takéhoto modelu interpretovať komplexné časové osi udalostí operačného systému Windows a identifikovať digitálne stopy súvisiace s bezpečnostnými incidentmi.

Počas návrhu riešenia bolo experimentálne overených viacero architektúr založených na metódach strojového učenia. Prvým skúmaným prístupom bol LSTM autoenkodér, ktorý využíva učenie bez učiteľa a detekuje anomálie na základe rozdielu medzi pôvodnou a rekonštruovanou sekvenciou udalostí. Následne bol skúmaný prístup založený na architektúre Transformer typu BERT [19], ktorá prostredníctvom mechanizmu self-attention umožňuje efektívnejšie zachytávať vzťahy medzi jednotlivými udalosťami aj vo väčších kontextoch.

Oba prístupy však narazili na významné obmedzenia vyplývajúce zo špecifickej povahy digitálnych stôp. Forenzné časové osi obsahujú vysoký podiel bežných systémových udalostí, pričom bezpečnostné incidenty predstavujú iba malú časť dostupných záznamov. Okrem výraznej nevyváženosti dát predstavoval problém aj nedostatok kvalitne označených a dostatočne rozmanitých datasetov potrebných na vytvorenie robustného modelu schopného generalizovať na nové typy incidentov.

Na základe získaných poznatkov bol ako finálny analytický prístup zvolený predtrénovaný veľký jazykový model. V rámci experimentov boli testované otvorené modely Qwen [20] a Gemma [21], pričom ako výsledný model bol vybraný Qwen z dôvodu stabilnejších a presnejších výsledkov pri interpretácii forenzných časových osí. Model bol využívaný prostredníctvom inferencie bez dodatočného tréningu na analyzovaných dátach. Pre spracovanie rozsiahlejších časových okien bol využitý kontextový rozsah 65 536 tokenov, ktorý umožnil analyzovať väčšie bloky udalostí naraz.

2.2.1 Dátové súbory a ich predspracovanie

Na overenie navrhnutého prístupu bol použitý dataset The Case of the Stolen Szechuan Sauce (SSS) [22], ktorý predstavuje realistický scenár kompromitácie prostredia operačného systému Windows. Dataset obsahuje obrazy diskov dvoch zariadení:

- Domain Controller,
- Victim Desktop.

Výhodou tohto datasetu je dostupnosť manuálne označených udalostí súvisiacich s bezpečnostným incidentom, ktoré umožňujú objektívne vyhodnotenie úspešnosti navrhnutého riešenia.

Prvým krokom spracovania bolo vytvorenie jednotnej chronologickej časovej osi pomocou nástroja **Plaso** [23]. Tento framework umožňuje extrahovať digitálne stopy z rôznych zdrojov operačného systému Windows, napríklad zo systémových registrov, súborového systému, systémových udalostí a ďalších forenzných artefaktov.

Výstup nástroja Plaso obsahuje veľké množstvo atribútov, z ktorých viaceré predstavujú technické metadáta alebo redundantné informácie, ktoré nemajú priamy význam pre interpretáciu bezpečnostných udalostí. Z tohto dôvodu bola vykonaná selekcia atribútov s cieľom vytvoriť kompaktnú reprezentáciu vhodnú pre spracovanie jazykovým modelom.

Počas predspracovania boli vykonané nasledovné kroky:

- vytvorenie jednotného časového atribútu **datetime** zlúčením pôvodných hodnôt dátumu a času,
- odstránenie nerelevantných metadát,
- výber významných atribútov reprezentujúcich charakter udalosti (sourcetype, type, MACB, user, host, desc),
- skrátenie textového atribútu **desc** na definovanú dĺžku s cieľom optimalizovať využitie kontextového okna modelu,
- rozdelenie časovej osi na menšie bloky udalostí, ktoré je možné spracovať v rámci jedného vstupu pre jazykový model.

Výsledná reprezentácia pozostávala zo štruktúrovaného textového vstupu obsahujúceho chronologicky zoradené udalosti systému Windows.

2.2.2 Analýza pomocou veľkého jazykového modelu

Navrhnutý prístup využíva veľký jazykový model ako analytický nástroj na interpretáciu digitálnych stôp. Každé časové okno je transformované do textovej podoby a spolu s kontextovými informáciami tvorí vstupný prompt pre model.

