

Co je předmětem zkoumání digitální forenzní analýzy?

What is the subject of digital forensic analysis?

Ing. Marián Svetlík, Sr.

Anotace:

Pro správné pochopení toho, čím se DFA zabývá, jaké metody a postupy používá a jak využívat její výstupy je klíčové správně definovat předmět DFA. Na základě poznání výjimečné vlastnosti digitální stopy, kdy její originál a kopie jsou identické, vzájemně zaměnitelné a tím také mající stejnou vypovídající a důkazní hodnotu, dospějeme k závěru, že digitální stopa je informace. Informace, jako nehmotná entita, nezávislá na její materiální interpretaci, tedy nezávislá na nosiči.

Annotation:

For a good understanding of what DFA is doing, what methods and procedures it uses, and how to use its reinforcements, it is crucial to define the subject of DFA correctly. Recognizing the exceptional nature of a digital trace, where its original and copies are identical, interchangeable and thus also of the same predictive and probative value, we conclude that the digital track is information. Information as an immaterial entity, independent of its material interpretation, is independent of the medium.

Klíčová slova:

digitální forenzní analýza, digitální stopa, digitální informace

Keywords:

digital forensics analysis, digital trace, digital information

ÚVOD

Jednou jsem už definici Digitální Forenzní Analýzy v našem časopisu uvedl. Pro jistotu ji tady ještě jednou zopakuji:

Digitální forenzní analýza (dále také pro zjednodušení DFA) je exaktní forenzní věda, která zkoumá procesy a zákonitosti vzniku, existence a zániku digitální stopy a interpretuje tyto poznatky na objasnění dějů a procesů s tím souvisejících pro účely potřeb orgánů státu (pro jejich rozhodovací činnosti) a potřeby dalších subjektů pro účely jejich právních úkonů.

Tedy je to pravé místo, abychom si tuto definici podrobněji rozebrali. Je to nutné, protože vlastní zkušenost mi říká, že intuitivně tu definici lze pochopit, ale ten hlavní háček je právě v té „intuitivnosti“. Každý si totiž pod tím představuje trochu něco jiného. Proč? Protože každý chápe jednotlivé pojmy použité v definici mírně odlišně a pak ve výsledku vznikají neshody. A opět musím zdůraznit, že dopady „neshod“ v oblasti tvorby a hodnocení důkazů v (zejména) trestním řízení mohou mít fatální důsledky.

Tím klíčovým pojmem v uvedené definici je samotný předmět zkoumání, kterým je „digitální stopa“. Vysvětlení, proč to je klíčovým pojmem, je jednoduché - **digitální stopa je informace.**

V počátcích svoji profesionální praxe (a ještě i dlouho poté) jsem stál před problémem, co to vlastně ta digitální forenzní analýza je. Čím se zabývá, co je jejím předmětem. Pokoušel jsem se tyto otázky zodpovědět pomocí známých teorií klasické definice¹⁾. Diskutovali jsme to s kolegy a paradoxně vznikalo víc otázek než odpovědí. Podvědomě jsme cítili, že někde není něco v pořádku v tom, jak chápeme problém digitálních stop. Pořád nás to, pod vlivem našich tehdejších zkušeností a vědomostí, táhlo k vysvětlení, že digitální

stopa je materiální, latentní mikrostopa (tedy že to jsou ty změny magnetického pole na magnetickém záznamovém médiu nebo drobné mikroskopické změny v odrazové vrstvě na CD nosiči a podobně). Podle klasické kriminalistické teorie stop se ani jiné řešení nenabízelo.

V předchozím článku²⁾, v části o opakovatelnosti zkoumání digitálních stop jsem ale psal, „že správně provedená kopie digitální stopy je identická se svým originálem a je s ním plně zaměnitelná. Nezávisí dokonce (ve většině případů) ani na nosiči, na kterém je digitální stopa zaznamenána, a tím vlastně (opět ve většině případů) také na okolním prostředí, které nemá na stopu žádný zásadní vliv“. Znamená to, že nezajišťujeme materiální stopu, tedy změny magnetického pole nebo dírky v děrné pásce. Předmětem zajištění digitálních stop je informace, která je digitálně kódována.

Ano, je pravdou, že pro zajištění digitálních stop musíme často (ale ne vždy) zajistit jejich nosiče. Důležité však je, že tyto nosiče (pevné disky, paměti RAM výpočetní techniky, síťová odposlechová zařízení a další) zpravidla nejsou předmětem následného zkoumání. Tím jsou informace v digitální podobě, které jsou na takových nosičích uloženy. Samotné prvky výpočetní techniky, jako materiální stopy, mohou pouze dokreslovat a doplňovat informace, které jsou výsledkem zkoumání digitálních informací. Nechci tím říct, že hardware počítačů (tedy prvky výpočetní techniky, jako materiální stopy) nemůže být předmětem zkoumání, ale to již není zkoumání DFA, ale spíše zkoumání v jiném oboru (např. elektronika?).

Abychom ale správně pochopili co to vlastně digitální stopa je, musím v dalším trochu podrobněji objasnit některé související pojmy.

¹⁾ Kriminalistické stopy jsou stopy paměťové a materiální

²⁾ Digital Forensic Review, 2/2018, str. 7

INFORMACE

Abychom pochopili širší souvislosti, musíme si nejdříve vyjasnit pojem „informace“. Důvod je jednoduchý, protože primárním zájmem DFA je analýza (digitálních) informací. Co to tedy informace je?

Zkusíme jeden příklad. Abychom se mohli rozhodovat správně, potřebujeme k tomu odpovídající objem relevantních a dle možností pravdivých nebo alespoň hodnověrných informací. Jako jednoduchý příklad, který dokresluje popisovanou problematiku, lze uvést, že pro rozhodnutí o tom zda byl spáchán nějaký trestný čin, musíme mít relevantní informace (jinak i odpovědi na základní kriminalistické otázky) o tom kdo, co, kdy, kde, jak a proč.

Informace je tedy pojem, který charakterizuje snížení neurčitosti určitého stavu³⁾.

Na začátku vyšetřování nemáme dostatek informací o domnělém trestním činu. Tak jak postupně získáváme relevantní informace, neurčitost našeho vyšetřování se snižuje. A na faktu nemění nic ani skutečnost, jestli průběh vyšetřování zapisujeme na papír do vyšetřovacího spisu nebo si jej necháváme pouze v hlavě.

Informace je nehmotná, nicméně pro předávání informace od zdroje k cíli (např. od podezřelého k vyšetřovateli nebo od vyšetřovatele k jeho nadřízenému) **je nutný nějaký materiální nosič**. Např. při výslechu to je zvukové vlnění, které je způsobeno řečí podezřelého, nebo papír s textem v případě získání listinných důkazů.

Kódování informace je dalším důležitým pojmem, který vyžaduje vysvětlení. Aby totiž bylo možné předávat (nebo jakkoliv jinak zpracovávat) informaci, výše uvedená materiální prezentace (interpretace) informace (její záznam na materiálním nosiči) není postačující. Mezi zdrojem a cílem informace musí existovat domluva na tom, v jaké „řeči“ si danou informaci předají. Jinak řečeno jak a za použití jakých prostředků budou předávanou informaci kódovat.

Tzv. ústní předávání informace předpokládá nejenom společný jazyk, kterému oba rozumí, ale také schopnost předávajícího mluvit (tedy že není němý), schopnost příjemce slyšet (tedy že není hluchý) a nakonec existence prostředí, které je schopno přenášet mluvenou řeč. A mohli bychom jmenovat i další podmínky pro přenos informací, jako např. to, že předávající bude schopen v daném jazyce informaci formulovat a že příjemce bude schopen (např. inteligenčně) předávanou informaci pochopit.

Jiným příkladem kódování informací je psaný text. I v tomto případě se lidé museli nejdříve domluvit na použitém jazyce, ale kromě toho také na použité abecedě (znakové sadě), systému skládání písmen do slov a vět. Tak vlastně vzniklo diskrétní kódování. To vše se zdá být samozřejmé, automatické a nikdo se nad tím vlastně ani nezamýšlí. Máme to zažité už stovky

a tisíce let. Tím spíše, že výše uvedené způsoby předávání informací (neboli způsoby komunikace) byly lidstvem vymyšleny tak, aby byly **přímo vnímatelné lidskými smysly**.

Lze tedy říct, že **předávaná informace musí být pro příjemce zachytitelná (zpracovatelná), relevantní a srozumitelná⁴⁾**. Darmo budeme poslouchat pomocí sofistikovaných zařízení kosmické signály a snažit se něco zjistit o mimozemských civilizacích, když nebudeme schopni jimi posílané informace odlišit od kosmického šumu a když, tak nebudeme schopni porozumět jejich způsobu kódování informace. Stejně tak zbytečně budeme koukat na perforaci děrné pásky (pamětníci počátků výpočetní techniky si pamatují), když nevíme, podle jakých pravidel jsou dírký v pásce tvořeny a co která perforace představuje.



(zdroj: wikipedia - https://cs.wikipedia.org/wiki/Děrná_páska)

Informace také musí být relevantní, protože jejím základním účelem je „snížení neurčitosti určitého stavu“. Jestliže je naším cílem vyšetřit nějakou trestnou činnost nebo nějaký bezpečnostní incident v informačním systému, potom zřejmě poznatek, že „počet slonů v Africe neustále klesá“ pro stav našeho vyšetřování nebude informací (za předpokladu, že nevyšetřujeme právě pytláctví na slonech), protože s vyšetřováním to nemá souvislost a neurčitost vyšetřování to nijak nesníží (tedy k vyšetřování to nepřispěje).

