Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačního inženýrství

Diplomová práce

Ochrana soukromí v digitálním prostředí

Jan SOBĚSLAV

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Provozně ekonomická fakulta

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Jan Soběslav

Informatika

Název práce

Ochrana soukromí v digitálním prostředí

Název anglicky

Protection of privacy in digital environment

Cíle práce

Primárním cílem práce je navržení a provedení experimentu, který má prokázat využívání dat ze soukromých elektronických konverzací pro potřeby cílení internetové reklamy. Pokud se podezření prokáže, autor navrhne opatření pro ochranu jednoho vybraného komunikačního kanálu.

Vedlejším cílem práce je podání uceleného pohledu na problematiku digitálních stop a úniků osobních informací, včetně navržených protiopatření.

Metodika

Teoretická část se opírá o rešerši literárních a internetových zdrojů.

Experiment spočívá v monitorování množství a obsahu zobrazených reklam v kontrolním období a poté, co jsou do komunikačních kanálů podstrčeny informace, které mají za cíl zvýšit počet reklam relevantních k obsahu konverzací. Výsledky experimentu budou vyhodnoceny statistickou analýzou.

Práce sestává ze dvou částí – teoretické rešerše a praktického experimentu.

Metodika zpracování teoretické části vychází ze studia odborných literárních a elektronických zdrojů. Na základě syntézy zjištěných poznatků budou popsána teoretická východiska práce.

Praktická část práce spočívá v návrhu, provedení a vyhodnocení experimentu spočívajícího v monitorování množství a obsahu zobrazených reklam v kontrolním období a poté, co jsou do komunikačních kanálů podstrčeny informace, které mají za cíl zvýšit počet reklam relevantních k obsahu konverzací.

Bude ověřována hypotéza, že poté, co je komunikace ovlivněna, bude objektivně zaznamenáno více relevantních reklam. Výsledky experimentu budou vyhodnoceny statistickou analýzou a interpretovány.

Doporučený rozsah práce

60-80 stran

Klíčová slova

soukromí, reklama, sledování, osobní informace, zneužití dat, datová stopa, data

---

Doporučené zdroje informací

B. Krebs, Spam nation. ISBN: 978-1402295614. Sourcebooks, 2014.
B. Schneier, Data and Goliath. ISBN: 978-0393244816. W. W. Norton & Company, 2015.
F. Ahearn, E. Horan, How to Disappear. ISBN: 978-1599219776. Globe Pequot Press, 2010.
G. Day, Security in the Digital World. ISBN: 978-1849289610. IT Governance Publishing, 2017.
J. Angwin, Dragnet Nation. ISBN: 978-0805098075. Times Books, 2014.
M. Bazzell, Hiding from the Internet. ISBN: 978-1522914907. CreateSpace Independent Publishing Platform, 2016.
M. Bazzell, Personal Digital Security. ISBN: 978-1491081976. CreateSpace Independent Publishing Platform, 2013.

Předběžný termín obhajoby

2018/19 LS – PEF

Vedoucí práce

Ing. Jiří Brožek, Ph.D.

Garantující pracoviště

Katedra informačního inženýrství

Elektronicky schváleno dne 26. 3. 2019

Ing. Martin Pelikán, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 26. 3. 2019

Ing. Martin Pelikán, Ph.D.

Děkan

V Praze dne 27. 03. 2019

Čestné prohlášení

Prohlašuji, že svou diplomovou práci "Ochrana soukromí v digitálním prostředí" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 29.3.2019

Poděkování

Děkuji Ing. Jiřímu Brožkovi, Ph.D. za vedení práce a poskytnuté konzultace a podněty.

Také děkuji Ing. Pavle Hoškové, Ph.D. za konzultaci statistických metod.

Dále děkuji Bc. Veronice Čechové za účast v experimentu a podporu.

Samozřejmě je také na místě poděkovat mým nejbližším, jejichž obecná podpora měla na vzniku práce nemalý podíl.

Zvláštní poděkování patří též opět Pánu Ivo Krátkému od Kostečků.

Ochrana soukromí v digitálním prostředí

Abstrakt

Objem dat publikovaných na internetu roste. Na tomto faktu se podílí zejména rozšíření internetu a uživatelsky přívětivých technologií mezi laickou veřejnost.

S tím se do digitálního prostředí přesouvá významné množství osobních dat, která jsou cenná pro nejrůznější obchodní modely i nekomerční využití. Právě proto také zároveň roste počet incidentů při kterých byla uživatelská data zneužita a tím jejich soukromí narušeno.

Je proto zapotřebí neustále sledovat a zkoumat digitální hrozby, které vyvstávají v oblasti ochrany soukromí, bezpečnosti a svobody jednotlivců.

Tato práce představuje obvyklé typy unikajících dat, popisuje důvody a možnosti jejich zneužití a navrhuje opatření pro ochranu koncových uživatelů. Nakonec skrze experiment ověřuje, zda jsou data elektronických komunikací na vybraných službách strojově čtena a využívána pro cílení reklamy, a navrhuje možnosti zabezpečení soukromých konverzací.

Klíčová slova

Digitální stopa, ochrana soukromí, úniky osobních údajů, data mining, sledování, dohled

Protection of privacy in digital environment

Abstract

Volume of data published on the internet is growing. What contributes to this fact is mainly spread of internet and user friendly technology amongst laic public.

Following this trend, significant amount of personal data, which is valuable for diverse business models and noncommercial use, is submitted to digital environment. That is why the count of incidents when users' data were abused and their privacy breached, is rising as well.

It is necessary to constantly watch and examine digital threats rising in field of privacy, security and liberty protection.

This thesis presents usual types of leaked data, describes reasons for and possibilities of their abuse, and suggests measures to protect end users. Eventually, the thesis verifies whether data of electronic communications on selected media are machine-read and used for advertisement targeting, and suggests possibilities of protecting private conversations.

Keywords

Digital footprint, privacy protection, personal data leaks, data mining, tracking, surveillance

Obsah

Seznam tabulek

Seznam ilustrací

Úvod

Objem vyprodukovaných dat roste: denně je vygenerováno dat v řádech desítek až stovek exabytů.[1]

Největší podíl na tomto faktu má rozšíření internetu a mobilních chytrých zařízení mezi veřejnost a rozvoj technologií, které umožňují vytváření obsahu laickou veřejností, nebo-li co bývá označováno termínem web 2.0. V důsledku této změny lze pozorovat i kulturní posun k digitalizaci: lidé tráví více profesního i volného času online.