Súčasťou vstupu môže byť okrem samotných digitálnych stôp aj dodatočný kontext opisujúci analyzované prostredie alebo organizáciu, čím sa modelu poskytujú informácie potrebné na správnu interpretáciu udalostí. Výstupom modelu je štruktúrovaná množina identifikovaných bezpečnostných udalostí vrátane ich popisu a ďalších charakteristík.

Keďže výstupy veľkých jazykových modelov nemusia mať vždy konzistentnú štruktúru, po ukončení inferencie nasledovala fáza validácie a normalizácie výsledkov. Tá zahŕňala kontrolu formátu výstupu, odstránenie neplatných alebo duplicitných záznamov

a zjednotenie názvov jednotlivých kategórií. Cieľom bolo zabezpečiť konzistentný a jednoznačný výstup vhodný na ďalšiu analýzu.

2.2.3 Vyhodnotenie úspešnosti

Úspešnosť navrhnutého prístupu bola hodnotená porovnaním identifikovaných udalostí s manuálne označenými bezpečnostnými incidentmi nachádzajúcimi sa v dátovej sade.

Na hodnotenie boli použité štandardné metriky binárnej klasifikácie:

- **Precision**, ktorá vyjadruje podiel správne identifikovaných bezpečnostných udalostí zo všetkých udalostí označených modelom ako podozrivé.
- **Recall**, ktorá udáva podiel správne identifikovaných bezpečnostných udalostí zo všetkých skutočných bezpečnostných udalostí.
- **F1-score**, predstavujúce harmonický priemer metrík Precision a Recall, ktoré poskytuje vyvážené hodnotenie úspešnosti modelu pri výrazne nevyvážených dátach.

Pri vyhodnocovaní boli samostatne analyzované časové úseky obsahujúce bezpečnostný incident a časové úseky reprezentujúce výlučne bežnú prevádzku systému. Takýto prístup umožnil nezávisle hodnotiť schopnosť modelu identifikovať bezpečnostné udalosti a zároveň sledovať počet falošne pozitívnych detekcií počas normálnej prevádzky.

2.2.4 Interpretovateľnosť výsledkov

Jednou z hlavných výhod navrhnutého prístupu je vysoká interpretovateľnosť výsledkov. Veľký jazykový model neposkytuje iba označenie podozrivých udalostí, ale zároveň vytvára ich textový opis a vysvetlenie, ktoré môže digitálnemu forenznému analytikovi pomôcť pri ďalšom vyšetrovaní.

Výsledkom analýzy preto nie je iba zoznam identifikovaných digitálnych stôp, ale aj ich kontextová interpretácia, ktorá umožňuje rýchlejšie pochopenie priebehu bezpečnostného incidentu a podporuje rozhodovací proces analytika.

Na rozdiel od klasických klasifikačných modelov, kde výsledkom býva iba pravdepodobnostné skóre alebo trieda, LLM poskytuje prirodzený jazykový opis analyzovaného správania. Tento opis obsahuje kontext udalosti, potenciálnu súvislosť s bezpečnostným incidentom a zdôvodnenie výberu konkrétnych digitálnych stôp. Takáto forma výstupu znižuje náročnosť manuálneho preskúmania rozsiahlych časových osí a umožňuje analytikovi rýchlejšie overiť relevantné udalosti.

Záver

Cieľom tohto výstupu bolo preskúmať možnosti využitia moderných metód strojového učenia a umelej inteligencie pri identifikácii digitálnych stôp v prostredí digitálnej forenzej analýzy. Hlavným zámerom nebolo vytvorenie univerzálneho automatizovaného systému, ktorý by nahradil prácu analytika, ale overenie, do akej miery môžu vybrané metódy podporiť proces analýzy rozsiahlych a komplexných dátových množín, v ktorých je manuálne vyhľadávanie relevantných udalostí časovo náročné a prakticky obmedzené.

V rámci riešenia boli skúmané dva odlišné prístupy – hierarchické zhľukovanie a analýza pomocou veľkých jazykových modelov. Tieto metódy predstavujú rozdielne pohľady na identifikáciu digitálnych stôp. Zatiaľ čo metódy zhľukovania pracujú najmä s podobnosťou medzi objektmi a umožňujú odhaliť prirodzene sa vyskytujúce štruktúry v dátach, veľké jazykové modely využívajú schopnosť interpretovať komplexné textové reprezentácie a poskytovať kontextové vysvetlenie identifikovaných udalostí.

