Sociální sítě

Sběr a analýza dat v souvislosti s bezpečnostními incidenty

Technická zpráva FIT VUT v Brně

Ing. Radek Burget, Ph.D.

Technická zpráva č. FIT-TR-2017-11
Fakulta informačních technologií, Vysoké učení technické v Brně

Last modified: 4. ledna 2018

Sociální sítě: Sběr a analýza dat v souvislosti s bezpečnostními incidenty

Ing. Radek Burget, Ph.D.

Vysoké učení technické v Brně, email: burgetr@fit.vutbr.cz

Abstrakt Sociální sítě jsou v současnosti velmi populární komunikační prostředek umožňující sdílení velkého množství informací různých typů mezi uživateli, ale i jejich veřejné publikování. Z pohledu zpracování dat z bezpečnostních incidentů představují sociální sítě jednak potenciální zdroj těchto incidentů, současně však i potenciálně bohatý zdroj informací, který je možno využít při jejich zkoumání. V tomto dokumentu představujeme analýzu relevantních bezpečnostní incidentů a následně pak relevantních zdrojů informací, které zahrnují zejména nejpoužívanější sociální sítě a data o využívání webových prohlížečů na lokálních počítačích. Dále pak popisujeme architekturu softwarového nástroje navrženého pro získání, uložení a vyhodnocení velkého množství dat z těchto velmi různorodých zdrojů. Tento nástroj byl navržen s ohledem na rozšiřitelnost, škálovatelnost a možnost práce v distribuovaném prostředí a navržené principy byly ověřeny na prototypové implementaci.

1 Úvod

Sociální sítě představují na jedné straně efektivní a široce využívaný prostředek mezilidské komunikace, na straně druhé se však mohou podílet na úniku osobních i firemních dat a mohou být zneužity pro širokou škálu kybernetických útoků. Dojde-li k takovému útoku, mohou sociální sítě představovat důležitý zdroj informací potřebných pro jeho další analýzu a to bez ohledu na to, zda byla sociální síť použita jako hlavní prostředek útoku, jako koordinační mechanismus nebo nebyla přímo využita vůbec, ale pachatelé útoku ji využívají při jiných aktivitách.

Při vyšetřování bezpečnostních incidentů mohou data získaná analýzou sociálních sítí pomoci se zodpovězením mimo jiné následujících otázek ohledně identity účastníků:

- Známe-li jméno osoby, který z profilů stejného jména na sociálních sítích patří dané konkrétní osobě?
- Jedná se o uměle vytvořený (podvržený) profil?
- Patří různé profily (obecně na různých sociálních sítích) stejné osobě nebo skupině osob?
- Které profily na sociálních sítích byly využívány pomocí daného zařízení (počítače, mobilního telefonu, apod.)?

– Existují nějaké kontakty mezi danými osobami?

Následně je pak možno tyto informace využít například pro konkretizování aktivity dané osoby v určitém časovém období, místa pobytu apod.

Typický scénář využití dat ze sociálních sítí pro vyšetřování je popsán např. v [20]: Je zajištěno digitální zařízení spojené s bezpečnostním incidentem (počítač, telefon, paměťové médium apod.) a cílem je shromáždit a analyzovat veškeré informace o online aktivitách uživatele tohoto zařízení. To předpokládá získání velkého množství informací z potenciálně mnoha heterogenních zdrojů, které zahrnují různé typy sociálních sítí, informace z lokálních zařízení a podobně a následně korelaci identit uživatelů napříč těmito zdroji. Mezi očekávané výsledky patří:

– Vzájemná souvislost profilů na sociálních sítích
– Souvislost s dalšími zdroji dat, jako např. zajištěný mobilní telefon nebo počítač používaný pro přístup k sociální síti, monitorování síťového provozu apod.

Vzhledem k obrovskému množství potenciálně zajímavých informací, které je v takovém případě třeba zpracovat je nezbytné mít k dispozici nástroje, které umožní následující operace:

– Získat a uložit rozsáhlá data z mnoha heterogenních zdrojů
– Automaticky detekovat strojově rozpoznatelné souvislosti v těchto datech
– Efektivně vizualizovat získané výsledky a umožnit jejich interaktivní procházení za účelem dalšího zkoumání expertem.

Je třeba mít na paměti rovněž škálovatelnost zvolených metod a možnost spouštět popsané úlohy v distribuovaném prostředí pro dosažení praktické použitelnosti na velkých objemech dat.

Cílem tohoto dokumentu je analyzovat různé podoby kybernetických útoků souvisejících se sociálními sítěmi, identifikovat dostupné zdroje dat, které jsou využitelné pro další zkoumání těchto útoků a navrhnout nástroje umožňující výše popsané operace nad těmito zdroji. V kapitole 2 shrnujeme využití jednotlivých sociálních sítí se zaměřením na stav v České republice. V kapitole 3 rozebíráme jednotlivé druhy útoků souvisejících se sociálními sítěmi a v kapitole 4 zmiňujeme existující softwarové nástroje. Dále popisujeme zkoumané zdroje dat, mezi něž patří informace dostupné na lokálních počítačích, které rozebíráme v kapitole 5 a jednotlivé sociální sítě a jejich aplikační rozhraní popsaná v kapitole 6. V kapitole 7 se pak zabýváme návrhem architektury softwarového nástroje vyhovujícího výše uvedeným charakteristikám.

2 Sociální sítě používané v České republice

Využití sociálních sítí v České republice kopíruje s jistým zpožděním celosvětové trendy. Jak vyplývá z výzkumů agentury AMI Digital provedených v roce

Pořadí	Sociální síť	Uživatelů v 04/2017	Uživatelů v 10/2017
1.	Facebook	4 800 000	4 900 000
2.	YouTube	4 750 000	4 800 000
3.	Instagram	1 500 000	1 600 000
4.	LinkedIn	1 300 000	1 400 000
5.	Twitter	400 000	400 000
6.	Snapchat	400 000	80 000

Tabulka 1. Počty aktivních uživatelů sociálních sítí v České republice v dubnu a říjnu 2017 (průzkum AMI Digital [14,15])

2017, nejpoužívanější sociální sítě jsou aktuálně Facebook a YouTube s velkým rozdílem oproti ostatním sociálním sítím. Aktuální počty uživatelů jsou shrnuty v tabulce 1.

Průzkum bohužel nezahrnuje poměrně populární sociální síť Google+. Pokud se však podíváme na starší průzkum stejné agentury z roku 2016 [2], který mapoval procenta českých uživatelů využívajících některou sociální síť, dostaneme čísla uvedená v tabulce 2.

Pořadí	Sociální síť	% uživatelů
1.	YouTube	94 %
2.	Facebook	93 %
3.	Google+	74 %
4.	Twitter	30 %
5.	Instagram	25 %
6.	LinkedIn	24 %

Tabulka 2. Procento uživatelů internetu v ČR využívající jednotlivé sociální sítě (průzkum AMI Digital [2])

V dalším výzkumu se proto zaměřujeme na výše uvedené nejpopulárnější sociální sítě. Současně však bereme v úvahu rychlý vývoj v této oblasti a tedy předpokládanou nutnost rychle zpracovat data i z dalších sociálních sítí a datových zdrojů.