Mimo výše uvedeného lze také dovodit další podstatná poznání, konkrétně že:

- **text** na papíru, stejně jako **díry** v děrné pásce, **magnetický záznam** na pevném disku, **zvukové vlnění** nebo **elektrický proud** v síťovém propojovacím kabelu (a další) jsou pouze různé **materiální záznamy (materiální interpretace) nehmotné informace**. Dokonce tyto velice různorodé materiální záznamy mohou obsahovat stejnou informaci, jinak řečeno ta stejná informace může být zaznamenána různým způsobem na různých materiálních nosičích. Různé materiální nosiče informace nemusí hrát žádnou významnou roli při posuzování a hodnocení informace, která je na nich zaznamenána (nebo jejich prostřednictvím přenášena)

- **papír, děrná páska, pevný disk v počítači nebo propojovací kabel (a mnoho dalších) jsou pouze (potenciálními) nosiči informací**. Typ nosiče sám o sobě dokonce ani nic nemusí vypovídat o způsobu záznamu a způsobu kódování informace.

Například na papíře může být zaznamenána jak informace v diskrétním kódování (text), tak analo-

³⁾<https://cs.wikipedia.org/wiki/Informace>

⁴⁾Dalo by se mluvit také o dalších attributech, jako je např. pravdivost, včasnost a další podle konkrétní situace.

gová informace (obrázek), tak i informace digitální⁵⁾ (resp. přesněji binární - tedy zápis nul a jedniček):

Kakosty (Geranium) je neuvěřitelně široká skupina vytrvalých rostlin z čeledi Geraniaceae, která je rozšířena asi 820 druhy v československé přírodě se jich objevuje okolo 20. Uyskylují se hlavně v mírném pásu a to ve všech stadiích, jisté druhy rostou i v horách tropických oblastí. Tak široká rozšíření mají díky různým ekologickým nárokům, teploplné druhy vyžadují propustné půdy a slunná stanoviště, stínolínové například hluší a těžší půdy které jsou dostatečně vlhké. Popis: Kakosty jsou většinou byliny a jiné řídce keře vysoké 10 až 70 cm, velmi často porostlé různými druhy chlupů. Listy vyrůstají nejčastěji vstřícně, mají palisty, bývají dlanitě dělené až dlanitě laločnaté a jednotlivě lichozeprné. Většina jsou s výjimkou několika druhů s jednoduchými listy, které mají 3 až 5 pářímí rostou v listové růžici. Většinou druhy listy na podzim opadají, některé, např. Kakost Endressův (Geranium endressii) jsou stálezelené i přes zimu období.

Květy se stopkami jsou pravidelné, pětičetné, oboupohlavné a obvykle vyrůstají v termínálních nebo úžlabních dvoukvětých vidlanecch, jiné řídce jednotlivé. Kalíšni i korunní lístky jsou volné, korunní jsou obvykle zúžené do netřikrát a nížeji barou nadrou, červenou, červenofialovou, fialovou, hnědofialovou, růžovou nebo i bílou.

V květu je deset tyčinek vyrůstajících ve dvou kruzích a suchrní, pětízpuzdří semeník obsahující v každém pouzdru po dvou vajíčkách, z nichž jedno zakrnl a pouze z druhého se po opylení vyvine semeno. Semeník vytvořený z pěti plodolistů nee zahrnutou čímkou rozštěpenou do pěti blízken. V báze tyčinek je pět nedočkých žlázek. Dozrálý zobaný plod, merikarp, se rozpoztí v pět pouzder obsahující po jednom semeni. Pouzdra se od trvalého středního sloupku oddělují a jsou drzena jen obloukovitě napruženými osami, ty se ve vhodný čas uvolní a vystřelí semena do vzdálenosti až 10 m.

Význam: Mnohé druhy jsou již po staletí využívány pro své léčivé květy, dekorativní listy a pokrovný účinek jako okrasné rostliny, např. kakost krvavá, kakost luční, kakost lesní a byly z nich vylečteny mnohé kultivary odolnější se habitem než barvou květů či listů. Populární se stávají i druhy stínolínové, těch se využívá jako půdorysných pro vytvoření uzavřeného porostu pod korunami stromů, např. kakost oddenkatý či kakost Endressův. Pro suchá vyschlá místa, skalky či suché zídky je vhodný kakost krvavý. Původní druhy kakostí se možná mohou stát i velmi cennými vzhledově, ale bohužel mnohé druhy kakostí se možná jsou i druhy považované za neppřijemné, z důvodů je to hlavně kakost sardutá, nepřijemně páchnoucí stínolinná rostlina. V české přírodě vyrůstají i druhy

(Informace v diskrétním kódování - text psaný na papíře - zdroj autor s použitím dat z wikipedie)



(Analogová informace - fotografie rostliny, která je popsána v předchozím textu - zdroj autor s použitím dat z wikipedie)

[illegible]

(Binární informace - část výše uvedeného textu v binární podobě (nuly a jedničky) s uvedením textové interpretace binárního kódu - zdroj autor s použitím dat z wikipedia)

Podobně i na magnetické pásce může být zaznamenána informace analogová (typicky kazeta s hudbou, známá z kazetových magnetofonů) nebo informace digitální (typicky zálohovací kazeta pro pořízení digitálních záloh⁶⁾). Jak je vidno z výše uvedeného příkladu, může dokonce existovat záznam té samé informace na tom samém typu nosiče informací jednou jako analogový záznam a podruhé jako záznam digitální.

Uvedené příklady jenom dokreslují to, že informace a jejich materiální interpretace nemají ve většině případů žádnou přímou souvislost.

DIGITÁLNÍ INFORMACE

Z výše uvedeného lze dovodit definici digitální informace: **Digitální informace je záznam informace v digitální podobě na materiálním nosiči, který je schopen takový druh záznamu nést nebo přenášet.**

Tato definice je intuitivně asi pochopitelná. Asi se ale shodneme na tom, že čím je definice stručnější, tím větší důraz je kladen na správné pochopení významu každého jejího slova. Proto teď jenom zopakují, že digitální informace je pořád informace ve své obecné podstatě, tedy primárně nehmotná. A aby byla informace podle definice, musí být pro příjemce **zachytitelná, srozumitelná a relevantní**. Proto potřebuje nějaký materiální nosič a vhodné kódování.

Pro přiblížení výše uvedeného použijí jeden velice jednoduchý příklad, ve kterém vyjdou z obecně akceptovatelných předpokladů, že ve výpočetní technice se operuje s jedničkami a nulami (tedy s „1“ a „0“) a že vzhledem k velkým objemům dat můžeme v každém počítadle bez problémů najít i mnohonásobný výskyt následující posloupnosti:

0010011100010000

Co to je? Má to vůbec nějaký vztah k digitální informaci? Uvedená posloupnost obsahuje nuly a jedničky, tedy splňuje nutnou podmínku existence vztahu k výpočetní technice (digitálnímu světu), ale není to podmínka postačující. Předpokládejme tedy, že se jedná o digitální záznam.

Je taková posloupnost pro příjemce (běžného, typického uživatele) zachytitelná (zjistitelná)? Tady je nutné připustit, že pro běžného uživatele s velkou pravděpodobností asi ne, běžný uživatel se s takovouto posloupností nul a jedniček nejspíše nikdy nesetká, pro něj je tato sekvence vlastně neviditelnou. Koncový uživatel musí využít speciálních prostředků nebo zprostředkovatelů, aby se mu tato sekvence zpřístupnila/zviditelnila.

Je taková sekvence srozumitelná? Opět musíme připustit, že pro běžného uživatele těch několik jedniček a nul určitě nebude nic znamenat. Uživatel opět potřebuje něco (nějaká pravidla, klíč ke kódu, automatický převodník), podle kterého mu systém danou

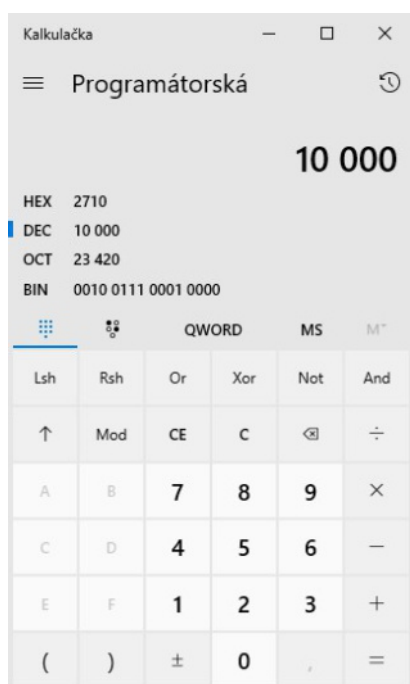
⁵⁾V současnosti pravděpodobně nej názornějším příkladem záznamu digitální informace na papíře je QR kód (https://cs.wikipedia.org/wiki/QR_kód)

⁶⁾Pamětníci si určitě vybaví, že v dávných počítačových dobách se pro uchování digitálních dat používaly obyčejné kazety, jak je známe z kasetáků

posloupnost „přeloží“. Takže přímo je taková posloupnost nesrozumitelná.

Připusťme, že existuje takový převodník, který danou posloupnost převede na základě nějakého předem smlouveného pravidla do srozumitelné podoby a uživateli zobrazí např. číslo 10 000 (přičemž skutečně platí, že binárně 0010011100010000 je v kódování „signed integer“ v desítkové početní soustavě⁷⁾ celé číslo s hodnotou 10 000). To už je pro uživatele srozumitelné, ale je to pro něj také relevantní? Jestliže uživatel (koncový příjemce informace) vyšetřuje např. dětskou pornografii, tak nejspíše ne, tedy posloupnost 0010011100010000 pro něj nebude informací ani po dekódování na číslo 10 000, protože mu to pravděpodobně nepomůže snížit stav nejistoty jeho vyšetřování dětské pornografie. Pro něj to pravděpodobně není relevantní informace, ale jenom nějaký digitální záznam. Jestliže ale vyšetřuje krádež 10 000 dolarů, tak ta posloupnost může být pro něj s velkou pravděpodobností důležitou (digitální) informací.