Spolu s tím se do digitálního prostředí přesouvá i velké množství osobních a citlivých údajů a jiných dat. Když si Rakouský student práva Max Schrems v roce 2011 na základě evropských práv na ochranu soukromí vyžádal od služby Facebook kompletní výčet dat, která o jeho osobě Facebook ukládá, obdržel přes 1200 stran dokumentu PDF, který obsahoval například přepisy všech publikovaných statusů, konverzací a všechny fotografie.[1][2]

Ve skutečnosti již není možné vést normální život bez uvolňování dat sběratelům, tj. poskytovatelům nejrůznějších služeb, mezi něž se počítají i služby veřejné správy.[1][3]

To vlastně ani není žádoucí: Služby založené na datech jsou pohodlné a užitečné a dokáží zprostředkovat daleko lepší servis, než jejich ekvivalenty, které data nevyužívají. Dokonce i cílené reklamy a doporučení mohou do určité míry zlepšovat zážitek z užívání internetu, na rozdíl od často hrubě nerelevantních necílených. Pohodlnost tohoto nového typu služeb pramení i z absence potřeby technologické gramotnosti; chytrá technologie dokáže bez přispění uživatele fungovat lépe, než její "hloupé" varianty a uživatelský zážitek je proto lepší.[1]

Proto samozřejmě není cílem (ani dostupnou možností) většiny společnosti se zcela odpojit od své digitální identity. Je ovšem důvod svou digitální identitu kultivovat a chránit. Osobní data uživatelů mají totiž ohromný ekonomický potenciál - ať už jsou monetizována legálně, nebo ne - i jiná využití.

Míra sledování - obchodní sférou i veřejnou správou - roste a více zasahuje soukromí.[2] Také se množí se případy úniků a zneužití osobních údajů: nové útoky se objevují téměř denně a zvlášť na základě událostí posledních let se tato problematika objevila i mezi hlavními zprávami mainstreamových médií.

To je ovšem pouze špička ledovce. Navzdory tomu, jak často se bezpečnostní incidenty objevují na specializovaných médiích, se do širokého povědomí dostávají jenom ty, které

se dotýkají jiných tradičních témat tisku, jako je například politiky, nebo ekonomiky. Nelze se to médiím vyčítat, protože se o technických tématech pro veřejnost (tj. ve srozumitelné a poutavé formě) píše obtížně.

Je zapotřebí na digitální bezpečnost dále upozorňovat veřejnost a poskytovat laickým uživatelům školení a návody informatické gramotnosti a zásad digitálního bezpečí. Ochrana osobních údajů je důležitá nejen z pohledu správce dat, ale i z pohledu uživatele, který musí umět určovat, kterým správcům která data poskytnout. Úniky dat - ačkoliv představují noční můru pro poskytovatele služeb - se samozřejmě týkají především uživatelů, neboť unikají data o nich, a to včetně dat, které jim mohou v nesprávných rukou způsobit újmu.

Odhlédneme-li od očividných praktických aspektů ochrany soukromí, dále vystává několik problémů, které je potřeba adresovat. Soukromí se nerovná tajemství. Říká se "nemám co skrývat", což ovšem musí zároveň znamenat "nikdo u mě nemá co hledat".

Právo na soukromí (co čehož za určitých podmínek spadá právo být zapomenut) je základní lidské právo a otázka internetové svobody je stejně relevantní otázka lidských práv, jako svoboda v offline světě. Soukromí je životně důležité pro osobní svobodu.[1] Nejvýraznější problém v tomto ohledu představuje fakt, že s určitou mírou nadsázky lze říci, že "vše co je potřeba pro policejní stát je dostupné na trhu".[2]

Práce na některých místech rozlišuje zneužití osobních dat:

- ze strany jejich majitele - původního sběratele, který data nějakým způsobem vytvořil a uložil
- ze strany útočníka, který data před zneužitím nějakým způsobem zcizí od sběratele
- ze strany útočníka, který data zcizí přímo uživateli

Zatímco jsou možnosti útoků na sběratele zmiňovány, hlavním směrem práce je ochrana koncových uživatelů a to zejména před zneužitím původními sběrateli dat. Pokud data nemůže zneužít ani jejich vlastník, tím spíš toho nedosáhne útočník na majitele útočící. Žádná databáze není bezpečná a úspěšné útoky na ně se dějí.[1] Přitom, jak bylo zmíněno, jsou to často právě data o uživatelích, která unikají.

Proto je z pohledu uživatele nutné ovlivňovat obsah databází a chránit se před nezodpovědností sběratelů. Ochrana před útoky cílenými přímo na koncové uživatele je doplňována pro úplnost a to především z pohledu rizik, která jsou reálná.

Kromě toho, některé body z ochrany uživatelů před útočníky mohou být použity i pro ochranu sběratele dat: zejména se jedná o ochranu před útoky směřujícím ke krádeži identity. Ty jsou často prostředkem pro proniknutí korporátní bezpečnosti. Zároveň

některá doporučení vycházející z ochrany před cílenými útoky mají platnost i v ochraně před sběrateli dat.

Práce obsahuje pouze obecná doporučení: někteří autoři se v radách na ochranu zaměřují detailně na tu či onu oblast hrozeb (M. Bazzell a G. Day na osobní útoky, B. Schneier na vládní sledování, J. Angwin na sledování obchodní sférou, F. Ahearn na cílené vyhledávání lidí, B. Krebs na organizovaný kyberzločin).[4][5][6][7] Tato práce má poskytnout obecný, ale komplexní souhrn jehož obsah je uplatnitelný nezávisle na situaci konkrétních čtenářů; literatura, ze které práce čerpá obsahuje konkrétnější návody pro ochranu.[8][9]

Stojí za zmínku, že někteří autoři (M. Perry, J.J. Luna, F. Ahearn, K. W. Royce) jdou často s radami do extrémů: doporučují občanům stěhovat se, hromadit zásoby a zbraně a podobně. Jiní (J. Angwin, B. Schneier, G. Day, M. Bazzell) volí pragmatičtější přístup, který nenarušuje běžný život, i za cenu toho, že jejich ochrana není stoprocentní.

Je rozdíl, jestli je cílem ochránit se před masovými úniky a hromadným zneužitím dat, nebo před cíleným sledováním jednotlivce.