3 Kybernetické útoky na sociálních sítích

Bezpečnostní hrozby na sociálních sítích zahrnují poměrně širokou škálu útoků a bezpečnostních incidentů. Typickým cílem útoků je zejména získání osobních údajů uživatelů, zneužití účtu na sociálních sítích za účelem šíření obtěžujících zpráv, vylákání přihlašovacích údajů k dalším službám, ale i nebezpečnějších jevů

jako obtěžování a vydírání. Obvykle používané dělení bezpečnostních hrozeb na sociálních sítích [10,21,23] je rozebráno v následujících kapitolách.

3.1 Tradiční útoky aplikované na sociální sítě

Jedná se o typy útoků známých i mimo oblast sociálních sítí, např. z webu nebo e-mailových zpráv. Sociální síť zde představuje alternativní komunikační prostředek, který je zneužitelný podobným způsobem. Cílem je obvykle získání důvěrných informací včetně přihlašovacích údajů k dalším službám, čísla kreditní karty apod.

Phishing Jedná se o útok využívající v oklamání uživatele pomocí podvržené webové stránky nebo zaslané zprávy, která na první pohled působí věrohodně a přiměje uživatele nějakým způsobem sdělit požadované citlivé údaje, např. vyplněním do podvrženého formuláře. V kontextu sociálních sítí útočník obvykle sbírá dostupné osobní informace o uživateli (zejména osobní data a jména přátel) s cílem vytvořit věrohodnou zprávu, která přiměje uživatele sdělit některé konkrétní informace nebo navštívit podvrženou webovou stránku.

Malware Cílem útoku je prostřednictvím sociální sítě přimět uživatele nainstalovat do počítače škodlivý program (trojského koně, virus, apod.), případně provést instalaci bez vědomí uživatele s využitím některé bezpečnostní chyby v jeho zařízení. Nejčastěji využívaným mechanismem je sdílení podvržených URL adres prostřednictvím sociální sítě [7].

Spamming Jedná se o hromadné zasílání nevyžádaných zpráv pomocí komunikační infrastruktury dané sociální sítě.

Clickjacking Útok využívá podvržení falešných ovládacích prvků, které vzhledem připomínají ovládací prvky dané sociální sítě (např. tlačítko Like na Facebooku), ve skutečnosti však skrývají nějaký druh malware popsaný výše.

Klonování profilů S využitím osobních informací včetně fotografií, které jsou veřejně dostupné nebo získané jinými metodami, vytvoří útočník podvržený profil konkrétního uživatele a pokusí se s jeho využitím získat přístupová práva k informací o dalších uživatelích. Typicky je cílem přesvědčit ostatní uživatele, aby si přidali falešný profil do "přátel" nebo "kontaktů", čímž majitel podvrženého profilu získá přístup k dalším informacím o dotčených uživatelích, které nejsou veřejně publikované. Klonovaný profil lze rovněž využít pro osobní útoky pomocí šíření nepravdivých informací, podvržených fotografií a podobně [10].

3.2 Útoky související s multimediálním obsahem

Všechny sociální sítě umožňují svým uživatelům sdílet fotografie a videa, často doplněné o podrobnější údaje, jako např. přesná poloha pořízení fotografie, jména zachycených osob a podobně. Tento obsah je častým cílem útočníků cílících na získání citlivých údajů o osobách. Jedná se zejména o následující bezpečnostní hrozby:

Sdílení nechtěných informací v multimediálním obsahu Sdílené fotografie a videa mohou obsahovat zneužitelné informace o uživateli. Z fotografií je možno usuzovat např. na aktuální místo pobytu uživatele, rodinnou situaci, vytipovat adresu bydliště a podobně. Existující technologie, jako automatické rozpoznávání obličejů nebo řeči společně s možností explicitně označovat osoby v multimediálním obsahu, může útočníkovy poskytnout informace i o třetích osobách [6]. Podobně pokročilé algoritmy vyhledávání obrázků na základě obsahu (Content-Based Image Retrieval, CBIR) umožňují přesněji identifikovat objekty na obrázku, zpřesnit lokaci a podobně [10].

Skrytá komunikace pomocí multimediálního obsahu Multimediální obsah může být využit jako komunikační kanál pro šíření skrytých, na první pohled obtížně odhalitelných správ (steganografie) [24].

Metadata Fotografie i videa mohou obsahovat dodatečné informace (metadata) zahrnující místo a čas pořízení, typ fotoaparátu a mnohé další informace, které jsou potenciálně zneužitelné. Zatímco síť Facebook automaticky odstraňuje metadata před publikací obsahu, síť Google+ zachovává vše kromě geografických údajů a síť Flickr cíleně využívá dostupná geografická metadata pro nabízení obsahu uživatelům [21].

Sdílení odkazů Častou vlastností sociálních sítí je sdílení webových odkazů. V takovém případě se může útočník pokusit nahradit cíl odkazu vlastním, potenciálně nebezpečným obsahem, využít útoků typu Cross-site scripting (XSS) [10] a podobně.

3.3 Sociální hrozby

Veškeré citlivé informace o uživatelích získané výše uvedenými metodami jsou dále využitelné pro některou z dalších forem útoků, které jsou typické pro sociální sítě a mohou zahrnovat vydírání, finanční podvody, sexuální obtěžování, stalking, firemní špionáž a další.

4 Související výzkum a softwarové nástroje

Vytvoření nástroje pro analýzu dat publikovaných na sociálních sítích představuje komplikovaný technický problém. Hlavní překážku, kterou je nutné překonat, představuje skutečnost, že jednotlivé sociální sítě mají diametrálně odlišná aplikační rozhraní umožňující přístup ke zveřejněným datům na různých úrovních a s různým stupněm zabezpečení. Některé sociální sítě, jako například Twitter, slouží především k veřejnému sdílení informací a pro tento účel poskytují plnohodnotné a dobře dokumentované aplikační rozhraní. Jiné sítě zavádí různá omezení s cílem ochránit soukromí svých uživatelů a v některých případech (například LinkedIn) není k dispozici prakticky žádné použitelné aplikační rozhraní. Bližší rozbor jednotlivých sociálních sítí z tohoto pohledu uvádíme v kapitole 6.

Patrně z těchto důvodů se nám nepodařilo nalézt téměř žádný softwarový nástroj, který by přímo pracoval s daty dostupnými prostřednictvím různých sociálních sítí. Jediné nalezené řešení [12,18] existuje ve stavu funkčního prototypu a je zaměřeno na síť Facebook. Tento nástroj umožňuje (za předpokladu získání přístupových údajů) pořídit kopii profilu uživatele zahrnující příspěvky, multimediální obsah, osobní informace a další data za účelem dalšího zkoumání těchto dat, které již však musí být provedeno jinými prostředky. Dále existují akademické studie nástrojů pro zpracování dat ze sítě Twitter [11], nejblíže našemu přístupu, který prezentujeme v kapitole 7 je rámcové řešení popsané v [20]. Toto zkoumané řešení rovněž počítá s modulární architekturou zahrnující větší množství lokálních zdrojů a sociálních sítí, nezabývá se však škálovatelností a možností distribuovaného zpracování a nepodařilo se nám rovněž nalézt zmínku o tom, že by daný přístup byl skutečně implementován.