Hierarchické zhľukovanie sa ukázalo ako vhodný nástroj na analýzu rozsiahlych dátových množín, v ktorých nie je možné vopred definovať počet alebo charakter jednotlivých skupín. V prostredí digitálnej forenzej analýzy je táto vlastnosť významná najmä pri práci s neznámymi aktivitami, kde analytik často nemá k dispozícii dostatočné informácie o analyzovanom prostredí alebo očakávaných vzoroch správania. Vytvorené zhľuky umožňujú identifikovať podobné skupiny digitálnych stôp, odhaliť opakujúce sa vzory a podporiť následnú analýzu súvisiacich udalostí.

Význam hierarchického zhľukovania spočíva najmä v schopnosti redukovať komplexnosť analyzovaných dát. Namiesto manuálnej kontroly jednotlivých záznamov môže analytik pracovať s menším počtom reprezentatívnych skupín, ktoré charakterizujú podobné správanie. Takýto prístup môže byť využiteľný napríklad pri analýze veľkého množstva príkazov vykonaných útočníkmi, identifikácii podobných spôsobov kompromitácie alebo odhaľovaní opakujúcich sa aktivít v rôznych incidentoch.

Druhým skúmaným prístupom bolo využitie veľkých jazykových modelov na interpretáciu digitálnych stôp. Experimenty ukázali, že moderné predtrénované jazykové modely dokážu pracovať s komplexnými foreznými časovými osami a identifikovať udalosti, ktoré môžu súvisieť s bezpečnostným incidentom. Významnou vlastnosťou tohto prístupu je schopnosť kombinovať jednotlivé udalosti do širšieho kontextu, čo predstavuje rozdiel oproti klasickým algoritmom strojového učenia, ktoré často pracujú iba s numerickými reprezentáciami alebo explicitne definovanými príznakmi.

Pri návrhu riešenia založeného na LLM bolo zároveň ukázané, že úspešná aplikácia veľkých jazykových modelov v oblasti digitálnej forenzej analýzy nezávisí iba od samotného modelu, ale významnú úlohu zohráva aj príprava dát, vhodná reprezentácia digitálnych stôp a kontrola výsledných výstupov. Spracovanie pôvodných forezných artefaktov do jednotnej

chronologickej časovej osi, výber relevantných atribútov a rozdelenie dát do spracovateľných blokov umožnili efektívnejšie využitie schopností jazykového modelu.

Výsledky zároveň poukázali na význam interpretovateľnosti pri aplikácii metód umelej inteligencie v digitálnej forenznej analýze. Automatizované metódy môžu výrazne urýchliť proces identifikácie relevantných udalostí, avšak ich výstupy musia zostať zrozumiteľné a overiteľné odborníkom. Z tohto dôvodu predstavuje schopnosť vysvetliť dôvody rozhodnutia modelu jeden z kľúčových faktorov praktického využitia týchto technológií. V prípade zhukovacích metód ide najmä o analýzu vlastností jednotlivých skupín a charakteristických atribútov, zatiaľ čo pri veľkých jazykových modeloch ide o vytvorenie kontextového vysvetlenia identifikovaných udalostí.

Experimentálne overenie na troch nezávislých datasetoch ukázalo, že navrhnutá metodika dokáže identifikovať skupiny digitálnych stôp súvisiace s bezpečnostným incidentom, odlíšiť ich od bežného správania systému a poskytnúť analytikovi interpretovateľné informácie o charaktere identifikovaných aktivít. Výsledky naznačujú, že hierarchické zhukovanie časových okien predstavuje vhodný nástroj na podporu prvej orientácie v rozsiahlych forenzných časových osiach.

Navrhnuté prístupy ukázali, že metódy strojového učenia a veľké jazykové modely môžu predstavovať významnú podporu pre digitálnu foreznú analýzu, najmä pri práci s rozsiahlymi dátovými množinami. Ich hlavný prínos spočíva v schopnosti znižovať množstvo manuálnej práce, zvýrazňovať relevantné digitálne stopy a poskytovať analytikovi informácie potrebné na rýchlejšie pochopenie vyšetrovaného incidentu.

Zároveň však tieto metódy nemožno chápať ako úplnú náhradu odborného rozhodovania. Výsledky automatizovanej analýzy musia byť vždy interpretované v širšom kontexte vyšetrovania a overené pomocou dostupných forenzných dôkazov. Budúcnosť využitia umelej inteligencie v digitálnej forenznej analýze preto nespočíva iba v zvyšovaní presnosti modelov, ale aj v tvorbe transparentných, vysvetliteľných a prakticky využiteľných nástrojov, ktoré budú efektívne spolupracovať s odborným analytikom.