Tento příklad jsem sice ohodnotil jako „velice jednoduchý“, ale pro mnohé to asi jednoduché není. Nicméně i jen z pouhého přečtení je zřejmé, že bylo použito mnoho zjednodušení. Jako příklad uvedu „převod nul a jedniček do srozumitelné podoby podle nějakého pravidla“, jak je uvedeno v předchozím odstavci. Pro ilustraci použiji kopii okna standardní kalkulačky MS Windows 10 v „programátorském módu“:



To pravidlo, které bylo v našem příkladě použito, byl převod binární posloupnosti 0010011100010000 (BIN) do čísla v desítkové soustavě (DEC) 10 000. Jak je vidět z obrázku, v osmičkové soustavě (OCT) bychom dostali číslo 23 420 a v šestnáctkové (HEX) by to bylo 2710. Všechna zde uvedená čísla přitom vyjadřují tu stejnou hodnotu, jenom v jiné početní soustavě.

⁷⁾Tedy v početní soustavě, která je nám nejpřirozenější

Tady ještě musím doplnit, že **ne každou posloupnost nul a jedniček lze dekódovat** a tím z ní získat nějakou informaci. V zásadě existují dvě situace, kdy toho nelze dosáhnout:

a) **Zajištěním se nepovede získat původní informaci v případě, kdy posloupnost byla získána v rozporu s pravidly, podle kterých byla původní informace uložena.**

Zjednodušené vysvětlení může být následující. Naše původní posloupnost (0010011100010000) byla „vyseknuta“ (tedy získána) z podstatně delší posloupnosti, např. takové, která mohla vypadat přibližně následovně:

10111001001110001000010110

Teoreticky (zdůrazňuji že jen čistě teoreticky) ale mohla být posloupnost nul a jedniček „vyseknuta“ i následovně (tedy o dvě pozice vlevo):

10111001001110001000010110

Tím získáme posloupnost ve tvaru pro laika podobném - 1100100111000100. Když na takto „posunutou“ posloupnost použijeme stejný dekódovací postup, dostaneme jiné číslo, v tomto případě v desítkové soustavě by to bylo číslo 51 652 (místo původního čísla 10 000).

Odborníci určitě mohou namítat, že tento příklad je hodně přitažen za vlasy a že taková situace nastat v podstatě nemůže. Ano, asi mají pravdu. Jenže je potřebné zdůraznit, že získání (zajištění) dat (sekvenci nul a jedniček) musí být přísně v souladu s pravidly pro práci s digitálními daty, nelze to dělat libovolně, protože další analýzy mohou dospět ke zcela zcestným (a tedy nepravdivým) závěrům. To platí jak na mikroúrovni nul a jedniček, tak na úrovni celých souborů a dalších sad dat, jak se v ICT prostředcích mohou nacházet.

Co je na tomto příkladě nesmírně důležité je skutečnost, že **nesprávným zajištěním a následným dekódováním můžeme získat informaci, která se může jevit jako správná, ale je nepravdivá** (místo čísla 10 000 jsme dostali číslo 51 652). Jenom tím, že jsme použili jiný způsob zajištění a dekódování nemáme často ani šanci zjistit, že to byl způsob nesprávný, protože výsledek je vnímatelný (číslo 51 652 je stejně hezké číslo jako číslo 10 000) a mohlo by se lehce zdát, že i pravdivý. Jenomže pravdivý není a tato situace může v konečném důsledku výrazně ovlivnit výsledek, v našem případě například výsledek soudního procesu.

b) **Dekódováním se nepovede získat informaci v případě, kdy původní informace byla kódována neznámým způsobem nebo způsobem, který je jiný**

než způsob její dekódování. V tomto případě se nabízí ještě jednodušší vysvětlení - původní informace byla do nul a jedniček sice kódována, ale pro její dekódování nemáme odpovídající algoritmus. Stručně řečeno, původní informace byla šifrována způsobem, který neumíme dešifrovat.

Z uvedeného příkladu se dá nicméně dovodit, že digitální informace je informace jako jakákoliv jiná, jenom není běžnými uživatelskými prostředky přímo zjištělná a také není přímo srozumitelná. Digitální informace je tedy **latentní a kódovaná**. Pro uvedení digitální informace do vnímatelného stavu (a tedy do stavu použitelného) je nutné digitální záznam zpřístupnit a dekódovat. Až poté, co je digitální záznam zpřístupněn a dekódován, lze přistoupit k posuzování jeho **relevantnosti**.

DIGITÁLNÍ STOPA

Výše jsem se na obecné úrovni snažil dospět k tomu, co to je digitální informace a že digitální informace je latentní a kódovaná. Až poté, co je získána a dekódována, je možné začít posuzovat její relevantnost, protože do té doby nevíme nic o jejím obsahu, tedy do té doby nevíme, jaká informace je vlastně pod těmi pověstnými nulami a jedničkami ukrytá.

Vraťme se ale k situaci, kdy se povedlo z poloupnosti nul a jedniček získat korektním způsobem smysluplnou informaci. Následuje proces posuzování její relevantnosti, neboli vhodnosti dané informace snižovat neurčitost stavu objektu, který zkoumáme. Když je takto získaná informace použitelná k danému účelu, nazýváme ji digitální stopu.

Digitální stopa je tedy informace, která má určitou vypovídající schopnost o zkoumaném objektu a tedy snižuje neurčitost našeho poznání o tomto objektu. Digitální stopa je tedy informace o tom, že v informačním systému proběhl nebo probíhá nějaký děj nebo proces.

Například když se v digitálních datech povede najít informaci (textový řetězec) typu:

UpdateURL=https://update.libreoffice.org/check.php

je to stopa o tom, že v informačním systému (počítači) byl nebo ještě stále je nainstalován kancelářský balík LibreOffice. Digitální stopou může být záznam v registru OS Windows⁸⁾, záznam v logu nějakého systému nebo skutečně cokoliv, co nese jakoukoliv relevantní informaci, včetně informace, která může být zaznamenána byť i jenom v jednom jediném bitu.

Důležitou vlastností digitálních stop je její vztah (tedy relevantnost) ke sledované skutečnosti.

Kriminalistická digitální stopa

Aniž bych chtěl bagatelizovat velice často použí-

vaný pojem „kriminalistická digitální stopa“, jeho vysvětlení je nabíledni. Kriminalistická digitální stopa je taková digitální stopa, relevantnost které se posuzuje ve vztahu k vyšetřování trestné činnosti.

Pro ještě lepší objasnění tohoto pojmu se pokusím ještě jasněji vysvětlit právě to co a jak se ve vztahu k digitální stopě posuzuje a kdo relevantnost digitálních stop (a vůči čemu) posuzuje.

Relevance digitálních stop

Na tomto místě se ještě musím podrobněji vrátit k pojmu, který jsem již několikrát použil a který asi všichni intuitivně chápeme - k pojmu relevantnosti. **Relevance znamená významnost či důležitost, obvykle ve vztahu k nějakému cíli nebo záměru⁹⁾**. Relevantnost ve vztahu k naší problematice můžeme chápat jako důležitou informaci, která nám pomůže dosáhnout náš cíl, tedy přispět k poznání dějů a procesů v informačním systému.

Důležité v tomto kontextu je to, kdo hodnotí relevantnost digitální informace. Jestliže neustále mluvím o digitální forenzní analýze, tak jejím cílem je „zkoumání procesů a zákonitostí vzniku, existence a zániku digitálních stop a interpretace těchto poznatků na objasňování dějů a procesů s tím souvisejících¹⁰⁾“ a logicky hodnotitelem relevantnosti digitálních informací je (vzhledem k cílům digitální forenzní analýzy) forenzní expert/soudní znalec. Jestliže je ale cílem vyšetřování bezpečnostního incidentu nebo trestného činu, tak hodnotitelem informací získaných z digitálních dat je vyšetřovatel¹¹⁾. Na první pohled se může zdát, že v tom není žádný rozdíl. Opak je ale pravdou, minimálně v našem právním prostředí. A není v tom žádný rozpor. **Naopak, rozdíl ve vymezení a pochopení rolí znalce a vyšetřovatele je nesmírně důležitý¹²⁾**. Pokusím se to zjednodušeně a ve stručnosti vysvětlit na příkladech:

K vyšetření trestné činnosti je nutné zkoumat počítač podezřelého s tím, že je potřebné zjistit, zda se v počítači nacházejí fotografie, které by prokazovaly určité skutečnosti, důležité pro vyšetřování. V případě nálezů takových fotografií je potřebné zjistit další doplňující informace o způsobu pořízení těchto fotografií.

Data z počítače byla zajištěna a byla dekódována. Znalec posuzoval relevantnost výsledných dat ve vztahu k požadavku identifikace formátů obrazových dat (digitálních fotografií). Byla nalezena souvislá série souborů fotografií s názvy IMG01001.JPG až IMG01020.JPG, celkem tedy 20 souborů. Z fotografií byly zjištěny EXIF informace, které potvrdzovaly, že daná série fotografií byla pořízena v souvislé časové řadě na jednom místě a jedním fotoaparátem. Tyto výsledky byly předány vyšetřovateli.

Vyšetřovatel následně posuzoval relevantnost fotografií k vyšetřované trestné činnosti a vybral tři fotografie, které se k dané věci vztahovaly.

Tady musím zdůraznit, že vyšetřovatel neposuzoval relevantnost digitálních stop ve formě digitální fotografie.