V oblasti forenzního zkoumání fyzických počítačů naopak existuje větší množství nástrojů, které jsou schopné vytvořit podrobnou zprávu o různých typech nalezených informací. Mezi nepopulárnější z nich patří komerční nástroje Magnet Internet Evidence Finder[1], EnCase Forensic[2], Internet Examiner Toolkit[3], z open source projektů např. log2timeline[4].

5 Informace o využívání sociálních sítí na lokálních počítačích

Všechny zkoumané sociální sítě jsou přístupné prostřednictvím vlastních klientských aplikací. V případě využití standardního počítače (stolní počítač, notebook) má klientská aplikace vždy podobu webové aplikace, která běží ve webovém prohlížeči uživatele. Uživatel tedy přistupuje k sociální sítě přes webové stránky provozovatele. V případě mobilních zařízení (mobilní telefon, tablet) jsou primárně využívány specializované aplikace jednotlivých sociálních sítí.

[1] https://www.magnetforensics.com/magnet-ief/
[2] https://www.guidancesoftware.com/encase-forensic
[3] http://www.siquest.com/
[4] https://github.com/log2timeline/plaso/wiki

V obou případech může používání sociálních sítí zanechat informační stopy na lokálním souborovém systému počítače nebo mobilního zařízení. Tyto stopy představují potenciálně zajímavý zdroj informací o využití sociálních sítí. V ideálním případě mohou lokální stopy pomocí se zodpovězením následujících otázek:

- **Využívá daný uživatel některé sociální sítě?** Jak často k nim přistupuje, jaké je datum poslední návštěvy? Zobrazil nějaký konkrétní profil nebo část obsahu?
- **Pochází některé lokální soubory ze sociální sítě?** Kdy byly staženy, kam byly uloženy?
- **Jaké je uživatelské jméno na sociální síti?** Další přístupové údaje?

Je však nutno poznamenat, že dané údaje jsou obvykle v mnoha ohledech nekompletní a poskytují pouze hrubý obrázek využití sociálních sítí na daném počítači. Je třeba vzít v úvahu následující omezení:

- Lokální stopy pokrývají pouze část interakcí uživatele s aplikací. V závislosti na konkrétní sociální síti může jít zejména o stahování souborů nebo záměrnou návštěvu jednoznačně identifikovatelné stránky (např. konkrétní profil jiného uživatele, zobrazení jednotlivé fotografie apod.)
- Uživatel může střídavě nebo i současně přistupovat k sociální síti z různých zařízení. Informace na jednom zařízení tedy nemusí být kompletní.
- Jedno zařízení může využívat více uživatelů a to i v rámci jednoho uživatelského účtu.
- Moderní prohlížeče disponují funkcemi pro dočasné nebo trvalé vypnutí ukládání historie prohlížení a dalších výše uvedených dat (tzv. *privátní režim*), další volitelná rozšíření prohlížečů umožňují selektivně zabránit ukládání určitých navštívených URL nebo modifikovat historii podle jiných kritérií.

V následujících kapitolách rozebereme typy informací, které jsou typicky k dispozici na lokálních počítačích. Vzhledem k tomu, že problematika mobilních zařízení je v rámci projektu Tarzan zkoumána samostatně [16], zaměříme se pouze na běžné počítače (stolní počítače, notebooky apod.)

5.1 Informace v lokálních profilech

Každý webový prohlížeč pří svém používání vytváří tzv. *uživatelský profil*. Jedná se o sadu konfiguračních a databázových souborů, které jsou obvykle uloženy v samostatném adresáři na pevném disku počítače a obsahují personalizovanou konfiguraci prohlížeče pro konkrétního uživatele počítače. Pro každého uživatele existuje typicky jeden profil, který je vytvořen při prvním spuštění prohlížeče. Pokročilí uživatelé mohou využívat více profilů, mezi kterými se při spuštění prohlížeče přepínají.

Konkrétní umístění a struktura uživatelských profilů se u jednotlivých prohlížečů do značné míry liší. U všech běžných prohlížečů však v profil obsahuje různé informace o minulé aktivitě uživatele na webu, které jsou primárně využívány pro usnadnění obsluhy prohlížeče (např. nabízení již navštívených URL) nebo pro zrychlení práce prohlížeče (dočasné soubory apod.) Z hlediska analýzy interakcí se sociálními sítěmi jsou zajímavé zejména následující informace.

Historie prohlížení Historie prohlížení obsahuje evidenci všech URL navštívených uživatelem často za celou dobu existence profilu (od prvního spuštění prohlížeče). Z této historie lze zjistit mimo jiné následující zajímavé údaje:

- Navštívená místa (URL): počet návštěv, datum a čas první návštěvy a poslední návštěvy.
- Stažené soubory: URL zdroje, velikost souboru, čas stažení a cílové místo, kam byl soubor uložen na lokálním počítači.

Podle URL lze tedy vybrat záznamy, které souvisejí s konkrétní sociální sítí.

Cookies *Cookie* je krátký datový záznam (s velikostí typicky do 4 kB) uložený v prohlížeči. V protokolu HTTP představují Cookies mechanismus, který umožňuje serveru uložit na straně klienta různé servisní informace, které má server následně k dispozici při dalších požadavcích přicházejících od stejného uživatele. Každý uložený cookie je vázaný na URL svého zdroje a má nastavitelnou dobu platnosti, po jejímž uplynutí je vymazán [8].

Kromě ukládání různých specifických konfiguračních informací jsou cookies využívány zejména pro rozpoznání dříve přihlášených uživatelů. Při přihlášení do sociální sítě (např. zadáním uživatelského jména a hesla) je uživateli vygenerován jedinečný identifikační řetězec (tzv. *identifikátor sezení*), který je následně uložen pomocí cookie v prohlížeči uživatele. Při následujících požadavcích přicházejících od stejného uživatele pak server rozpozná, o kterého uživatele se jedná. Platnost identifikátoru končí odhlášením uživatele nebo vypršením platnosti cookie.

Z hlediska analýzy práce ze sociální sítí lze z uložených cookies získat informaci o tom, zda a případně kdy byl uživatel k sociální síti přihlášen. V případě, že se uživatel důsledně neodhlašuje při skončení práce se sociální sítí, může být platný cookie s identifikátorem sezení využit pro získání přístupu k účtu uživatele na sociální síti [12].

Cache Jedná se o dočasné úložiště částí webových stránek (např. obrázků, skriptů, stylových předpisů, apod.), jehož primárním účelem je zrychlení reakce prohlížeče a úspora přenosové kapacity při prohlížení webu. Prohlížeč si zde ukládá opakovaně používaná data, což umožňuje využít lokální kopii namísto opakovaného stahování stejného souboru z webových serverů.