Bibliografia

- [1] C. Hargreaves and J. Patterson, "An automated timeline reconstruction approach for digital forensic investigations," *Digital Investigation*, vol. 9, pp. S69–S79, 2012.
- [2] L. F. daCruzNassif and E. R. Hruschka, "Document clustering for forensic analysis: An approach for improving computer inspection," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 46–54, 2012.
- [3] X. Du and M. Scanlon, "Methodology for the automated metadata-based classification of incriminating digital forensic artefacts," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, pp. 1–8, 2019.
- [4] M. AbuJarour, R. B. Tareaf, and V. Hafner, "Using machine learning to prioritize forensic artifacts in email evidence," in *2025 International Conference on Information Networking (ICOIN)*, pp. 201–206, 2025.
- [5] H. Studiawan, C. Payne, and F. Sohel, "Graph clustering and anomaly detection of access control log for forensic purposes," *Digital Investigation*, vol. 21, pp. 76–87, 2017.
- [6] E. Marková, P. Sokol, and K. Kováčová, "Detection of relevant digital evidence in the forensic timelines," in *2022 14th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1–7, 2022.
- [7] H. Studiawan and F. Sohel, "Anomaly detection in a forensic timeline with deep autoencoders," *Journal of Information Security and Applications*, vol. 63, p. 103002, 2021.
- [8] F. Breiteringer, H. Studiawan, and C. Hargreaves, "Sok: Timeline based event reconstruction for digital forensics: Terminology, methodology, and current challenges," *Forensic Science International: Digital Investigation*, vol. 53, p. 301932, 2025.
- [9] R. Adila, H. Studiawan, and F. Breiteringer, "Forensic event reconstruction with process mining," *IEEE Access*, 2026.
- [10] L. Schmidt and S. Schinzel, "Towards automated and robust forensic event reconstruction," in *2025 IEEE 24th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pp. 200–210, 2025.
- [11] P. Sokol, Ľ. Antoni, O. Krídlo, E. Marková, K. Kováčová, and S. Krajčí, "Formal concept analysis approach to understand digital evidence relationships," *International Journal of Approximate Reasoning*, vol. 159, p. 108940, 2023.
- [12] H. Studiawan, F. Breiteringer, and M. Scanlon, "Towards a standardized methodology and dataset for evaluating llm-based digital forensic timeline analysis," *Forensic Science International: Digital Investigation*, vol. 54, p. 301982, 2025.

- [13] Y. Wei, K.-P. Chow, and S.-M. Yiu, "Insider threat prediction based on unsupervised anomaly detection scheme for proactive forensic investigation," *Forensic Science International: Digital Investigation*, vol. 38, p. 301126, 2021.
- [14] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint arXiv:1109.2378*, 2011.
- [15] M. R. Ackermann, J. Blömer, D. Kuntze, and C. Sohler, "Analysis of agglomerative clustering," *Algorithmica*, vol. 69, no. 1, pp. 184–215, 2014.
- [16] Magnet CTF 2019 Windows Desktop, 2019, [online 2026-05-12]. Dostupné na: <https://digitalcorpora.s3.amazonaws.com/corpora/scenarios/magnet/2019%20CTF%20-%20Windows-Desktop.zip>.
- [17] Magnet CTF 2022 Windows, 2022, [online 2026-05-12]. Dostupné na: <https://digitalcorpora.s3.amazonaws.com/corpora/scenarios/magnet/2022%20CTF%20-%20Windows.zip>.
- [18] Data LeakageCase, [online 2026-05-12]. Dostupné na: https://cfreds-archive.nist.gov/data_leakage_case/data-leakage-case.html.
- [19] DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*, 2019 [online 2025-09-25]. Dostupné: 10.48550/arXiv.1810.04805.arXiv:1810.04805 [cs]
- [20] QWEN TEAM. Qwen3.6-35B-A3B. 2026. [online 2026-05-12]. Dostupné: <https://qwen.ai/blog?id=qwen3.6-35b-a3b>.
- [21] GOOGLE DEEPMIND. Gemma 4. 2026. [online 2026-05-12]. Dostupné: <https://deepmind.google/models/gemma/gemma-4/>.
- [22] Case 001 – the stolen szechuan sauce, 2020, [online 2026-05-12]. Dostupné na: <https://dfirmadness.com/the-stolen-szechuan-sauce/>.
- [23] Plaso Team. (n.d.). *Plaso documentation*. [online 2026-05-12]. Dostupné: <https://plaso.readthedocs.io/en/latest/>.