V prvním případě stačí mít zabezpečení lepší, než ostatní. To projde sběrateli buďto bez povšimnutí, nebo přiměje útočníky hledat snazší cíle.[1] Ojedinělé průniky ochranou také nemají významný dopad.[2][1]

Ve druhém případě je nutné sáhnout po opatřeních, které nejsou vhodné pro každodenní život, nebo většinovou populaci.[3] V tom případě stačí k jedna chyba (přihlašování se k e-mailu z hotelů, kde je osoba přihlášena, EXIF data ve fotografii, jedna návštěva stránky během aktivního přihlášení na sociálních médiích) a důsledky bývají serióznější.[1]

Tato práce se bude zaměřovat spíše na první případ a pro druhou kategorii hrozeb doporučuje zmiňované autory.

Nakonec, v rámci rešerše jsou často příklady úniků a zneužití zobecněny. Citovaná literatura oplývá konkrétními příklady a čísly.

[1] Trefné je přirovnání k domovním lupičům, kterým nezáleží, který dům vykradou, tolik jako na tom, ve kterém mají menší šanci na neúspěch.

Cíl práce a metodika

Cíl práce

Primárním cílem práce je navržení a provedení experimentu, který má prokázat využívání dat ze soukromých elektronických konverzací pro potřeby cílení internetové reklamy.

Pokud se podezření prokáže, autor navrhne opatření pro ochranu jednoho vybraného komunikačního kanálu.

Vedlejším cílem práce je podání uceleného pohledu na problematiku digitálních stop a úniků osobních informací, včetně navržených protiopatření.

Metodika

Teoretická část se opírá o rešerši literárních a internetových zdrojů.

Experiment spočívá v monitorování množství a obsahu zobrazených reklam v kontrolním období a poté, co jsou do komunikačních kanálů podstrčeny informace, které mají za cíl zvýšit počet reklam relevantních k obsahu konverzací.

Hypotéza říká, že poté, co je komunikace ovlivněna, bude objektivně zaznamenáno více relevantních reklam. Tento rozdíl bude ověřen příslušnou statistickou analýzou.

Experiment spočívá v bodech:

1. Příprava scénáře experimentu
2. Příprava prostředí experimentu
3. Kontrolní měření
4. Ovlivnění výsledků
5. Vlastní měření
6. Ověření hypotézy

Teoretická východiska

Definice pojmů

Data a metadata

Prvními termíny, které je potřeba definovat, jsou data a metadata.

Data

Podle Cambridge Dictionary: "Informace, zvláště fakta a čísla, shromážděné za účelem zkoumání a považované za pomoc při vytváření rozhodnutí, nebo informace v elektronické formě, které mohou být uloženy a použity počítačem."

(v originále "*information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer*")[10]

Metadata

Podle Cambridge Dictionary: "Informace, které jsou poskytnuty k popisu, nebo pomoci k užití jiných informací."

(v originále "*information that is given to describe or help you use other information*")[11]

Rozdíl

Zatímco data nesou konkrétní informace o jejich předmětu, metadata jsou nejčastěji popisovány jako "data o datech".

B. Schneier vysvětluje rozdíl následovně: "data jsou obsah, metadata jsou kontext". Kontext je důležitý pro hromadné sledování, obsah pro jednotlivé. Proto by měla být metadata, kdykoliv je to možné, chráněna stejně jako data samotná.[1]

Typy citlivých dat

Práce operuje s několika přídavnými jmény užívanými pro popis dat, které jsou jejím předmětem. Většina definicí nejčastěji vychází z legislativy, a to zejména z českého zákona 101/2000 a nařízení GDPR, které je v době psaní práce pravděpodobně nejdůležitějším mezinárodním ujednáním vztahujícím se k tématu ochrany citlivých dat.

Citlivá data

Podle zákona 101/2000: "citlivým údajem osobní údaj vypovídající o národnostním, rasovém nebo etnickém původu, politických postojích, členství v odborových organizacích, náboženství a filozofickém přesvědčení, odsouzení za trestný čin, zdravotním stavu a sexuálním životě subjektu údajů a genetický údaj subjektu údajů; citlivým údajem je také biometrický údaj, který umožňuje přímou identifikaci nebo autentizaci subjektu údajů".[12]

Podle University of North Carolina: "Citlivá data jsou definována jako informace, které jsou chráněny před neodůvodněným zveřejněním. Přístup k citlivým datům by měl být chráněn. Ochrana citlivých dat může být vyžadována z legálních, nebo etických důvodů, kvůli problémům souvisejícím s osobním soukromím, nebo pro proprietární důvody".

(v originále "*Sensitive data is defined as information that is protected against unwarranted disclosure. Access to sensitive data should be safeguarded. Protection of sensitive data may be required for legal or ethical reasons, for issues pertaining to personal privacy, or for proprietary considerations.*").[13]

Osobní údaje

Podle zákona 101/2000: "osobním údajem jakákoliv informace týkající se určeného nebo určitelného subjektu údajů. Subjekt údajů se považuje za určený nebo určitelný, jestliže lze subjekt údajů přímo či nepřímo identifikovat zejména na základě čísla, kódu nebo jednoho či více prvků, specifických pro jeho fyzickou, fyziologickou, psychickou, ekonomickou, kulturní nebo sociální identitu".[12]

Podle GDPR: "„osobními údaji" (se rozumí - pozn. aut.) veškeré informace o identifikované nebo identifikovatelné fyzické osobě (dále jen „subjekt údajů"); identifikovatelnou fyzickou osobou je fyzická osoba, kterou lze přímo či nepřímo identifikovat, zejména odkazem na určitý identifikátor, například jméno, identifikační číslo, lokační údaje,

síťový identifikátor nebo na jeden či více zvláštních prvků fyzické, fyziologické, genetické, psychické, ekonomické, kulturní nebo společenské identity této fyzické osoby".[14]

Behaviorální data

Zákon 101, ani GDPR neobsahují definici behaviorálních dat a autorovi se nepodařilo najít přímou obecně platnou definici tohoto termínu, ačkoliv je často skloňován. Několik definicí poskytují weby zabývající se digitálním marketingem.

Podle Business Dictionary: "Informace, které jsou použity v marketingu pro návrh propagačních kampaní založených na nákupních zvycích, preferencích značky a používání produktu uživatelem".

(v originále "*Information employed in marketing for designing promotional campaigns based on consumers' buying habits, brand preferences, and product usage*")[15]

Podle 2040 Digital: "Behaviorální data je záznam interakcí, kliků, zapojení, pohybů skrze webovou stránku, akcí podniknutých skrze e-mailový marketing, uživatelův pohyb skrze digitální obsah a více".

(v originále "*Behavioral data is the capture of interaction, click, engagement, movement through a website, actions taken via e-mail marketing across the buyer journey, user movement through digital content and more.*")[16]

Pro potřeby práce autor používá zobecněnou pracovní definici: "Veškerý záznam chování uživatele, což obsahuje zejména všechny typy interakcí s digitálními produkty".