Na rozdíl od výše zmíněné historie prohlížení, která může být velmi dlouhá, trvanlivost souborů v dočasném úložišti je u moderních prohlížečů poměrně krátká (řádově jednotky dnů). Z přítomnosti některých souborů v úložišti lze usuzovat na návštěvu konkrétních webových stránek, je však poměrně obtížné spojit tuto informaci s nějakou konkrétní aktivitou uživatele na sociální síti. Vzhledem k tomu, že strategii ukládání dočasných souborů určují vnitřní mechanismy prohlížeče, nelze existenci konkrétního dokumentu (např. fotografie, videa, apod.) v úložišti přímo interpretovat tak, že uživatel s tímto dokumentem pracoval.

5.2 Operační systémy a webové prohlížeče

Na desktopových počítačích jsou v současnosti používány téměř výhradně operační systémy Windows (90,77 % uživatelů), Mac OS X (6,25 %) a Linux (2,98 %) [4]. V závislosti na operačním systému se liší umístění a způsob uložení uživatelských profilů. Nicméně všechny hlavní webové prohlížeče lze provozovat na všech zmíněných operačních systémech a uložené uživatelské profily obsahují vždy stejnou množinu informací.

Pokud jde o prohlížeče, za aktuálně nejvíce používané lze považovat Chrome (59,84 % uživatelů), Internet Explorer (15,09 %), Firefox (13,14 %) a Microsoft Edge (4,58 %) [3]. Dále proto rozebíráme, jakým způsobem jsou dostupné informace z lokálních profilů popsané v kapitole 5.1 u jednotlivých prohlížečů na různých operačních systémech.

5.3 Chrome

Umístění uživatelského profilu na různých operačních systémech ve výchozím nastavení prohlížeče je uvedeno v tabulce 3.

Operační systém	**Umístění profilu**
Windows	`C:\Users\<username>\AppData\Local\Google\Chrome\User Data\Default`
Mac OS X	`Users/<username>/Library/Application Support/Google/Chrome/Default`
Linux	`/home/<username>/.config/google-chrome/Default`

Tabulka 3. Standardní umístění uživatelských profilů u prohlížeče Chrome.

Uživatelský profil obsahuje množství souborů a adresářů. Pro získání výše uvedených informací jsou zajímavé zejména soubory *History* a *Cookies*. Jedná se o relační databázové soubory ve formátu SQLite[5], které lze s využitím vhodných knihoven poměrně jednoduše načíst a analyzovat na různých platformách.

5.4 Firefox

Podobně jako u prohlížeče Chrome, i u Firefoxu se umístění profilu liší v závislosti na operačním systému, jak je patrné z tabulky 4 níže.

Struktura uživatelského profilu je do značné míry obdobná, jako u prohlížeče Chrome. Relevantní jsou zejména soubory *places.sqlite* a *cookies.sqlite*, které opět využívají databázový formát SQLite. Jejich struktura, jako např. jména tabulek, sloupců a celkové relační schéma jsou v porovnání s Chrome poněkud odlišné, rozsah dostupných informací je však stejný.

[5] https://www.sqlite.org/

Operační systém	Umístění profilu
Windows	C:\Users*<username>*\AppData\Roaming\Mozilla\Firefox\Profiles*xxx*.default
Mac OS X	Users/*<username>*/Library/Application Support/Firefox/Profiles/*xxx*.default
Linux	/home/*<username>*/.mozilla/firefox/*xxx*.default

Tabulka 4. Standardní umístění uživatelských profilů u prohlížeče Firefox (*xxx* je náhodně vygenerované jméno profilu).

5.5 Microsoft Internet Explorer

Na rozdíl od prohlížečů Chrome a Firefox jsou prohlížeče Microsoftu k dispozici pouze pro platformu Windows. U prohlížeče Internet Explorer nejsou data o historii používání umístěna v oddělených profilech, ale jsou součástí uživatelských dat systému Windows.

Historie prohlížení je typicky umístěna ve složce
C:\Users*<username>*\AppData\Local\Microsoft\Windows\History.
Starší verze Internet Exploreru (do verze 9 včetně) používají proprietární formát souboru *index.dat*. Novější verze pak používají databázové soubory *Extensible Storage Engine (ESE)* (soubory EDB).

Obdobně cookies jsou uloženy obvykle v adresři
C:\Users*<username>*\AppData\Roaming\Microsoft\Windows\Cookies.
Tento adresář obsahuje prosté textové soubory s informacemi o cookies z jednotlivých navštívených domén.

5.6 Microsoft Edge

Prohlížeč Edge používá obdobně jako nové verze Internet Exploreru soubory Extensible Storage Engine. Veškerá data prohlížeče zahrnující historii prohlížení i cookies jsou nyní uložena v souboru
C:\Users*<username>*\AppData\Local\Packages
\Microsoft.MicrosoftEdge_xxxxx\AC\MicrosoftEdge\User\Default
\DataStore\Data\nouser1\xxxxx\DBStore\spartan.edb.

Struktura tohoto souboru je velmi odlišná od ostatních prohlížečů, jedná se však o strukturovaný soubor s relačním modelem dat, pro jehož zpracování jsou k dispozici obecně použitelné knihovny a nástroje [17].

6 Získávání dat ze sociálních sítí

Jednotlivé sociální sítě umožňují publikovat a sdílet různé typy informací různými způsoby. Existují však jisté zažité zvyklosti, které platí napříč sociálními sítěmi a které umožňují publikované informace obecněji popsat.

– Každý uživatel publikuje své příspěvky ve své vymezené oblasti, kterou můžeme v závislosti na konkrétní sociální síti nazvat profil (*profile*), stránka (*page*), případně časová řada (*timeline*).
– V rámci této oblasti jsou příspěvky publikovány chronologicky, čas publikace je jedním ze základních atributů příspěvku.

Jednotlivé příspěvky potom mohou obsahovat následující nejběžnější typy informací:

– Text příspěvku. U některých sociálních sítí je délkově omezen (např. u sítě Twitter), u jiných naopak může být i velmi dlouhý a strukturovaný (Facebook, LinkedIn).
– Fotografie, případně video. Některé sociální sítě prezentují fotografie jako primární obsah příspěvků (např. Instagram), u jiných mají spíše doplňkovou úlohu (Twitter).
– Odkazy na web (URL). Sdílení obsahu publikovaného jinde na webu je velmi časté téměř u všech sociálních sítí.
– Geografické informace (místa). Geografické souřadnice místa mohou být přiřazeny příspěvku případně jen některým jeho částem (typicky fotografii) a mohou se vztahovat k místu, kde byl obsah pořízen, nebo o kterém pojednává.

Jednotlivé publikované příspěvky mohou obsahovat jednu z výše uvedených informací nebo jejich prakticky libovolnou kombinaci.

Všechny sociální sítě umožňují čtenářům reagovat na zveřejněné příspěvky, ať již ve formě komentářů, nebo jen pomocí stručného ohodnocení (Líbí se mi apod.) Rozšířená je také možnost sdílet ve svém profilu příspěvek, který publikoval někdo jiný.