⁸⁾Viz např.: Ing. J. Hološka, "Důkazy o spuštění aplikace", Digital Forensic Review, 1/2018, str. 21-23, ISSN 2570-5059

⁹⁾<https://cs.wikipedia.org/wiki/Relevance>

¹⁰⁾ https://www.researchgate.net/publication/277871336_ZAZRAKY_FORENZNIHO_ZKOUMANI

¹¹⁾ V tomto kontextu je pojem "vyšetřovatel" v tom nejširším kontextu jako osoba, která provádí (a je zodpovědná za) vyšetřování

¹²⁾ Pro bližší vysvětlení situace v trestním řízení viz zákon 141/1961 Sb.

Vyšetřovatel posuzoval informace na poskytnutých obrázcích nezávisle na tom, zda to byla digitální fotografie nebo ne. Úplně správně by to mělo znít, že „vyšetřovatel posuzoval relevantnost informací, které byly zobrazeny na vybraných fotografiích (tedy bez vztahu k tomu, že fotografie byly digitální). Z druhé strany je nutné zdůraznit, že znalec nikdy nemůže posuzovat relevantnost digitálních stop k vyšetřované skutečnosti.

Vyšetřovatel se následně obrátil na znalce s dodatečným požadavkem podrobněji prozkoumat hodnověrnost veškerých informací, které se k vybraným fotografiím vztahovaly.

Znalec proto provedl analýzu obrazových dat pomocí speciálního SW, který vypočítává specifické charakteristiky snímání obrazového čipu pro skupinovou identifikaci fotografického přístroje, který digitální fotografie pořídil. Znalec následně zjišťoval relevantnost získaných dat ve vztahu k datům, která získal z EXIF informací všech snímků a dospěl k závěru, že vyšetřovatelem vybrané tři fotografie byly s velkou pravděpodobností padělané, protože vypočtené charakteristiky snímání obrazových čipů u daných třech fotografií se výrazně odlišovaly od ostatních fotografií. Bylo zjištěno, že předmětné tři fotografie byly pořízeny jiným fotoaparátem za jiných světelných podmínek a že EXIF informace těchto tří fotografií byly pozměněny tak, aby se zdálo, že všechny fotografie byly pořízeny jedním fotoaparátem v jedné souvislé řadě.

Je mi nad slunce jasné, že hned se k tomuto příkladu snese zásadní připomínka - jak je možné, že znalec hned neodzkoumal možnost padělání fotografií? Připomínka vskutku zásadní, ale její zodpovězení má velké množství proměnných a v této fázi výkladu i velké množství neznámých. Snad se k vysvětlení doberu v některé z dalších částech povídání o digitální forenzní analýze. Ve stručnosti se dá říct, že v prostředí současných digitálních technologií je ukryto takové nesmírné množství informací, že dohledání a analýza všech by nebyla prakticky v reálných podmínkách (kvalifikačních, vědomostních, časových a materiálních) možná. Proto je potřebné se soustředit jen na ty podstatné. Je pravdou, že možnost padělání fotografií je asi podstatnou informací, ale takových potenciálně podstatných informací je velké množství. Daly by se zkoumat dodatečné zásahy do obrazové informace, možnost skrývání jiných informací v obrázku pomocí steganografie a provádět mnoho dalších, vesměs náročných analýz. To ale primárně závisí od konkrétních potřeb, časových a finančních možností zadavatele, materiálního a přístrojového vybavení zpracovatele a také od efektivní spolupráce vyšetřovatele se znalcem. Ale to je na jiné povídání.

Z uvedeného příkladu je zcela zřejmá rozdílnost rolí znalce a vyšetřovatele. Znalec posuzoval výlučně odborné otázky, které obdržel od vyšetřovatele. V datech hledal taková data, která mají formát obrazových dat, protože vyšetřovatel potřeboval zjistit, zda jsou v datech obrázky, které by mohly mít vztah v vyšetřované skutečnosti. Znalcová kvalifikace neumožňuje dělat závěry o tom, zda informace na obrázku je nebo není relevantní pro vyšetřovatele. Znalec (konkrétně specialista na DFA) v zásadě nemůže posuzovat, co na daných fotografiích je, tedy nemůže posuzovat např.

informace ve vztahu k individuální nebo skupinové identifikaci subjektů vyobrazených na obrázku. To může být úkolem znalce v jiném oboru nebo závěr k tomu musí učinit sám vyšetřovatel.

Z druhé strany vyšetřovateli je v principu úplně jedno, v jaké formě mu znalec předal výsledky zkoumání (v tomto případě digitální obrázky), pro vyšetřovatele je podstatná informace, která na těch obrázcích je. Je jedno, zda mu byly obrázky poskytnuty jako soubory (např. v JPEG nebo PDF formátu) nebo jako wordový dokument s vloženými obrázky a troufám si tvrdit, že ve většině případů by bylo pro vyšetřovatele nejvhodnější ty obrázky vytisknout. Úlohou vyšetřovatele je totiž **posoudit relevantnost obsahu poskytnutých informací** (tedy toho, co je na obrázcích zobrazeno) k vyšetřované skutečnosti.

Jsem si jist, že vymezení rolí je z tohoto příkladu je zcela zřejmé. Znalec posuzuje relevantnost digitálních dat ve vztahu k položeným znaleckým otázkám, které by měly mít ze své podstaty pouze odborný charakter. Vyšetřovatel zase posuzuje relevantnost poskytnutých informací (výstupů znaleckého zkoumání) ve vztahu k vyšetřované události. Minimálně v našem právním prostředí jsou tyto dvě role striktně odděleny. Jestliže však se jedná o vyšetřování např. bezpečnostního incidentu, tedy o řízení mimo orgány státní moci, lze pravděpodobně připustit sloučení obou rolí. Vše potom závisí od smluvního nebo jiného ujednání zúčastněných subjektů.

TRANSFORMACE DIGITÁLNÍ STOPY

K určitým formám zpracování digitálních stop jsem se už vyjadřoval, když jsem zmiňoval požadavky na integritu digitálních dat¹³⁾. Tam jsem vysvětlil, co je originální digitální stopa a co je její identická kopie a jak se zajišťuje integrita kopie. A také to, že veškeré další činnosti, které se s digitální stopou provádějí, jsou v zásadě prováděny pouze na identické kopii (resp. kopiích) digitální stopy.

V průběhu práce dochází k mnoha transformacím digitální stopy. Některé transformace jsou nutné prakticky ve všech případech, jiné závisí od možností a způsobů získání/zajištění prvotních kopií digitálních stop, ale také od řešení konkrétního problému a od metody analýzy, která byla pro ten který případ zvolena.

Binární kopie

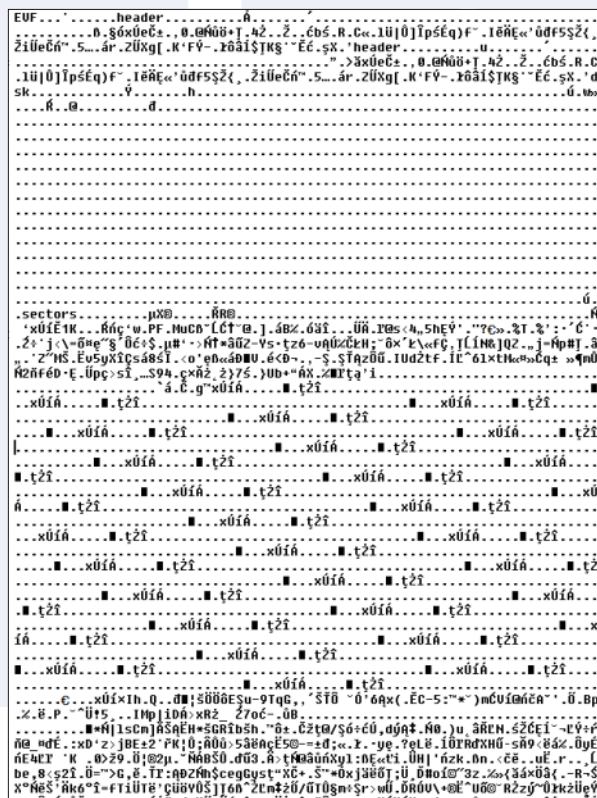
Binární kopie (někdy je položen důraz na forenzní účel a proto se nazývá „forenzní binární kopie“) je nejčastěji základním východiskovým zdrojem digitální stopy. Vzniká jako prvotní kopie digitálních dat v procesu jejich zajištění. Jinak se také často nazývá „obraz disku“, „identická kopie“, „forenzní obraz“, „otisk disku“ a podobně. Historicky se lze asi nejčastěji setkat s názvem „forenzní obraz disku“ (nebo jiného paměťového úložiště, např. forenzní obraz RAM

¹³⁾ Digital Forensic Review, 2/2018, str. 11-13

paměti počítače). Pochází to z anglického překladu „Forensic Disk Image“. Jak jsem se již zmiňoval v povídání o integritě dat, pro zajištění hodnověrnosti musí být pro binární kopii spočtena kontrolní suma, která je v případě využití digitálních stop jako důkazů její nedělitelnou součástí.

Co je významným charakteristickým rysem binární kopie? Binární kopie je jeden soubor¹⁴⁾ bez jakékoliv přímo identifikovatelné nebo viditelné struktury obsahu. Je to sice kopie všech dat z daného zdrojového paměťového úložiště, ale bez další speciální analýzy a interpretace obsahu je vidět pouze změt dat a nelze z takové kopie přímo vyčíst jakékoliv soubory, data, obrázky a pod. **Obsah binární kopie je tedy ještě pořád latentní a kódovaná digitální stopa.**

Obsah binární kopie si sice můžeme prohlédnout, ale obecně uvidíme pouze změt znaků (říká se tomu „rozsypaný čaj“), která se bude podobat tomu, co je vidět na následujícím obrázku:



(Náhled na obsah začátku forenzního obrazu disku)

Je zřejmé, že samotná binární kopie neumožňuje bez dalších nástrojů a analýz získat přístup k informacím, které jsou jejím obsahem.