Biometrická data

Podle GDPR: "„biometrickými údaji" (se rozumí - pozn. aut.) osobní údaje vyplývající z konkrétního technického zpracování týkající se fyzických či fyziologických znaků nebo znaků chování fyzické osoby, které umožňuje nebo potvrzuje jedinečnou identifikaci, například zobrazení obličeje nebo daktyloskopické údaje".[14]

Genetická data a údaje o zdravotním stavu

Podle GDPR: "„genetickými údaji" (se rozumí - pozn. aut.) osobní údaje týkající se zděděných nebo získaných genetických znaků fyzické osoby, které poskytují jedinečné informace o její fyziologii či zdraví a které vyplývají zejména z analýzy biologického vzorku dotčené fyzické osoby".[14]

Podle GDPR: "„údaji o zdravotním stavu“ (se rozumí - pozn. aut.) osobní údaje týkající se tělesného nebo duševního zdraví fyzické osoby, včetně údajů o poskytnutí zdravotních služeb, které vypovídají o jejím zdravotním stavu".[14]

Digitální stopa

Podle Dictionary.com: "unikátní set digitálních aktivit, akcí a komunikací, které zanechávají datovou stopu na internetu, počítači, nebo na jiném digitálním zařízení, a které mohou identifikovat konkrétního uživatele, nebo zařízení".

(v originále: "*one's unique set of digital activities, actions, and communications that leave a data trace on the Internet or on a computer or other digital device and can identify the particular user or device*")[17]

Podle Technopedia: "Digitální stopa je unikátní datová stopa uživatelových aktivit, akcí, komunikací a transakcí na digitálním médiu. Tato datová stopa může být zanechána na internetu, počítačích, mobilních zařízeních, nebo jiných médiích. Digitální stopa může být využita ke sledování uživatelových aktivit a zařízení. Uživatel může zanechat digitální stopu aktivně, nebo pasivně, ale jakmile je jednou sdílena, je téměř permanentní".

(v originále "*A digital footprint is a unique data trace of a user's activities, actions, communications or transactions in digital media. This data trace can be left on the internet, computers, mobile devices or other mediums. A digital footprint can be used to track the user's activities and devices. A user can leave digital footprint either actively or passively, but once shared, a digital footprint is almost permanent in nature*")[18]

Digitální stopa jsou tedy elektronické, digitální informace, které jsou o uživateli zanechávány používáním obecně všech IT zařízení, ale typicky především počítače, smartphone a jiných chytrých zařízení a počítačových sítí - hlavně samozřejmě internetu. Literatura zpravidla uvádí dělení na aktivní stopy (vytvořené za vědomí a přičinění uživatele) a pasivní (vznikající bez vlastního přičinění uživatele).[8][18]

Typy digitálních stop

Pro potřeby členění následující kapitoly jsou digitální stopy členěny celkem do čtyř kategorií. Zatímco se v literatuře objevuje pouze výše popisované rozdělení, v otázkách stop vytvořených třetí stranou (které se týkají uživatele) nemusí být zařazení jednoznačné: z pohledu jejich tvůrce se jedná o výsledek vlastní akce, ale z pohledu osoby, jíž se obsažená data týkají, jde o pasivní stopu, neboť vznikla zcela bez jejího přičinění. Tento případ tedy pro přehlednost práce vyčleňuje mimo obecně uváděnou kategorizaci.

Dalším případem, který taktéž autor vyčleňuje z tradičního rozdělení, jsou stopy vynucené, které vznikají na uživatelově zařízení, případně i jeho přičiněním (tedy daly by se za normálních okolností klasifikovat mezi pasivní, nebo aktivní), ale na základě chování software, který tato data vytváří bez uživatelova vědomí či souhlasu a to právě za účelem sběru dat (tedy přímým důsledkem akce třetí strany).

Na další problém naráží občas uváděné rozdělení podle vědomí uživatele: pokud si například uživatel uvědomí, že komunikací na internetu odhaluje svou IP adresu, ta nemůže změnit kategorii.

Autor proto navrhuje následující rozdělení stop nikoliv podle vědomí uživatele, ale podle míry přičinění a původce dat. Ve všech případech si vlastní uživatel může nebo nemusí být vědom toho, do jaké míry jsou data zaznamenána.

- Aktivní: Data vytvořil uživatel; jsou přímým důsledkem uživatelovy akce (vytvoření stopy je cílem akce)
- Pasivní: Data vytvořil uživatel, nebo jeho zařízení; jsou nepřímým důsledkem uživatelovy akce (vytvoření stopy není cílem akce)
- Třetích stran: Data nevytvořil uživatel; jsou přímým i nepřímým důsledkem akce třetí strany
- Vynucené: Data vytvořil uživatel, nebo jeho zařízení; jsou přímým důsledkem akce třetí strany

Například: Pokud uživatel zadá do vyhledávače určitý řetězec, je to podle autora aktivní stopa: autor tento řetězec zcela cíleně napsal a odeslal. IP adresa asociovaná jako původce požadavku je pasivní stopa, protože je nepřímým důsledkem předchozí akce. Pokud vyhledávací engine na základě historie vyhledávání vytváří profil uživatele, jsou to nová data vytvořená o uživateli třetí stranou a proto se neklasifikuje ani do jedné z původních kategorií. Pokud je navíc na uživatelově stroji nainstalován malware typu keylogger, který zachytí a odešle vyhledávaný řetězec, je tato stopa považována za vynucenou.

Autor na tomto místě připomíná, že toto rozdělení nemá oporu v literární rešerši, ale je jím využito pro dekompozici dlouhého textu a pro oddělení stop, které by mohly nabývat znaků obou kategorií.

Aktivní stopy

Aktivní stopy jsou takové, které uživatel vědomě vytváří a vkládá do digitálního prostředí; je si obvykle vědom, že data, která vytvořil budou někde zaznamenána, někam putovat a někde zpracována – to ale nemusí být podmínkou. Některé z později uvedených stop uživatel zanechává skrze používání produktu a nemusí si, obzvláště, pokud není technicky vzdělaný, uvědomovat, že jsou někde uchovány a sledovány.[8]

Vlastní příspěvky

Do této kategorie spadá veškerý obsah (textové příspěvky - články, fotografie, videa, biografické údaje), který autor publikuje na vlastním webu, blogu, Wiki apod.