Kromě vlastních publikovaných příspěvků jsou ke každému profilu přiřazeny informace o jeho vlastníkovi, jejichž podrobnost je rovněž závislá na konkrétní sociální síti a sdílnosti konkrétního uživatele. Tyto informace mohou zahrnovat osobní údaje (jméno, datum narození, kontakt) nebo i profesní životopis uživatele a jeho zájmy.

Pokud chceme analyzovat příspěvky a profily uživatelů napříč sociálními sítěmi, je třeba se zabývat tím, jak lze programově získat obsah publikovaných příspěvků a další informace o každém uživateli. Některé sociální sítě k tomuto účelu poskytují dokumentované aplikační rozhraní, které umožňuje aplikacím třetích stran využívat zveřejněné informace za daných podmínek. Rovněž některé informace mohou být přístupné pouze omezenému okruhu uživatelů a je tedy se zabývat použitým systémem přístupových práv. Dále popisujeme tyto detaily pro jednotlivé sociální sítě.

6.1 Facebook

Facebook umožňuje publikovat příspěvky v rámci *profilu*, *skupiny* nebo *stránky*. Profil (*profile*) je dostupný pouze pro fyzické osoby a umožňuje sdílet publikovaný

obsah vybraným skupinám uživatelů (viditelnost jednotlivým uživatelům může být nastavena pro každý příspěvek). Podobně skupina (*group*) je místo, kam může přidávat příspěvky více fyzických osob, obvykle bývá tématicky zaměřená. Konečně stránka (*page*) slouží pro publikaci příspěvků o libovolném subjektu (osobě, firmě, produktu apod.) a je určena pro veřejné sdílení příspěvků.

Facebook poskytuje dokumentované aplikační rozhraní (API)[6], které umožňuje aplikacím přistupovat k obsahu profilů a také publikovat nový obsah. Oproti přístupu přes standardní webové rozhraní je přístup přes API do značné míry omezen:

- Aplikace, která přistupuje k profilu nějakého uživatele musí být schválena uživatelem profilu. Součástí schvalování jsou i detailní přístupová práva, která aplikace požaduje (např. přístup k osobním informacím, ke kontaktům, k obrázkům apod.) Po schválení získává aplikace unikátní přístupový řetězec (token), který pak využívá pro autorizaci každého jednotlivého požadavku.
- Na rozdíl od přístupu přes webové rozhraní, které umožňuje procházet veřejné příspěvky všech uživatelů, aplikační rozhraní umožňuje přístup pouze k profilu uživatele, který aplikaci schválil.

Aplikační rozhraní je tedy převážně určeno pro práci s profilem nebo profily uživatelů, kteří si instalovali danou aplikaci a rozsah dostupných informací je výrazně menší, než při použití standardního webového rozhraní. Nelze například přistupovat k údajům o spřátelených profilech a podobně. Při statistickém zpracování dat publikovaných na Facebooku se proto někteří výzkumníci uchylují k tvorbě specializovaných skriptů, které získávají data prostřednictvím webového prohlížeče [5].

Přístup k obsahu stránek prostřednictvím API na druhou stranu takovým omezením nepodléhá vzhledem k veřejné povaze stránek.

6.2 Google+

Na síti Google+ mohou být příspěvky publikovány na osobních profilech jednotlivých uživatelů, nebo na *stránkách značky*, které může spravovat více uživatelů a které reprezentují veřejný profil libovolného subjektu (např. firmy nebo organizace). V obou typech profilů je možno publikovat buď veřejné příspěvky viditelné všem uživatelům, nebo příspěvky sdílené jen s omezených okruhem uživatelů, který je možno detailně určit pro každý příspěvek.

Poskytované aplikační rozhraní sítě Google+[7] umožňuje plný přístup k profilu vlastníka; vlastník musí každou aplikaci explicitně autorizovat podobně jako v případě sítě Facebook. Dále aplikační rozhraní zpřístupňuje veřejné příspěvky všech uživatelů a veřejně dostupné informace o jednotlivých profilech.

[6] https://developers.facebook.com/docs/graph-api/
[7] https://developers.google.com/+/web/api/rest/latest/

6.3 Twitter

Síť Twitter je primárně určena pro veřejné publikování příspěvků na profilech, které mohou patřit jak fyzickým osobám, tak i dalším subjektům. Ve výchozím nastavení jsou tedy všechny publikované příspěvky viditelné všem uživatelům, včetně těch, kteří nemají účet na Twitteru [1]. Uživatel má možnost viditelnost příspěvku explicitně omezit.

Twitter poskytuje velmi dobře dokumentované aplikační rozhraní[8], které umožňuje efektivní přístup k obsahu všech viditelných příspěvků. Tyto vlastnosti z něj činí populární zdroj dat pro různé analýzy (např. [9]).

6.4 Instagram

Síť Instagram je primárně zaměřena na sdílení fotografií a videa. Příspěvky jsou standardně publikovány jako veřejné, existuje však možnost viditelnost příspěvků omezit.

Podobně jako ostatní zmiňované sociální sítě nabízí i Instagram aplikační rozhraní[9]. Toto rozhraní zpřístupňuje veřejná data o uživatelích a veřejně dostupné příspěvky. Přihlášenému uživateli umožňuje plný přístup k jeho profilu včetně možnosti přidávání příspěvků.

Na rozdíl od ostatních sociálních sítí je na Instagramu vyžadováno schválení aplikace provozovatelem. Každá aplikace je vyvíjena v tzv. *sandbox* (vývojovém) režimu a přidělení jednotlivých přístupových práv (např. přístup k příspěvkům, k uživatelským datům, apod.) je podmíněno schválením provozovatele Instagramu. Hotová aplikace musí být následně znovu posouzena provozovatelem, aby přešla z vývojového do standardního režimu a mohla být používána všemi uživateli.

6.5 LinkedIn

Sociální síť je oproti ostatním výše uvedeným sítím orientovaná spíše na pracovní kontakty a sdílení příspěvků jako takové hraje spíše podružnou roli.

Aplikační rozhraní sítě LinkeIn je aktuálně poměrně omezené[10]. Uživatel musí každou aplikaci schválit včetně jednotlivých přístupových práv. Aplikace potom může přistupovat k profilu uživatele, který ji schválil, číst základní data o uživateli a přidávat sdílené příspěvky. API není možno využít k přístupu k profilům ostatních uživatelů, a to ani k veřejným informacím.

7 Nástroj pro analýzu dat ze sociálních sítí

Jak vyplývá z předchozí analýzy informačních zdrojů v kapitolách 5 a 6, ať již se jedná o relevantní záznamy o používání prohlížečů na lokálním počítači, nebo

[8] https://developer.twitter.com/en/docs
[9] https://www.instagram.com/developer/
[10] https://developer.linkedin.com/docs/rest-api

o data o příspěvcích ze sociálních sítí, ve všech případech je nutno analyzovat potenciálně velmi rozsáhlou množinu informací. Data o používání prohlížečů mohou obsahovat tisíce položek i za poměrně krátké časové období několika dnů. Obdobně u sociálních sítí je často zapotřebí projít velké množství potenciálně relevantních profilů napříč různými sociálními sítěmi, které mohou obsahovat stovky příspěvků. Získané informace pak mohou mít vazby na další informační zdroje zkoumané dosud odděleně, jako např. monitorování síťového provozu [19] nebo analýza mobilních zařízení [16].