Z technického pohledu je to zřejmé, protože binární kopie sama o sobě neumožňuje interpretovat přímo přístup k jednotlivým souborům. K tomu je nutné realizovat vstupní analýzu s cílem identifikovat, jak je disk rozdělen na oddíly a jaký souborový systém je

v jednotlivých oddílech použit. Zjednodušeně řečeno, je nutné dekódovat a analyzovat vnitřní strukturu obrazu disku. Bez takové analýzy je přístup k jednotlivým souborům prakticky nemožný.

Je to stejné, jako kdybychom se snažili podívat se na soubor MS Wordu, aniž bychom měli MS Word (nebo nějaký vhodný prohlížeč) k dispozici. Aniž bych chtěl příliš zaplňovat text nesmyslnými obrázky, uvedu příklad náhledu na dokument MS Wordu (.docx):



(Náhled na obsah začátku dokumentu MS Word - .docx)

Myslím, že je zřejmé, že z pohledu informačního obsahu jsou oba příklady přibližně na stejné úrovni. Při troše snahy sice můžeme v té zmeti znaků najít semtam nějaké to slovo nebo jiný relativně smysluplný čitelný text, nicméně je to spíše náhoda.

Proč to tedy tak zdůrazňuji? Je to důležité z procesního pohledu na digitální stopy. **Pořízení binární kopie disku neznamena automaticky, že majitel binární kopie získá přímý přístup k informacím, které jsou na tomto médiu uloženy.** Získat přístup k informacím na binární kopii je dokonce ještě náročnější než získat přístup k datům přímo z disku. Jestliže současné systémy nemají problém přečíst data (tedy soubory) z disku, který k počítači připojíme například přes USB rozhraní, není standardní součástí běžných systémů, aby bylo možné stejně jednoduše připojit do systému binární obraz¹⁵⁾ a tímto způsobem zpřístupnit jeho obsah (tedy soubory).

Jestliže získání přímého přístupu k souborům na pevném disku patří již ke standardním dovednostem běžného uživatele (určitě v případě připojení přes USB rozhraní), pro získání přístupu k datům binární kopie musí být počítač vybaven specifickým programovým vybavením¹⁶⁾, případně musí být uživatel pro tyto účely speciálně proškolen¹⁷⁾.

¹⁴⁾ Samozřejmě v případě, že není rozdělen na několik menších (dělalo se to v minulosti z důvodu kapacitních limitů cílových disků v průběhu zajištění).

¹⁵⁾ Relativně nejjednodušším způsobem to umí operační systémy od společnosti Apple a pak také operační systémy na bázi linuxu

¹⁶⁾ Například asi nejpoužívanější nástroj pro MS Windows: <https://marketing.accessdata.com/ftkimager4.2.0>

¹⁷⁾ Například použití příkazu „mount“ pro operační systém Linux, více na: https://www.forensicswiki.org/wiki/Mounting_Disk_Images

Binární kopie disku ale hraje důležitou roli i z jiného důvodu. Výše jsem prohlásil, že cílem při zajištění dat by mělo být zajištění co možná kompletního obsahu všech dostupných dat, protože předem prakticky nikdy nelze říct, která data bude potřebné zkoumat. Forenzní binární kopie obsahuje veškerá data, která jsou na disku uložena. Ale tady je nutné upozornit, že nejenom soubory, které jsou na původním, originálním disku viditelná uživatelem, ale i data uživatelem smazaná (v případě MS Windows i většinu dat, která byla vysypána z koše).

Je tedy nutné si zapamatovat, že forenzní binární kopie obsahuje veškerá data, která kdy byla na disku zaznamenána a nebyla přepsána jinými daty.

Forenzní binární kopie obsahuje veškerá data, která kdy byla na disku zaznamenána a nebyla přepsána jinými daty

Dekódování

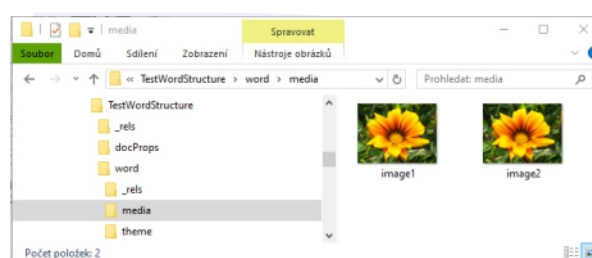
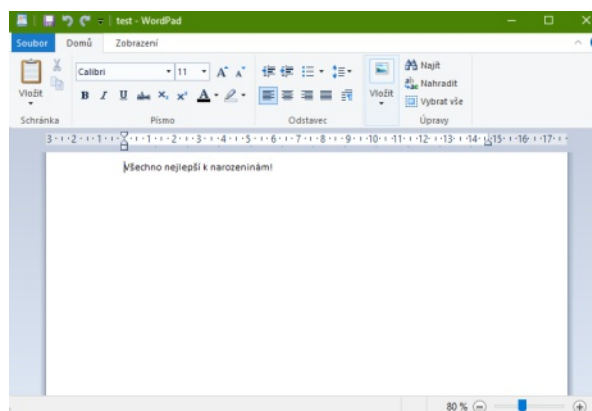
Vraťme se teď k tomu, co jsem už uvedl výše, tedy že „digitální informace je latentní a kódovaná. Až poté, co je získána a dekodována, je možné začít posuzovat její relevantnost“.

V předchozím jsem popsal základní standardní způsob získání digitální informace pomocí pořízení forenzní binární kopie. Teď se podíváme v hrubých rysech na to, co je potřebné udělat, abychom obsah binární kopie zpřístupnili ve vnímatelné podobě. Stručně by se dalo říct, že obsah musíme dekodovat.

Prakticky pro veškerou digitální techniku v současnosti platí, že data jsou ukládána v nějaké formě strukturovaných záznamů (snad kromě prostých textových dat - říká se jim „plain text format“). Pro získání informací, které jsou uloženy v nějaké formě strukturovaných dat je nezbytnou podmínkou znalost struktury takových dat. Dobrým příkladem takového strukturovaného záznamu je předchozí obrázek, kde je vidět náhled na obsah dokumentu napsaného v MS Wordu. Bez specifického software, který zná strukturu dat a je schopen ji převést do vnímatelné podoby, mnohdy neexistuje možnost, jak se k datům dostat. Navíc již ze samotné podstaty strukturovaných dat vyplývá, že taková data obsahují ve své specifické struktuře různá data, jinak by asi nemělo smyslu je strukturovat. Proto je také důležité si určit, jaký typ (nebo druh) dat požadujeme zpřístupnit.

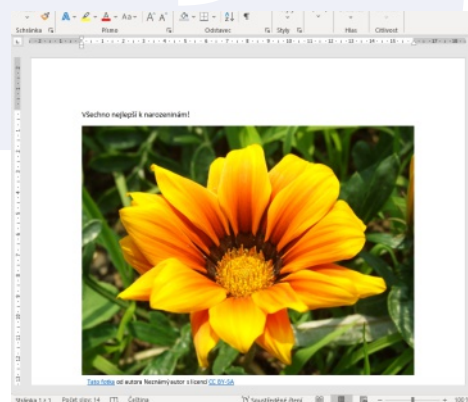
Ukážeme si to opět na příkladu souboru test.docx vytvořeného programem MS Word. K zobrazení části jeho dat lze použít např. program WordPad a výsledek takového zobrazení bude přibližně následující:

Existuje však ale i jiný pohled na obsah souboru vytvořeného v programu MS Word, který poskytne jiný pohled na jiný obsah dat tohoto souboru:



Jestliže první obrázek dokumentuje textová data, která jsou obsahem souboru (vytvořeného MS Wordem) pomocí programu WordPad, druhý obrázek dokumentuje vnitřní strukturu tohoto souboru (pomocí programu „Průzkumník“) a dokazuje například to, že obsahem souboru test.docx je kromě výše uvedeného textu (a kromě určitě dalších informací) také obrázek (obrázky?), který program WordPad neumí zobrazit. Tento příklad neuvádím proto, abych vysvětlil do detailu vše, co je na obrázcích vidět a proč to tak je. Uvádím jej proto, abych dokumentoval možnosti a úskalí procesu získání informací, které jsou uloženy ve strukturovaných dokumentech.

Pro úplnost uvedu také, jak vypadá obsah souboru test.docx tak, jak byl vytvořen programem MS Word:

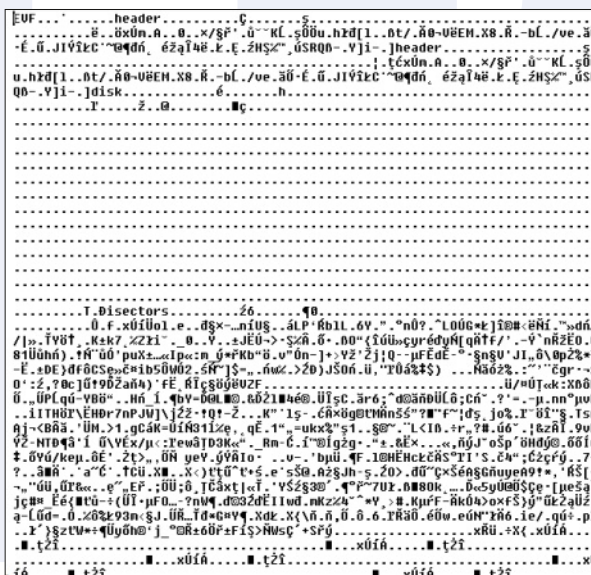


Když se vrátím k původní problematice forenzní binární kopie a její dekodování, je nutné uvést, že i forenzní binární kopie je ve své podstatě strukturovaným souborem. Je tedy pouze otázkou znalosti vnitřní struktury takového souboru a schopnosti vhodného programového nástroje tuto strukturu