Uživatel obvykle má zájem jej mít spojený se svou reálnou identitou, nebo digitálním pseudonymem, neboť tak naprosto vědomě buduje vlastní mediální obraz. Zároveň také zpravidla má nad těmito médii kontrolu.

Rizikem je nezodpovědné vyzrazení citlivých informací, které lze použít proti autorovi (to je zvláště nebezpečné u dětí) a nebo, v případě, že autor publikuje anonymně, spojení pseudonymu se skutečnou identitou.

Také, jednou publikovaný obsah nelze nikdy permanentně odstranit z historie, protože může zůstat zachycen službami typu Web Archive, nebo indexován roboty vyhledávačů.[1][8]

Sociální média

Vyplněné profily a přidané příspěvky, komentáře a multimediální obsah jsou opět typicky přímým důsledkem budování vlastní digitální identity.

Důležité je zde uživatelské nastavení (a technická implementace) ochrany soukromí, neboť ne všechny informace mají být veřejné mimo okruh autorizovaných kontaktů (přátel).[4]

I při správné bezpečnostní politice ovšem zůstávají data v moci samotného provozovatele sociální sítě, který jich může využívat v marketingu (Facebook, například, má velmi dobrou nabídku cílení reklam podle lokace, demografie, zájmových skupin apod.), nebo je může, v nejhorším případě, poskytnout třetím stranám (v případě útoku, nebo soudního příkazu nedobrovolně, jinak v rámci obchodních aktivit vědomě).[1]

Útoky na sociální sítě nejsou nijak vzácné: v minulosti unikly data například sítí MySpace, LinkedIn a seznamek AdultFriendFinder a Ashley Madison. Při nich unikaly typicky uživatelská jména, e-maily a hesla.[19]

Méně zřejmé může být to, že obsahem na sociálních médiích jsou i uživatelem sledované zájmové skupiny: práce "Computer-based personality judgments are more accurate than those made by humans" vypracovaná univerzitami v Cambridge a Stanfordu ukázala, že po korelaci uživatelských dotazníků a jejich Facebookových profilů byli výzkumníci schopni poměrně přesně odhadovat vlastnosti jiných uživatelů Facebooku jenom na základě jejich profilů. Jednalo se přitom například i o otázku sexuality, která byla odhadnuta na základě sledování kosmetických produktů, nebo různých umělců. Na nejlepší predikce – lepší, než které vytvořili nejbližší přátelé – stačilo zhruba 300 označení „to se mi líbí" u stránek.[20]

Rizikem sociálních sítí jsou tedy zejména vyzrazení a neplánované šíření osobních údajů, názorů, povahových rysů, zájmů, mapy přátel a známostí a dalších citlivých dat mimo okruh zamýšleného publika a zneužití poskytovatelem. Všechny tyto informace jsou cenné pro cílení reklamy, ale především pro potenciální státní represi, nebo civilní zločince. Starý obsah sociálních médií je také velice dobrým materiálem pro bulvární tisk.[2][1][8][6][5][7]

Elektronická komunikace

Do této kategorie spadají například e-mail, chat (instant messaging), SMS a telefonní a video hovory.[1][8]

Jsou to vědomě odeslané informace určeny pouze adresátovi (adresátům). V tomto případě ale zpravidla není adresát zodpovědný za infrastrukturu komunikace – doručení zprávy je zodpovědnost poskytovatele (či poskytovatelů v součinnosti) služby.

Pokud není mezi koncovými zařízeními nastavená nějaká forma šifrování nepřístupného prostředníkům, svěřují uživatelé svá citlivá data třetí straně (poskytovateli služby), které věří, že jejich zprávy nebude analyzovat, ani nezprostředkuje další straně, která bude (zejm. bezpečnostním složkám), a bude je chránit před únikem. Kromě toho vznikají i

na izolovaných zařízeních (např. telefonech, nebo e-mailových klientech využívajících POP3 protokol) historie komunikace, ke kterým lze při jejich získání přistupovat.

Tato data lze podle platného trestního zákoníku považovat za chráněné listovní tajemství ("tajemství dopravovaných zpráv") a i policie k jejich oficiálnímu sledování potřebuje souhlas soudu (podle údajů ministerstva vnitra bylo v roce 2017 takto sledováno 3241 osob).[21][22] Přesto, neetické sledování může probíhat ilegálně, případně v jurisdikci státu, který takovou ochranu elektronických komunikací zakotvenou v zákoně nemá, a tak vzhledem k tomu, jakou část lidské komunikace tvoří digitální přenos dat, je nutné zvážit, jak citlivé informace si lidé těmito kanály sdělují.[1]

Menší ochraně než obsahu zpráv jsou podrobena metadata o komunikacích (už z důvodu, že z technologických důvodů je není možné šifrovat ani jinak skrývat). V případě e-mailu se jedná o například o identifikaci původce a cíle, datum, velikost zprávy a historii předávání zprávy mezi e-mailovými servery; v případě telefonu pak jsou zaznamenány telefoní čísla původce a cíle, čas hovoru a jeho délka a lokace obou telefonistů, na základě lokace vysílače, k němuž jsou připojeni.[1][5]

Problémy využívání metadat se dále zabývá kapitola "Data mobilních sítí".

Kontaktní formuláře

Podobný typ dat, jako elektronická komunikace, představují zprávy doručené skrze kontaktní formuláře poskytované adresátem. Rozdíl oproti elektronické komunikaci, jak byla vylíčena v předchozí kapitole, je ten, že infrastrukturu obvykle nespravuje třetí strana, ale poskytovatel služby a na něm proto leží zodpovědnost za uchování dat poskytnutých uživatelem v soukromí.

Existují problémy vztahující se k povaze dat, které člověk sdílí a k potenciálně nedostatečnému zabezpečení kontaktních rozhraní. Například při objednání dovážky občerstvení typicky zákazník uvádí své celé jméno, e-mail, telefon a adresu bydliště a tyto mimořádně cenné osobní údaje jsou – v nejhorším možném případě po odeslání nezabezpečeným HTTP protokolem ze špatně zabezpečené, nebo veřejné Wi-Fi – poté obvykle uloženy do přinejlepším standardně zabezpečené databáze a také vyslány skrz internet nešifrovaným e-mailem (z web serveru hostujícího objednávkovou aplikaci na mail servery adresáta i zákazníka). Možností k odposlechnutí těchto informací je proto více než dostatek.