Aby bylo možno toto množství dat efektivně zobrazit a analyzovat, navrhli jsme rozšiřitelný nástroj pro analýzu sociálních sítí, který bere v úvahu zejména následující požadavky:

- Možnost jednotným způsobem získávat data z heterogenních zdrojů (ať již sociálních sítí, nebo dalších zdrojů).
- Jednotná reprezentace získaných dat nezávisle na jejich zdroji a jejich uložení pro další analýzu.
- Dostupnost nástrojů pro automatizované vyhledání různých vztahů mezi zkoumanými datovými zdroji – např. sdílení stejné informace nebo fotografie, vzájemné odkazování, tématické zaměření příspěvků a podobně.
- Přehledné zobrazení shromážděných dat s vizualizací detekovaných vztahů mezi jednotlivými datovými zdroji, profily nebo přímo příspěvky v těchto profilech.
- Snadná rozšiřitelnost o další datové zdroje a analytické nástroje podle momentální potřeby.

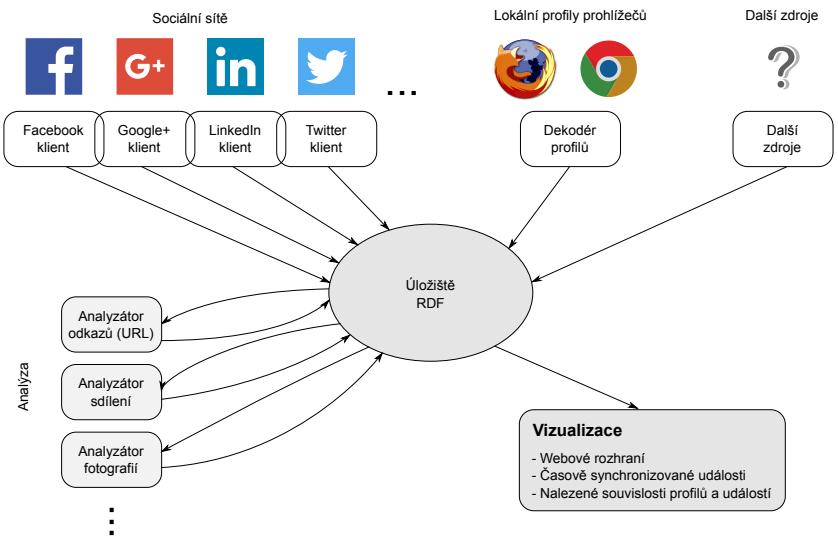
Cílem nástroje tedy je dát uživateli možnost shromáždit všechna dostupná data z různých zdrojů v jednotném úložišti, spouštět automatizované analytické úlohy nad těmito daty s cílem nalézt vztahy mezi různými datovými zdroji a přehledně graficky zobrazit detekované vztahy pro podrobnější zkoumání.

Na základě uvedených požadavků jsme navrhli architekturu nástroje s pracovním názvem *TimelineAnalyzer*, zvolili jsme vhodné technologie pro realizaci jednotlivých jeho částí a implementovali prototypové řešení pro ověření navrženého konceptu.

7.1 Architektura nástroje

Navržená architektura nástroje *TimelineAnalyzer* je na obrázku 1. Jádrem nástroje je jednotné úložiště, které jednotným způsobem ukládá informace z různých zdrojů. Byl zvolen datový model založený na technologii RDF, jak popisujeme v sekci 7.2.

Další komponentu tvoří zdroje informací reprezentované moduly, které implementují získávání informací z jednotlivých sociálních sítí, lokálních profilu webových prohlížečů a případně i dalších zdrojů. Jelikož různé sociální sítě používají různá aplikační rozhraní, předpokládá se implementace samostatného modulu pro každý informační zdroj. Architektura nástroje je přitom navržena s cílem umožnit jednoduché rozšiřování množiny zpracovávaných zdrojů dat.



Obrázek 1. Architektura systému *TimelineAnalyzer* pro analýzu dat ze sociálních sítí

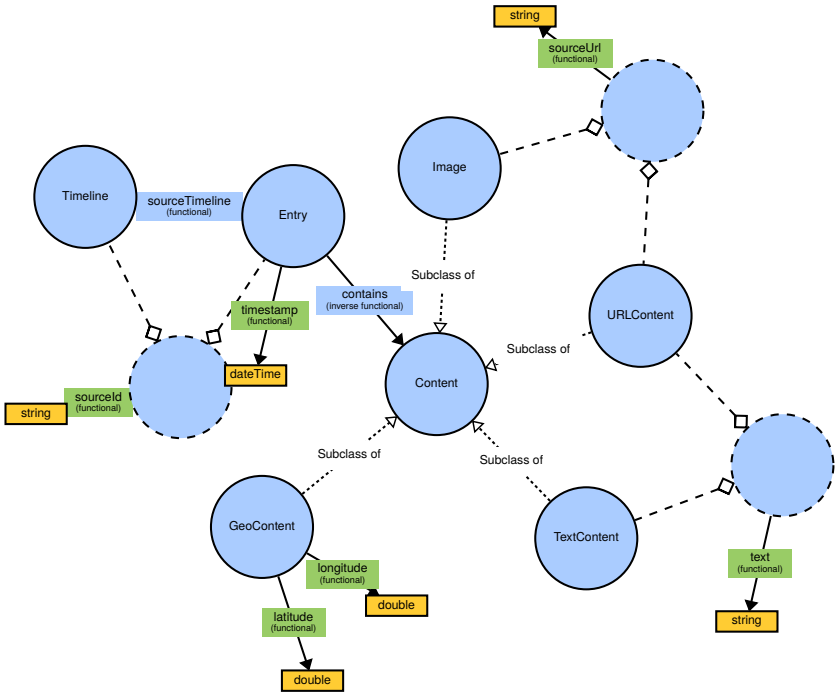
Automatizovanou analýzu získaných dat pak reprezentují analytické moduly, které čerpají data z úložiště provádí nad ním konkrétní analýzu a získané údaje ukládají ve formě dodatečných informací zpět do úložiště. Konečně pro vizualizaci a manuální analýzu získaných dat je k dispozici grafické uživatelské rozhraní běžící ve webovém prohlížeči.

7.2 Datový model

Použitý datový model vychází z analýzy informací dostupných na sociálních sítích, jak je diskutováno v kapitole 6. Obecná reprezentace zvoleného modelu ve formě ontologie je patrná na obrázku 2. Ontologie popisuje jednotlivé identifikované entity (objekty) v cílové doméně a jejich vzájemné vztahy.

Každý datový zdroj, který odpovídá jednomu profilu na konkrétní sociální síti nebo jednomu lokálnímu profilu prohlížeče je reprezentován časovou osou – entitou *Timeline*. Ta sdružuje jednotlivé příspěvky (*Entry*), které jsou chronologicky řazeny. Každý příspěvek pak obsahuje různé druhy obsahu (*Content*), konkrétně text (*TextContent*), obrázky (*Image*), odkazy URL (*URLContent*) nebo geografické údaje (*GeoContent*).