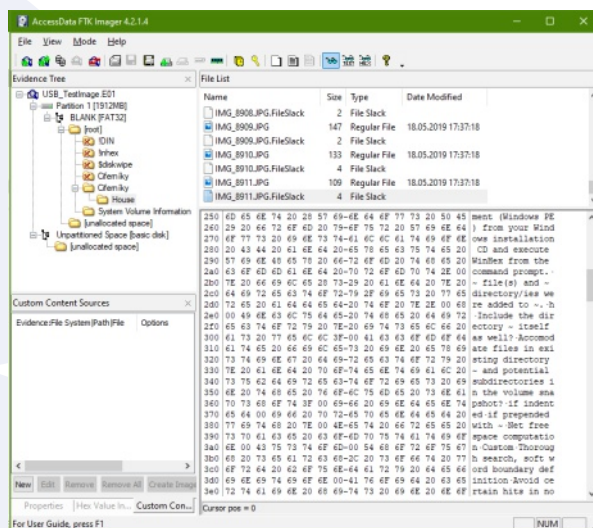
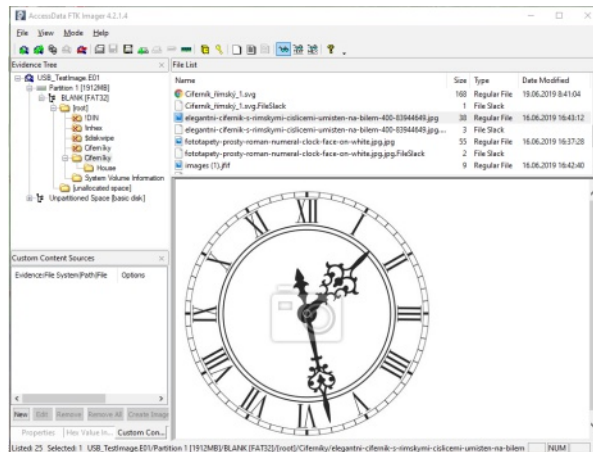
korreктně analyzovat (dekódovat) a zobrazit obsah této struktury ve vnímatelné podobě. A protože forenzní binární kopie se používá prakticky výhradně pro účely digitální forenzní analýzy, programy, které umí zobrazit ve vnímatelné podobě její obsah, nejsou běžně k dispozici a tudíž získat přístup k datům, které jsou obsahem forenzní binární kopie lze v zásadě pouze pomocí speciálních forenzních nástrojů.

Podobně jako v předchozím příkladě se souborem test.docx, který byl vytvořen pomocí MS Wordu, i u forenzní bitové kopie je důležité, jaké informace potřebujeme z něj získat a jaký nástroj k tomu použijeme. Je jasné, že asi nejlepšími nástroji pro dekódování forenzní bitové kopie by měly být takové, které jsou schopny dekódovat a zpřístupnit veškeré informace, které se na forenzní bitové kopii nacházejí. To je samozřejmě ideál, takové nástroje pravděpodobně existovat nikdy nebudou.

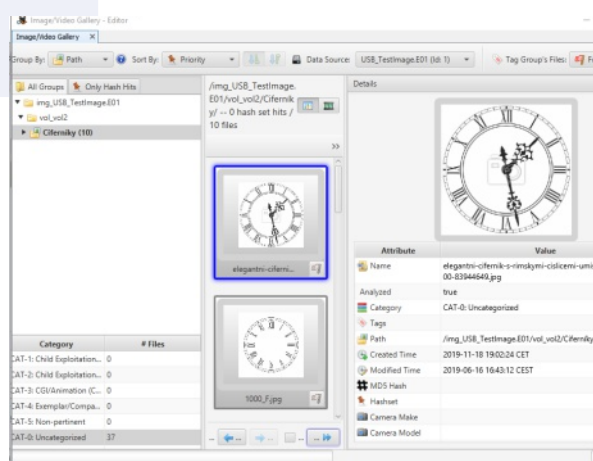
Už jen pro názornost předchozího tvrzení zde uvedu několik ukázek, jak lze zobrazit do vnímatelné podoby vnitřní strukturu forenzní binární kopie disku (v tomto konkrétním případě USB Flashdisku). Nejprve si ukážeme náhled na začátek takové kopie bez analýzy její struktury, jak to bylo uvedeno již v předchozím textu:

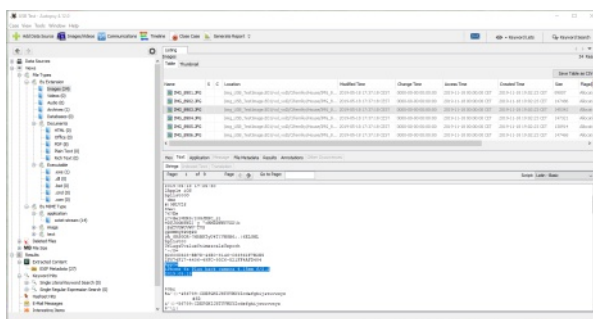


Když se ale na taková data podíváme nějakým vhodným programovým nástrojem, který umí analyzovat jejich vnitřní strukturu (v tomto konkrétním případě jsem použil program FTK Imager), můžeme získat například informace o souborech, které jsou na daném USB Flashdisku uloženy (u některých souborů včetně jejich náhledu). Nicméně těch informací, které lze z tohoto souboru získat je podstatně více. Na následujících obrázcích uvidíte například text, který byl někdy v minulosti na USB Flashdisku zaznamenán. A když se podíváte do horní levé části screenshotu, uvidíte tam přeškrtnuté adresáře, tedy adresáře, které byly v minulosti smazány.

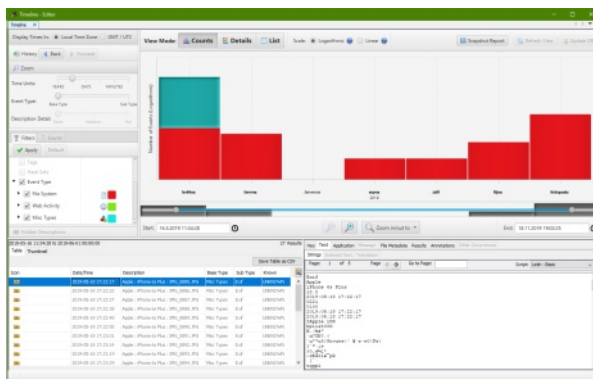


Ale při použití kvalitnějších forenzních analytických programů (aktuálně byl použit program Autopsy) lze převést do vnímatelné podoby množství dalších informací, které jsou ve forenzní binární kopii zaznamenány, např. informace o tom, kdy a jakým fotoaparátem byly pořízeny některé z fotografií:





A ještě jedna ukázka, kde jsou nalezené soubory seřazené podle časové osy včetně některých detailů o souborech:



Jsem přesvědčen, že už jen zde uvedená a pouze malá ukázka z toho, jaké všechny informace může forenzní binární kopie disku obsahovat, je dostatečně ilustrativní. Dokazuje to v našem kontextu minimálně to, že **forenzní binární kopie dat je složitý strukturovaný soubor, který při správné analýze a dekodování může poskytnout velké množství informací** pro účely digitální forenzní analýzy primárně a v konečném důsledku i pro samotný vyšetřovací proces.

Jestliže se nám povede správně dekodovat forenzní binární kopii, jako výsledek dostaneme v zásadě tři základní skupiny dat.

Jsou to tzv. **platná data**, tedy data a soubory, které v době pořízení forenzní binární kopie existovaly na paměťovém médiu a byly zároveň běžným způsobem přístupná i uživateli standardními prostředky operačního systému. Zjednodušeně řečeno se jedná o ta data a soubory, které uživatel mohl sám běžným způsobem práce s počítačem vidět a pracovat s nimi.

Druhou skupinou jsou **smazaná data**. Když vyjdeme z výše uvedeného tvrzení, že na paměťovém médiu jsou uchována veškerá data, která kdy na něj byla zaznamenána a nebyla přepsána jinými daty, tak jsou to vlastně tato veškerá data bez platných dat. Pro ilustraci to přiblížím tak, že to jsou všechna data/soubory, které kdy byly na paměťovém médiu uloženy, následně v průběhu doby smazány a nebyly přepsány jinými daty.

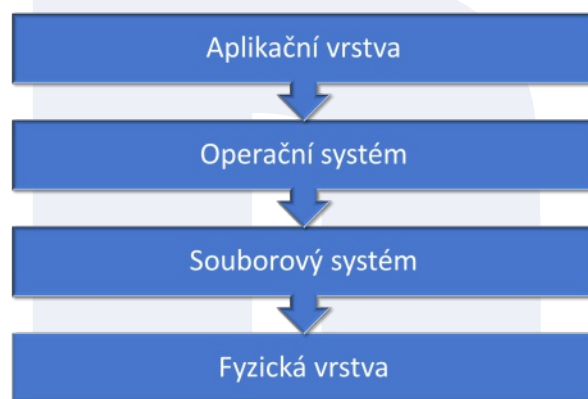
Tou poslední skupinou dat jsou **odvozená data**. Jak už se samotného názvu vyplývá, tato data nejsou na

forenzní binární kopii uložena přímo tak, jak je možné je zpřístupnit ve vnímatelné podobě, ale jsou nějakým definovaným způsobem z takových dat odvozena nebo vypočtena.

Hierarchie systémů

Než se pokusím v obecné rovině vysvětlit základní rozdíl mezi platnými, smazanými a odvozenými daty, pro lepší pochopení musím odbočit a stručně a zjednodušeně objasnit to, čemu se říká hierarchie systémů, které se vyskytují prakticky ve všech běžně používaných výpočetních systémech.

Pro tento účel si hierarchii systémů v počítačích rozdělíme do čtyř základních vrstev:



1. **Fyzická vrstva**. To je samotné záznamové médium, fyzická reprezentace paměťového média. Je to například magnetická páska, samotné čipy paměti RAM, paměti v USB Flash discích nebo na SSD discích, plotny na harddisku a podobně.