Tristní zabezpečení takových formulářů je dáno často tím, že se menší podniky spoléhají na levné out-of-the-box řešení, zpravidla typu pluginu pro CMS WordPress. Takováto řešení jsou často terčem hromadných útoků roboty: pokud se objeví jedna bezpečnostní

mezera a plugin je nainstalován na milionu webů, nemusí být konkrétní instalace, která se stane obětí útoku, pro útočníka ani nijak zajímavá.[7]

Využívání komponent se známými zranitelnostmi je na seznamu nejvýznamnějších zranitelností webových aplikací vydaném pro rok 2017 organizací OWASP umístěno na devátém místě a právě pluginy pro WordPress se v minulosti staly terči úspěšných XSS útoků (číslo 7) a miskonfigurace zabezpečení, do čehož spadá i absence aktualizací systému, je dalším častým prohřeškem levných řešení.[23][24]

Historie nákupů

Stejné problémy, jako ty popisované v souvislosti s kontaktními formuláři, vyvstávají v oblasti eCommerce. Nadto je ovšem potřeba adresovat několik dalších myšlenek.

V první řadě do řešení obvykle vstupují platební karty, které představují jeden z nejčastěji unikajících typů dat.[2][3][1][6][7] Kromě toho s digitální identitou propojují i offline nákupy - včetně lokace terminálu.[1]

Dále díky častějším online nákupům vzniká historie nákupních zvyků, která je cenným zdrojem pro online marketing. Kromě existence samotných objednávek, nejen, že platební společnosti i kryptoměny musí udržovat historii finančních transakcí, kterou lze sledovat (v případě kryptoměn je pro sledování nutné je ještě spojit elektronickou peněženku s reálnou identitou), také služby typu Steam, Google Books, nebo App Store musí udržovat databázi nákupů digitálních komodit (her, e-knih, a software, v uvedených příkladech) spojených s konkrétním uživatelem.[2][1]

V západním světě jsou například záznamy o nákupech, nebo výpůjčkách knih považovány za citlivé informace, s ohledem na ochranu osobní svobody. Čtení dejme tomu životopisů kontroverzních osob, nebo jinak „nevhodných“ děl totiž může přitahovat pozornost bezpečnostních složek.[2][25]

Podobným případem jsou potom služby založené na periodickém předplatném - ať už se jedná o digitální obsah, nebo fyzická periodika. Ty prozrazují zájmy a případně bydliště odběratele.[3]

Všechna tato data jsou užitečná pro cílení reklamních kampaní a to především tehdy, jsou-li data nákupů z více obchodů spojována podle kontaktních e-mailů, nebo čísel platebních karet.[1]

Historie vyhledávání

Člověk si nemusí vždy uvědomovat, že je jeho historie vyhledávání tvořena nejen v prohlížeči, ale že i vyhledávací engine ukládá hledané výrazy a snaží se je spojovat s konkrétními osobami.

To umožní například společnostem Google, nebo Microsoft vytvářet modely chování osob, určit zájmové skupiny, demografické údaje a další lidské vlastnosti, podle kterých lze také velmi dobře cílit reklamu.[2]

B. Schneier vyslovuje názor, že uživatelé jsou k vyhledávači upřímnější a svěřují mu citlivější informace, než ostatním lidem - zejména jedná-li se o choulostivá témata typu zdravotních komplikací.[1] Kromě toho stojí za zmínku myšlenka, že vyhledávané termíny jsou odrazem myšlenkových pochodů uživatelů, kteří hledají, na co právě myslí.[2][1]

Osobní sledování

Fitness náramky jsou oblíbeným typem nositelné chytré elektroniky, který umožňuje uživateli sbírat a vizualizovat data o jeho fyzické aktivitě a zdravotním stavu. Mnoho webů, které obsahují například mapu prodejen provozovatele, umožňuje uživateli sdělit svou aktuální polohu skrze GPS čip v chytrých telefonech pro nalezení nejbližší provozovny. Facebook umožňuje u příspěvků označovat lokalitu, které se týkají.

Riziko je zřejmé: uživatel aktivně zpřístupňuje svou lokaci a informace o denním režimu. Kuriózním, ač seriózním, případem byla fitness aplikace Strava, jejíž autoři zveřejnili tepelnou mapu aktivit jejich uživatelů. Problémem bylo to, že ji využívali vojáci americké armády a proto se na mapě blízkého východu jasně rozsvítily americké vojenské základny.[26]

Dále do této kategorie stop spadají také všechny digitální způsoby osobní organizace: kalendáře, osobní seznamy úkolů a podobně, které dále odhalují agendu, zájmy a další odrazy myšlení jejich správce.[1][2]

Pasivní stopy

Do pasivních stop se řadí ty, které uživatel nezanechává vědomě, nebo přinejmenším ne cíleně. Zatímco shromažďování aktivních stop vyžaduje lidskou kooperaci, nebo

alespoň souhlas, pasivní stopy bývají často vedlejším produktem technologického řešení produktu; jejich vznik může být nezbytný pro fungování sítě, nebo služeb.[8]

IP adresa, MAC adresa

Obě adresy identifikují zařízení, které je původcem, nebo cílem komunikace: jsou nezbytné pro praktickou realizaci komunikace, a právě proto je lze sledovat, nebo podvrhnout (tzv. IP spoofing a MAC spoofing). Pokud útočník získá přístup k síťovému zařízení, nebo k celé síti kdekoliv během cesty, po které protéká klientova komunikace, může monitorovat veškerou jeho aktivitu na internetu (sledováním obou adres, portů i obsahu datových packetů), nebo se za něj vydávat (disponováním falešnými adresami přimět aktivní síťové prvky, aby uživatelova data nesprávně směrovala na útočníkův stroj) a tím aktivně do komunikace zasahovat. (tzv. DNS spoofing, route injection)[27][28][29]

Oba příklady lze provést za předpokladu, že komunikace není šifrována a protéká zcela tímto napadeným zařízením, či sítí; i ze šifrované, nebo neúplné komunikace lze ovšem z principu vyčíst identita adresáta – tedy například web server, se kterým uživatel komunikuje. Reverzním DNS vyhledáváním lze potom zjistit, jaké weby se na tomto serveru nachází.

IP adresu lze mimo jiné běžně lokalizovat na úroveň měst a poskytovatel připojení potom dokáže komunikaci přiřadit k naprosto konkrétní přípojce, jejíž ID má spojené se smlouvou s konkrétním zákazníkem.[3][6][8]

Tyto adresy se také například zaznamenávají v hlavičkách e-mailů. Lze tak za určitých podmínek do určité míry stopovat původ e-mailu.[1]

Poskytovatel připojení

Poskytovatel připojení (ISP) je identifikován na základě IP adresy a opět tedy vychází z běžné technické implementace.