Takto navržený obecný model umožňuje reprezentovat obsah konkrétního profilu na libovolné sociální síti. Současně je však použitelný na reprezentaci dalších zdrojů dat, která jsou chronologicky řazena a lze v nich nalézt některé ze zmíněných druhů obsahu. Například historii používání prohlížeče získanou z lo-



Obrázek 2. Datový model reprezentovaných znalostí (ontologie)

kálního profilu (viz kapitolu 5.1) je možno reprezentovat jako samostatnou časovou osu obsahující chronologický řazené položky odpovídající návštěvám stránek, staženým souborům a dalším akcím doplněným o příslušné URL.

Získaná data popsaná pomocí uvedené ontologie lze jednoduše reprezentovat pomocí obecného datového modelu RDF (*Resource Description Framework* [13]). Existuje celá řada databázových produktů, které umožňují takto popsaná data uložit a dále s nimi pracovat například pomocí dotazování. Konkrétní implementaci popisujeme dále v kapitole 7.6. To umožňuje vybudovat společné úložiště dat, kam se soustřeďují informace získané ze všech datových zdrojů a nad kterým je potom možno provádět detailní analýzu získaných dat.

7.3 Zdroje dat

Navržený nástroj počítá s rozšiřitelnou množinou zdrojů dat. Každý zdroj dat představuje softwarový modul, který implementuje získání dat z jednoho konkrétního zdroje, tedy konkrétní sociální sítě, lokálního profilu konkrétního prohlížeče a dalších. Získaná data jsou pak reprezentována pomocí výše popsaného datového modelu a uložena do společného úložiště pro další analýzu. Jednotlivé datové zdroje tedy představují rozhraní mezi sociální sítí a zbývajícími částmi systému, které pak již nemusí zohledňovat specifické vlastnosti jednotlivých sociálních sítí.

Základní uvažované zdroje dat jsou patrné z obrázku 1. Předpokládáme implementaci datových zdrojů pro hlavní sociální sítě používané v ČR a pro lokální profily hlavních prohlížečů. Nabízí se však i možnost integrovat informace získané monitorováním síťového provozu [19] nebo analýzou mobilních zařízení [16].

7.4 Analytické nástroje

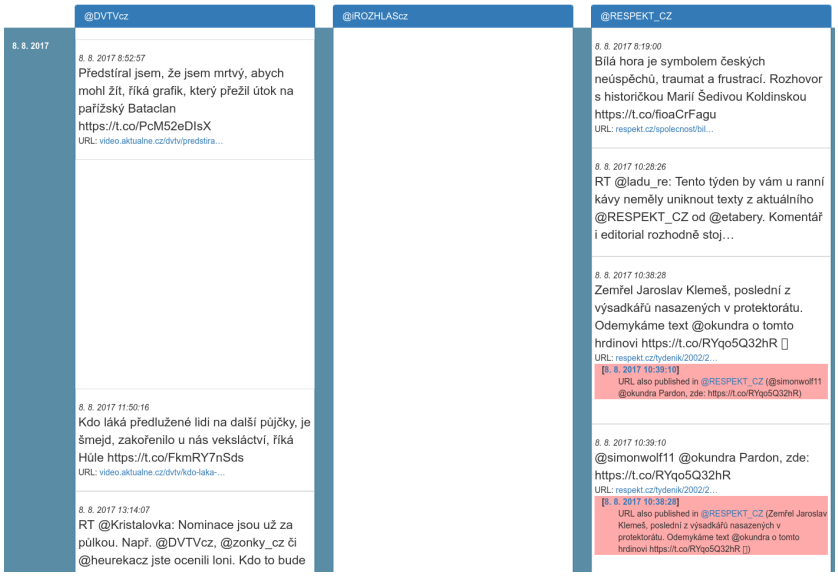
Analytické moduly naznačené na obr. 1 pracují nad společným úložištěm dat. Jejich úkolem je nalézt potenciálně zajímavé vztahy mezi jednotlivými uloženými časovými řadami. Každý analytický modul vyhledává jeden určitý typ vztahu a informace o nalezených vztazích ukládá zpět do společného úložiště. Mezi potenciálně zajímavé vztahy vyhledatelné v datech patří:

- Shody v publikovaném obsahu – např. uživatelé sdíleli stejnou informaci, obrázek, odkaz apod.
- Shody v URL – např. lokální uživatel navštívil stránku odkazovanou v některém profilu, stáhnul nebo nahrál zveřejněnou fotografii nebo jiný obsah.
- Shody v čase nebo místě využívání profilů – např. na více profilech byly publikovány odkazy na stejné místo, publikace probíhá často ve stejnou dobu apod.
- Vzájemné interakce mezi profily – explicitní odkazování, opakované sdílení příspěvků apod.
- Tématická podobnost příspěvků – profily publikují příspěvky zaměřené na podobná témata.

Dané informace jsou využitelné např. pro přiřazení profilů nebo konkrétních publikovaných informací ke konkrétním osobám, identifikaci profilů obsluhovaných stejnými osobami, identifikaci "spřátelených" profilů apod. Cílem analytických modulů je nalézt a zaznamenat algoritmicky zjistitelné souvislosti. Dalším předpokládaným krokem je pak ruční vyhodnocení těchto souvislostí odborníkem, který je dokáže interpretovat. K tomu lze použít grafickou reprezentaci nalezených vztahů prostřednictvím grafického uživatelského rozhraní.

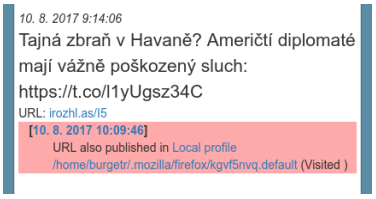
7.5 Grafické uživatelské rozhraní

Grafické uživatelské rozhraní zobrazuje jednotlivé časové osy a vztahy mezi nimi. Jeden z možných pohledů je na obrázku 3. Zvýrazněný vztah je viditelný na obrázku 4.



Obrázek 3. Grafická prezentace časových os

Nad centrálním úložištěm je možno generovat i více grafických reprezentací. Do budoucna proto předpokládáme možnost alternativních zobrazení (např. ve formě grafu), která by umožňovala abstrahovat od konkrétního obsahu příspěvků a soustředit se na topologii vztahů mezi profily, jejich četností apod.



Obrázek 4. Detekovaný vztah (publikovaný odkaz navštívený v lokálním profilu prohlížeče)

7.6 Implementace a integrace s platformou Tarzan

V současné době je implementován prototyp nástroje TimelineAnalyzer[11] implementovaný na platformě Java. Využívá centrální úložiště dat RDF postavené na otevřené databázi Rdf4j[12]. Implementace zahrnuje datové zdroje pro sociální síť Twitter, lokální prohlížeč Firefox a implementaci dvou analytických metod založených na porovnání použitých URL odkazů a obrázků. Součástí je i grafické uživatelské rozhraní běžící ve webovém prohlížeči implementované v jazyce JavaScript (viz ukázku na obr. 3).