2. **Souborový systém**¹⁸⁾. Souborový systém je vlastně předpis, soubor pravidel, jakými jsou data na fyzickou vrstvu ukládána tak, aby bylo možné s nimi efektivně pracovat. Souborový systém určuje množství parametrů, například to, že lze soubory pojmenovávat, že je lze ukládat do nějaké struktury adresářů, že lze s nimi ukládat další dodatečné informace (tzv. meta-data) jako velikost souboru, datum pořízení, vlastníka dat a podobně. Souborový systém také v určité míře určuje efektivitu využití prostoru fyzické vrstvy, rychlost přístupu k datům, spolehlivost dat a množství dalších parametrů. Souborový systém musí také zohlednit různou fyzikální podstatu fyzické vrstvy, protože je zřejmé, že data na magnetickou pásku je potřebné ukládat jiným způsobem než data na harddisk.

3. **Operační systém**¹⁹⁾. Operační systém plní v počítačích funkci základního programového vybavení, které zajišťuje vlastní funkčnost počítače. Kromě mnoha dalších důležitých funkcionalit, v naší konkrétní situaci (ve vztahu k datům uloženým na paměťových nosičích) zajišťuje operační systém vlastní

¹⁸⁾ Problematika různých souborových systémů je ovšem mnohem složitější. Existuje velké množství souborových systémů a také velké množství důvodů pro použití toho kterého souborového systému. Úvod do této problematiky lze najít např. v https://en.wikipedia.org/wiki/File_system

¹⁹⁾ Komplexitu úloh operačního systému popisuje v obecné rovině ve stručnosti například https://en.wikipedia.org/wiki/Operating_system

realizaci pravidel souborového systému (ať už přímo, tedy v jádru operačního systému, nebo pomocí dodatečných programových modulů/driverů).

4. Aplikační vrstva. Aplikační vrstvou rozumíme vlastně všechny programy, které umožňují uživateli provádět s pomocí počítače úlohy, tedy obecně zpracovávat data. Jsou to ty programy, které uživatel používá k řešení jeho problémů, např. webové prohlížeče, kancelářské balíky, grafické editory, databáze a další. Mezi ně patří např. i program „Průzkumník“ z MS Windows, byť se jeví spíše jako součást operačního systému a s operačním systémem se také instaluje.

To, jak s daty uživatel pracuje, určuje právě aplikační vrstva. Jestliže aplikace požaduje zapsat nový soubor, v rámci aplikace se vytvoří sada dat, která daný soubor tvoří (např. data dokumentu MS Word) a zadá požadavek na operační systém, aby tato data uložil. Operační systém data z MS Wordu vezme a podle pravidel souborového systému je připraví tak, aby je bylo možné zapsat na dané fyzické paměťové médium.

Při standardní práci výpočetního systému se na práci z daty podílejí hierarchicky všechny čtyři subsystémy/vrstvy. Nicméně to, zda je ten který soubor, který je uložen na fyzickém paměťovém médiu použitelný/viditelný pro vyšší systémy/vrstvy, je určeno právě pravidly souborového systému. Jestli je daný soubor zaznamenaný v něčem podobném, jako je „**katalog souborů**“ souborového systému, tak je soubor viditelný a použitelný. Jestli je ale z tohoto „katalogu“ vyřazen, operační systém již nemá žádnou možnost, jak se k danému souboru standardními postupy dostat. Vyřazením souboru z „katalogu“ souborového systému ztrácí tento souborový systém veškeré informace o souboru a již jej tedy nemůže prostřednictvím operačního systému nabídnout ke zpracování aplikační vrstvě.

Platná data

Platná data je skupina dat, která jsou jako platná označena použitým souborovým systémem a tedy jsou viditelná a použitelná daným operačním systémem a standardními aplikacemi. Jinak řečeno, jsou to soubory a adresáře, které dostaneme při výpisu adresářové a souborové struktury (výpis "katalogu souborů"). U našeho výše uvedeného příkladu dat na USB Flashdisku by takový výpis v prostředí MS Windows mohl vypadat například následovně:

Název	Datum úpravy	Typ	Velikost
Hrout	18.11.2019 18:02	Šablona souborů	
1000_F_1.jpg	17.06.2019 13:27	Soubor JPG	113 kB
1000_F_1.jpg	17.06.2019 13:39	Soubor JPG	139 kB
1000_F_28166028_hQWgD8081VrytWakighuONLZL01049.jpg	16.06.2019 17:48	Soubor JPG	122 kB
1000_F_7031961_Fa27232a7a0L8fMhMhcyKXAK3pdu.jpg	16.06.2019 17:45	Soubor JPG	139 kB
70234732-horloge-a-chiffres-romains.jpg	16.06.2019 17:36	Soubor JPG	122 kB
Cifernik_rimsky_1.svg	17.06.2019 17:11	Chrome HTML Document	191 kB
Cifernik_rimsky_2.svg	18.06.2019 18:41	Chrome HTML Document	169 kB
elegantni-cifernik-v-rimskym-cislocem-umisten-na-bilem-400-83944649.jpg	16.06.2019 17:40	Soubor JPG	38 kB
fortatagety-prosty-roman-numeral-clock-face-on-white.jpg.jpg	16.06.2019 17:37	Soubor JPG	55 kB
images (1).jif	16.06.2019 17:42	Soubor JIF	9 kB
images.jif	16.06.2019 17:42	Soubor JIF	6 kB
stazeny soubor.jif	16.06.2019 17:39	Soubor JIF	9 kB

Respektive textový výstup výpisu souborové a adresářové struktury by mohl vypadat takto:

```
[Ciferniky]
Cifernik_rimsky_1.svg          171 328 18.06.2019 09:41  -
Cifernik_rimsky_2.svg          194 699 17.06.2019 17:11  -
1000_F_1.jpg                   142 187 17.06.2019 13:39  -
1000_F_1.jpg                   114 762 17.06.2019 13:27  -
1000_F_28166028_hQWgD8081VrytWakighuONLZL01049.jpg 124 111 16.06.2019 17:48  -
1000_F_7031961_Fa27232a7a0L8fMhMhcyKXAK3pdu.jpg 141 876 16.06.2019 17:45  -
elegantni-cifernik-v-rimskym-cislocem-umisten-na-bilem-400-83944649.jpg 38 888 16.06.2019 17:40  -
images (1).jif                 9 932 16.06.2019 17:42  -
images.jif                     6 814 16.06.2019 17:42  -
9 121 16.06.2019 17:39  -
fortatagety-prosty-roman-numeral-clock-face-on-white.jpg.jpg 66 289 16.06.2019 17:37  -
70234732-horloge-a-chiffres-romains.jpg 122 938 16.06.2019 17:36  -
1 104 B v 12 souborech

[Ciferniky\Hrout]
Hrout_1000_F_1.jpg             73 870 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             72 794 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             88 292 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             85 221 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             78 123 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             85 746 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             89 978 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             89 819 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             88 271 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             86 410 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             110 127 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             101 288 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             108 734 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             124 193 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             93 087 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             126 118 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             89 807 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             147 486 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             148 342 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             147 721 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             138 914 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             147 468 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             117 828 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             113 099 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             180 361 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             138 883 18.06.2019 18:37  -
Hrout_1000_F_1.jpg             111 297 18.06.2019 18:37  -
2 848 B v 27 souborech
```

Na tomto místě je nutné zmínit, že ne všechna platná data (soubory) jsou vždy viditelná pro běžného uživatele. Do toho, kolik a jaká data uživatel vidí, vstupuje v současných systémech možnost přidělování uživatelských oprávnění a přístupových práv. Obecně nemá běžný uživatel přístup ke všem datům (souborům), taková práva má pouze administrátor systému. Takže jestliže mluvím o datech (souborech), tak ve své podstatě mluvím o datech/souborech, ke kterým má oprávnění a tedy je vidí administrátor systému.

Abych byl úplně korektní, tak pod pojmem platná data lze chápat veškerá data (soubory), která jsou registrována/zaznamenaná v použitém souborovém systému (mají záznam v "katalogu souborů" použitého souborového systému).

Platná data mohou být dále členěna podle různých kritérií. Těch kritérií může být relativně hodně a vždy závisí od účelu, pro který dané kritérium použijeme, resp. podle charakteristik, které lze u nich zjistit nebo podle atributů, které si o datech (souborech) vede použitý souborový systém nebo použitý operační systém.

Situace s rozdělením platných dat do různých kategorií je však často velice komplikovaná, protože kritéria, podle kterých by bylo možné jednotlivé kategorie dat jednoznačně určit v aktuálně používaných systémech nejsou jednoznačně stanoveny, kategorizace typů a druhů souborů není standardizována. Spolehlivé určení jednotlivých kategorií dat je prakticky vždy úkolem pro podrobnou a někdy i časově a výpočetně náročnou analýzu.

Smazaná data

Smazaná data je skupina dat, která jsou jako smazaná označena použitým souborovým systémem

a tedy nejsou již daným souborovým systémem dále viditelná a tedy také již nejsou dostupná daným operačním systémem a standardními aplikacemi.

Ke smazaným datům je nutné dodat dvě poznámky:

1. Z fyzikálních principů práce paměťových médií vyplývá, že **proces zápisu dat na tato média je vždy pomalejší, než proces čtení**. Změnit záznam tím, že je přepsán něčím „neutrálním“, např. nulami nebo jedničkami, tedy že původní data jsou „přemazána“, je vždy časově náročnější než pouhé jeho čtení. Z toho důvodu v počítačích při smazání souboru prakticky standardně nikdy nedochází ke (pře/s)mazání dat, které daný soubor obsahoval (to by zabralo v počítači hodně času), ale smazaný soubor je pouze vyřazen ze seznamu souborů („souborového katalogu“) daného souborového systému a datový prostor, který tento soubor zabíral, je z pohledu souborového systému označen jako neobsazený a tedy vhodný k zápisu nových dat/souborů. **Je tedy nutné si pamatovat**, že smazáním souboru nedochází k přepsání dat tohoto souboru, data samotná na paměťovém médiu zůstávají, jen je takový datový prostor označen jako volný, jako použitelný pro zápis nových dat.