Znalost ISP jednak zlepšuje možnosti geolokace komunikujícího zařízení (ISP zpravidla uvádí lokaci svého sídla) a také poskytuje informace pro phishing útoky: například pro akce, kdy se útočník vydává za poskytovatele uživatelova připojení a snaží se z něj vylákat informace, které by zneužil ke krádeži jeho identity, nebo naopak pro akce, kdy se vydává za klienta a snaží se z poskytovatele přimět, aby mu zpřístupnil informace o skutečném klientovi.[3]

Data mobilních sítí

Data vytvořená v mobilní síti obsahují zejména telefoní čísla původce a cíle, čas hovoru a jeho délku a lokaci obou telefonistů, na základě lokace vysílačů, k nimž jsou připojeni.

Tento typ dat také vychází z technické nezbytnosti: zařízení přirozeně musí být připojeno k signálu, aby mohlo přijímat hovory a zprávy, a identifikace zdroje a cíle komunikace je nutná pro ustanovení spojení.[1]

Jak jsou tato data zneužitelná ilustruje existence zařízení známých pod názvy "Cell-Site Simulators", "Stingrays", nebo "IMSI Catchers", které imitují chování vysílačů mobilního signálu, čímž přesvědčují mobilní telefony, aby se k nim připojovaly a veškerá zmíněná data posílaly skrze Stingray.[30][1]

Lokace

Na základě síly signálu vysílaného mobilním zařízením a přijímaného několika anténami mobilní sítě, lze průnikem dosahu těchto antén určit přibližnou polohu vysílajícího zařízení, a to s přesností na zhruba 600 metrů.[1][2][3][5]

Lepší přesnost (v řádech desítek metrů) má lokace na základě Wi-Fi - pokud je router poskytující Wi-Fi nakažen sledovacím malware.[2][3][5]

Na základě lokace telefonu v určitý čas lze identifikovat bydliště a pracoviště uživatele, případně oblíbené způsoby trávení volného času. Pokud má člověk relativně stabilní denní a týdenní režim, lze tak s určitou pravděpodobností předvídat jeho lokaci v budoucnosti.[1]

Sociální dynamika

I na základě lokace lze vytvářet mapy známostí: pokud se dva telefony vícekrát připojily ve stejný čas ke stejné věži, indikuje to propojení jejich majitelů.[1][2]

Další informace, které jsou mobilními operátory zaznamenávány, jsou informace o hovorech (datum a čas, délka trvání a samozřejmě adresáti), zaslaných SMS (datum a zdroj a adresát) a o mobilním připojení k internetu. Tyto informace musí operátoři podle zákona udržovat po dobu půl roku.[31][32]

Tato data přitom mohou prozrazovat dostatek informací i bez znalosti přesného obsahu komunikací: především lokaci osob, sociální dynamiku a v případě známých čísel také téma hovoru.

V případě známých čísel, výzkumníci během pokusu identifikovali - i bez znalostí obsahu hovoru - pacienty s roztroušenou sklerózou a infarktem, sběratele zbraní, domácího pěstitele marihuany a osobu, která podstoupila umělé přerušení těhotenství: vše na základě znalosti jednoho konce hovoru - veřejné služby, jejíž telefoní číslo a popis služeb jsou veřejně dostupné.[1]

Data družicových polohových systémů

Každé moderní mobilní zařízení obsahuje čip globálních družicových polohových systémů; nejčastěji GPS, méně často také GLONASS, nebo Galileo.

Využití dat o lokaci bylo popisováno v sekci "mobilní sítě", zde je pouze na místě dodat, že zatímco přesnost lokalizace mobilního signálu je v řádu stovek metrů, v případě těchto systémů se jedná o jednotky metrů i méně.[1][6][5]

Uživatelovo zařízení

Do této kategorie spadají všechny informace o webovém prohlížeči, operačním systému, velikosti zařízení a dalších vlastnostech uživatelova zařízení (tzv. browser fingerprint, nebo obecně device fingerprint).

Tyto informace, které standardně odesílá webový prohlížeč jako hlavičky HTTP požadavku (zejm. tzv. User-Agent, ale i další), nebo které detekuje javascriptový kód spuštěný v prohlížeči, lze využít pro optimalizování uživatelského zážitku: například se layout webu přizpůsobí podle velikosti displeje, uživatel je přesměrován na správnou jazykovou variantu webu, nebo při stahování software web nabídne přímo verzi pro detekovaný operační systém.

Na druhou stranu zároveň do určité míry zlepšují potenciálnímu hackerovi, nebo škodlivému kódu možnosti proniknutí do zařízení, neboť je v první řadě v hackerské komunitě známé, ve kterém software a ve kterých jeho verzích jsou dostupné jaké zranitelnosti a nezná-li útočník prostředí, musí nejprve tyto zranitelnosti odhalit a vyzkoušet, čímž na sebe může sám upozornit.[1]

Kromě toho je možné tyto stopy využít v phishingovém útoku: představme si zprávu týkající se „bezpečnostního hrozby“ uživatelova programu a nabízející „bezpečnostní záplatu“, která má přimět uživatele spustit útočníkův kód na svém stroji. Taková zpráva by neměla valný účinek, pokud by se týkala software, který uživatel nepoužívá.[8]

Nejdůležitější je ovšem opět hrozba jednoznačné identifikace uživatele.

K tomuto účelu je vhodný právě řetězec User-Agent, který vypadá například následovně:

```
Mozilla/5.0 (Windows NT 10.0; WOW64)
AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/71.0.3578.98 Safari/537.36
OPR/58.0.3135.79
```

Tento řetězec je přeložen zhruba jako "Opera 58 na Windows 10", či více specificky "Opera 58.0.3135.79". To je velice přesná informace, která velice zužuje množinu potenciálních původců komunikace. Ve skutečnosti, studie EFF objevila, že pouze 1 z 286 777 klientů sdílí User-Agent s jiným uživatelem.[33][34]

Dále lze identifikaci zpřesnit zvláště v kombinaci s dalšími údaji získanými z dalších HTTP hlaviček, nebo sesbíranými javascriptovým kódem (mezi něž patří zejm. zdrojová IP adresa, rozlišení displeje zařízení, výchozí jazyk, nebo časové pásmo apod.). Toho lze využít ke sledování aktivit uživatele napříč weby.[33]

Hlavička User-Agent (ani další) není technicky nutná a je možné jí falšovat, nicméně je také možné ji na straně serveru využívat při autentizaci uživatele, jako ochranu před tzv. cookie hijacking, takže její falšování může znepřístupnit legitimní přístupy k některým službám.[1]