Prototypová implementace ukazuje funkčnost zvoleného konceptu. V dalších měsících předpokládáme další rozšiřování množiny datových zdrojů, analytických metod i doplnění alternativních grafických zobrazení získaných informací.

Při návrhu nástroje byla rovněž uvažována možnost integrace s dalšími částmi vyvíjené Integrované platformy pro zpracování digitálních dat z bezpečnostních incidentů [22] a souběžně zkoumaných oblastí analýzy síťového provozu [19] a mobilních zařízení [16]. Tomu odpovídá i volba výše uvedených technologií, které je možno nasadit i v distribuovaném prostředí využívajícím technologie Apache Hadoop a Spark využívané v integrované platformě.

8 Závěr

Vzhledem k množství a různorodosti dostupných sociálních sítí a souvisejících zdrojů dat a vzhledem k množství informací dostupných v těchto zdrojů představuje zpracování informací v souvislosti s bezpečnostními incidenty poměrně náročný technický problém. V tomto dokumentu jsme provedli analýzu relevantních bezpečnostní incidentů a dostupných zdrojů informací, se zaměřením na nejpoužívanější sociální sítě a data o využívání webových prohlížečů na lokálních počítačích.

Následně jsme představili návrh nástroje, který umožňuje efektivně získat a reprezentovat dostupná data ze sociálních sítí, provádět nad nimi různé druhy analýz a interaktivně procházet získané informace. Při návrhu byl kladen důraz

[11] Dostupný na https://github.com/nesfit/timeline-analyzer
[12] http://rdf4j.org/

na rozšiřitelnost a škálovatelnost řešení. Hlavní koncepty návrhu byly ověřeny na prototypové implementaci.

V dalších fázích výzkumu předpokládáme rozšíření množiny implementovaných datových zdrojů, integraci s dalšími částmi platformy Tarzan a testování prototypu na reálných scénářích použití v souvislosti s bezpečnostními incidenty.

Reference

1. About public and protected tweets. Twitter Help Center, [Online; navštíveno 28.11.2017].
   URL https://help.twitter.com/en/safety-and-security/public-and-protected-tweets
2. V Česku má nejvíce uživatelů YouTube a Facebook. Marketingové noviny.cz, Březen 2016, [Online; navštíveno 30.11.2017].
   URL http://www.marketingovenoviny.cz/v-cesku-ma-nejvice-uzivatelu-youtube-a-facebook/
3. Desktop Browser Market Share. NetMarket Share, Říjen 2017, [Online; navštíveno 28.11.2017].
   URL http://www.netmarketshare.com/
4. Desktop Operating System Market Share. NetMarket Share, Říjen 2017, [Online; navštíveno 28.11.2017].
   URL http://www.netmarketshare.com/
5. Slušní lidé na Facebooku. Data Boutique, Červen 2017, [Online; navštíveno 28.11.2017].
   URL http://databoutique.cz/post/161312186128/slušní-lidé-na-facebooku
6. d. Andrade, N. N. G.; Martin, A.; Monteleone, S.: "All the better to see you with, my dear": Facial recognition and privacy in online social networks. *IEEE Security Privacy*, ročník 11, č. 3, May 2013: s. 21–28.
7. Faghani, M. R.; Saidi, H.: Malware propagation in Online Social Networks. In *2009 4th International Conference on Malicious and Unwanted Software (MALWARE)*, Oct 2009, s. 8–14.
8. Fielding, R.; Gettys, J.; Mogul, J.; aj.: RFC 2616, Hypertext Transfer Protocol – HTTP/1.1. 1999.
   URL http://www.rfc.net/rfc2616.html
9. Gannon, W.: Build a Sentiment Analysis Tool for Twitter with this Simple Python Script. Aylien Ltd., Říjen 2017, [Online; navštíveno 6.12.2017].
   URL http://blog.aylien.com/build-a-sentiment-analysis-tool-for-twitter-with-this-simple-python-script/
10. Hasib, A. A.: Threats of online social networks. *International Journal of Computer Science and Network Security*, 2009: s. 288–293.
11. Howden, C.; Liu, L.; Ding, Z.; aj.: Moments in Time: A Forensic View of Twitter. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, Aug 2013, s. 899–908.

12. Huber, M.; Schmiedecker, M.; Leithner, M.; aj.: Social Snapshots: Digital Forensics for Online Social Networks. In *Annual Computer Security Applications Conference (ACSAC)*, 12 2011.
13. Lanthaler, M.; Wood, D.; Cyganiak, R.: RDF 1.1 Concepts and Abstract Syntax. W3C recommendation, W3C, Únor 2014, http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/.
14. Lorenc, J.: Jak se daří jednotlivým sociálním sítí v České republice? AMI Digital s.r.o., Duben 2017, [Online; navštíveno 23.11.2017].
URL https://www.linkedin.com/pulse/jak-se-daří-jednotlivým-sociálním-sítí-v-české-republice-jakub-lorenc
15. Lorenc, J.: Jak se změnily počty uživatelů jednotlivých sociálních sítí za poslední půlrok? AMI Digital s.r.o., Říjen 2017, [Online; navštíveno 23.11.2017].
URL https://www.linkedin.com/pulse/jak-se-změnily-počty-uživatelů-jednotlivých-sítí-za-poslední-lorenc
16. Matoušek, P.; Havlík, J.: Detekce mobilních zařízení v síťové komunikace. Technická Zpráva FIT-TR-2017-08, FIT VUT v Brně, 2017.
17. Muir, B.: Windows 10 – Microsoft Edge Browser Forensics. Září 2015, [Online; navštíveno 28.11.2017].
URL http://bsmuir.kinja.com/windows-10-microsoft-edge-browser-forensics-1733533818
18. Mulazzani, M.; Huber, M.; Weippl, E. R.: Social Network Forensics: Tapping the Data Pool of Social Networks. In *Eighth Annual IFIP WG 11.9 Internaional Conference on Digital Forensics*, 2011.
19. Pluskal, J.: Netfox Detective 2.0 – Nástroj pro síťovou forenzní analýzu. Technická Zpráva FIT-TR-2017-06, FIT VUT v Brně, 2017.
20. Polakis, I.; Ilia, P.; Tzermias, Z.; aj.: Social Forensics: Searching for Needles in Digital Haystacks. In *4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, Nov 2015.
21. Rathore, S.; Sharma, P. K.; Loia, V.; aj.: Social network security: Issues, challenges, threats, and solutions. *Information Sciences*, ročník 421, č. Supplement C, 2017: s. 43 – 69.
22. Ryšavý, O.; Rychlý, M.: Platforma pro zpracování dat síťové forenzní analýzy: Návrh a implementace prototypu. Technická Zpráva FIT-TR-2017-07, FIT VUT v Brně, 2017.
23. Shaw, U.; Das, D.; Medhi, S. P.: Social Network Forensics: Survey and Challenges. *International Journal of Computer Science and Information Security*, ročník 14, č. 11, 11 2016: s. 310–316.
24. Viejo, A.; Castellà-Roca, J.; Rufián, G.: *Preserving the User's Privacy in Social Networking Sites*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, s. 62–73.