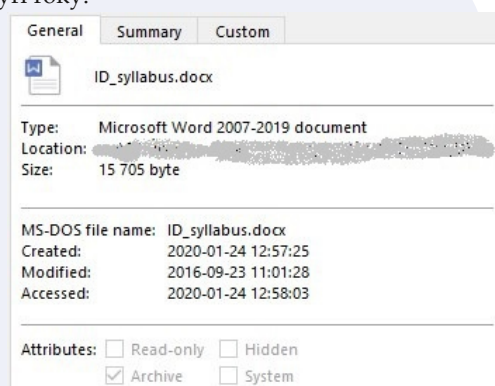
2. Prakticky všechny současné operační systémy nabízejí uživatelům „obezličku“, - **místo tzv. „tvrdého“, mazání souborů standardně přesouvají tyto soubory do tzv. „koše**„. Je to proto, aby se uživateli nabídla možnost omylem smazaný soubor obnovit. **Je tedy nutné si pamatovat**, že běžným smazáním souboru (do „koše“) nedochází k jeho skutečnému smazání (tedy uvolnění souborem obsazeného prostoru pro zápis nových dat), ale jenom k „přesunu“, těchto dat (ani to není přesný popis procesu, ve skutečnosti se samotná data nikam nepřesouvají, ale pro zjednodušení to nebudu popisovat detailně) do zvláštního adresáře, který slouží jako mezistupeň před „tvrdým“, vymazáním souboru. Určitou stinnou stránkou obezličky zvané „koš“, je to, že jeho obsah časem zabírá hodně místa na paměťovém médiu a je jej nutné čas od času tzv. „vysypat“.. **A až „vysypání koše“, je proces, kterým jsou data reálně smazána - tedy soubory z „koše“, jsou vyřazeny ze seznamu souborů daného souborového systému a prostor, který zabírají, je poskytnut k využití pro nová data/soubory.**

Zopakuji ještě jednou výše uvedenou definici: „*smazaná data je skupina dat, která jsou jako smazaná označena použitým souborovým systémem*“. Neznamena to, že jsou ztracena absolutně, jenom prostor, který zabírají, je označen jako použitelný pro nová data. Do doby, než je tento prostor přepsán novými daty, jsou data pořád zaznamenána na záznamovém médiu, jen již nejsou přístupná standardním způsobem přes souborový a operační systém²⁰⁾. Jestliže ale lze použít nějaký speciální, nestandardní způsob, který obchází pravidla souborového (a operačního) systému, lze získat přístup i k datům, která byla smazána tzv. „natvrdo“, tedy která byla vyřazena z „katalogu“ souborového systému.

Odvozená data

Jak je asi zřejmé z předchozího, v platných a smazaných datech je potenciálně možné nalézt velmi velké množství různých informací. Některé z nich jsou přímo aplikovatelné a přímo relevantní pro vyšetřování, jiné však je potřebné interpretovat, případně uvést do vzájemných vztahů.

Asi typickým a obecně známým příkladem mohou být informace, které jsou prezentovány např. v dokumentech MS Word a vztahující se časovým informacím, které jsou zobrazovány ve vlastnostech souborů. Velice často se můžete setkat se situací, kdy časové údaje, které zobrazuje MS Word, jsou zcela nesmyslné. Jak je uvedeno na následujícím obrázku, datum (poslední) úpravy souboru (Modified) předchází datum vytvoření (Created) souboru téměř o čtyři roky.



Výše uvedený příklad je typickou ukázkou situace, kdy zdánlivě přímo aplikovatelná informace vyžaduje vysvětlení, resp. interpretaci informací, které jsou (zcela zřejmě) v podání MS Word nepravdivé (objektivně musím uvést, že v poslední verzi programu Word pro Office 365 jsem už napodobně nesmysly z dílny Microsoftu nenarazil).

Mezi odvozená data patří například také výsledek analýz vzájemné komunikace různých subjektů a jejich zobrazování ve formě např. interaktivních síťových grafů²¹⁾.

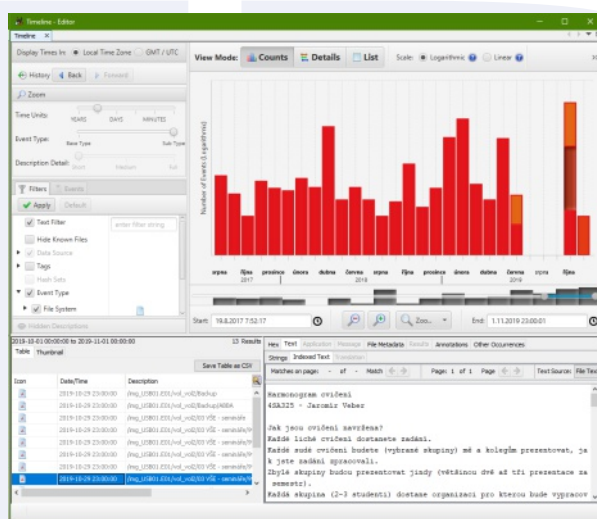


²⁰⁾ Viz výše, část věnována "Hierarchii systémů"

²¹⁾ Zdroj: https://www.analyticsvidhya.com/blog/2019/11/graph-feature-extraction-deepwalk/socialnetworkanalysis_graph/

V tomto případě (výše uvedený síťový graf) ani tak nejde o zcela novou informaci, kterou by nebylo možné (teoreticky) dovodit i z jednotlivých informací, ale jejich vizualizace podává zcela nový a přehledný pohled na věc a umožňuje efektivně a rychle dospět k dalším požadovaným informacím a závěrům.

Nicméně i obyčejná statistická data mohou přinést relevantní informace pro vyšetřování, jak je vidět např. i z následujícího grafu - časové osy (podle časových informací ze souborového systému), kdy v tomto konkrétním případě je zřejmé, že v době červenec až říjen 2019 nebylo s daty pracováno. Tedy závěr, že s daty nebylo v srpnu 2019 pracováno, je odvozená informace, která vychází z časových informací (metadat) souborového systému.



Odvozená data jsou tedy taková data, která vznikla z platných nebo smazaných dat. Způsob, jakým jsou odvozená data získána je různý, od elementárních operací až po výsledky aplikace umělé inteligence (AI). Nicméně výběr konkrétního způsobu získání odvozených dat je vždy na znalci a měl by být podložen relevantním zdůvodněním a použita metoda by měla být hodnověrná a ověřitelná.

Na toto téma by se dalo hodně diskutovat, zejména problematika využití prvků AI je z pohledu dosahování exaktnosti výsledků forenzní analýzy značně diskutabilní a seriózní diskuse kolem použití AI v oblasti forenzních věd se zatím moc nevede.

ZÁVĚR

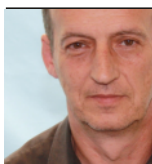
Pro správné pochopení toho, čím se DFA zabývá, jaké metody a postupy používá a jak využívat její výstupy, je klíčové správně definovat předmět DFA. Jen tak lze správně pochopit definici DFA a tím i správně pochopit a použít výstupy, které DFA poskytuje.

Na základě poznání výjimečné vlastnosti digitální stopy, kdy její originál a kopie jsou identické, vzájemně zaměnitelné a tím také mající stejnou vypovídající a důkazní hodnotu, dospějeme k závěru, že digitální

stopa je informace. Informace, jako nehmotná entita, nezávislá na její materiální interpretaci, tedy nezávislá na nosiči.

Jediným specifikem zůstává skutečnost, že tato informace je kódována binárním kódem, stručně řečeno že informace je digitální, tedy že je reprezentována kódem, který je tvořen nulami a jedničkami. Pro digitální informaci, jako pro nehmotnou entitu tedy platí, že nezávisí na tom, jaká je její materiální interpretace, tedy jakým materiálním způsobem jsou realizovány ty nuly a jedničky.

Ve své podstatě není rozhodující, jak a na jakém fyzikálním základě ty jedničky a nuly vznikly (světlo nebo tma, kladný nebo záporný elektrický náboj nebo magnetické nebo elektromagnetické pole nebo cokoliv jiného, co je schopno rozlišit dva stavy určité veličiny). Podstatné je, na základě jakého procesu v rámci výpočetního systému vznikly a jaká informace je pomocí nich kódována.



Ing. Marián Svetlík, Sr., soudní znalec, pracuje jako specialista na digitální forenzní analýzu Czech Cybercrime Centre of Excellence na Masarykově univerzitě v Brně, přednáší informační bezpečnost a počítačovou kriminalitu na vysokých školách. Je šéfredaktorem časopisu Digital Forensic Review, členem řídicího výboru Institutu pro Digitální Forenzní Analýzu, podílí se na mezinárodních projektech spojených s problematikou Digital Forensics a Cybercrime.

FORENSIC IMAGE AND VIDEO ENHANCEMENT



AMPED
FIVE

THE COMPLETE SOFTWARE SOLUTION

PROCESS AND ANALYZE DIGITAL IMAGES FROM CCTV, BODY WORN CAMS,
MOBILE PHONES, DRONES, LATENT FINGERPRINTS, AND MORE

SCIENTIFIC WORKFLOW DRIVEN BY FORENSIC NEEDS

Distribuci, prodej a školení v ČR a SR zajišťuje:



FIREWIRE REVOLUTION CZ
by forensee
Belgická 19
120 00 Praha 2

www.firewire-revolution.cz
info@firewire-revolution.cz

+420 315 55 88 70

forensee
digital forensics