Historie prohlížení webu

Jak bylo zmíněno u digitálních adres, cíl komunikace lze standardně odposlouchávat i pokud je obsah komunikace šifrovaný a to v případě, že útočník ovládá některý síťový prvek. Existují ale jiné způsoby, jak může historie prohlížení unikat a být zachycena na straně serveru.[1][6][8]

HTTP referer

Webové prohlížeče odesílají s požadavky hlavičku HTTP referer pro označení poslední navštívené URL, z níž pocházel odkaz, který odkazoval na adresu současného požadavku. To programátorům webových služeb umožňuje vytvářet určité jednoduché statistiky chování návštěvníků webu – map toků prokliků skrz web, nebo zdrojů návštěv (sociální média, vyhledávač, odkazy z jiných webů...) –, nebo na základě chování upravit uživatelské rozhraní (přidáním tlačítka „zpět“, nebo zobrazením nějaké zprávy).

Tato hlavička zpravidla není nutná, za několika podmínek není odesílána vůbec, a opět je možné ji podvrhovat, nebo vůbec neodesílat, nicméně opět na ní některé aplikace

mohou záviset, například při ověřování validity požadavků. Protože na referer nelze ze strany programátora příliš spoléhat, používají se v určitých případech pro podobný výsledný efekt i jiné techniky. Ty jsou dále popisovány v části věnující se vynuceným stopám.[35]

Záznamy aktivity

Nejrůznější aplikace pravidelně vytvářejí logy aktivit uživatelů, aby bylo možné jejich chování, nebo výkon aplikace analyzovat pro potřeby optimalizace systému (případně pro potřeby marketingu), nebo v případě nežádoucího chování systému chyby, nebo potenciální zneužití vysledovat ke konkrétnímu člověku. Toto je standardně součástí podnikových i komerčních aplikací a liší se v závislosti na zamýšlených cílech sledování.

Podnikové aplikace standardně vyžadují historii větších událostí – typicky změn v datech – z důvodu vyvozování personálních zodpovědností. Komerční aplikace na druhou stranu budou pravděpodobněji potřebovat data o běhu aplikace pro optimalizace výkonu, nebo uživatelská behaviorální data pro optimalizace uživatelského dojmu.

Takový monitoring je zakódován v aplikaci samotné, není možné se mu vyhnout a většinou ho ani nelze jednoduše detekovat. Záleží na architektuře aplikace: sledování zabudované v desktopových nebo mobilních aplikacích, nebo na serverové straně webových aplikací uživateli přístupné není; pro odhalení sledování je nutné sledovat síťovou komunikaci a v ní hledat podezřelá data.[1][6]

Komerční sledování aktivity na webu se prolíná se sledováním HTTP cookie, bylo načrtnuto už v předchozí kapitole a je dále rozvíjeno v části popisující vynucené stopy.

Metadata

Metadata jsou součástí velkého množství souborů vytvořených uživatelem. Dobrým příkladem jsou metadata fotografií a skenů, které obsahují tzv. EXIF data, která mohou obsahovat fyzickou lokaci (GPS) místa, kde byla fotografie pořízena, nebo ID kamery a další parametry, které mohou identifikovat zdroj.[1][6][5][3][8]

Dalším příkladem jsou dokumenty kancelářských balíků, které obsahují metadata o zařízení a uživateli, který je vytvořil.[6][8]

Zvláštním případem hodným zmínky je systém Machine Identification Code (MIC) - vodoznak, který u fyzických dokumentů identifikuje tiskárnu původu a je vytištěn ve

formě nepatrných žlutých teček. Tento vodoznak byl vyvinut již kolem osmdesátých let a objeven teprve v roce 2004.[3][36]

Nízkoúrovňové záznamy

Velkou skupinu pasivních záznamů tvoří všechny vyrovnávací paměti, dočasné soubory a zálohy vznikající na všech úrovních výpočetní techniky. Například:

- Aplikace si samy vytvářejí dočasné soubory pro plynulejší fungování; z internetu stahují statický obsah, nebo si poznamenají výsledky náročných operací, aby je nebylo nutné znovu stahovat/vykonávat pro všechny následující operace.[1]
- Také je obvyklá funkce ukládání konceptů, revizí a automatických záloh pro případ selhání aplikace.[1]
- Jakékoliv zařízení si může vytvářet interní logy pro případy, kdy je potřeba diagnostikovat problémy, nebo zjišťovat jejich status. Jedním příkladem jsou černé skříňky automobilů a to zejména těch samořiditelných.[1]
- Zálohy databází, nebo techniky škálování výpočetního výkonu (duplikování obsahu na více instancí aplikace, nebo tzv. Content Delivery Networks apod.) mohou uchovat data, která se uživatel snažil odstranit.
- Stejně tak mohou nežádoucí obsah zachovat všechny služby, které jej indexují (například v prostředí databází, nebo vyhledávacích nástrojů).
- DNS cache routeru, nebo operačního systému obsahuje domény, ke kterým zmíněný stroj, nebo jiný v síti routeru přistupoval.
- Smazané soubory z počítače ve skutečnosti na pevném disku zapsané zůstanou. Při smazání ze souborového systému se pouze smaže záznam o souboru a blok paměti, kde byl zaznamenán, je označen jako volný k zápisu: dokud do něj ale není nic zapsáno, data souboru zůstávají nedotčená.[3]
- Stejně tak se může dočasně uchovat otisk souboru v operační paměti a obecně hierarchie počítačových pamětí víceméně počítá s tím, že každé pomalejší médium může používat rychlejší technologii jako cache (webové úložiště používá pevný disk, ten má vlastní diskovou cache, dále následují operační paměť, cache procesoru a registry).

Rekonstruování smysluplných informací z takovýchto dat je jistě obtížné, ale jinak je teoreticky možné. V těchto datech patrně nebudou ukryté marketingově zajímavé

informace, ale mohou posloužit při policejním vyšetřování, nebo při různých hackerských útocích.[3][6]

Stopy vytvořené třetími stranami

Mezi stopy vytvořené třetími stranami tato práce počítá všechny stopy, které nesou informace o konkrétním uživateli, a které vytvořili a uchovávají jiní uživatelé, nebo instituce, a to nezávisle na míře uživatelova vědomí.

Obsah publikovaný známými

Prvním okruhem třetích stran, které produkují data o uživateli, jsou jeho přátelé a známí, zájmové skupiny, nebo zaměstnavatel. Ti mohou data nevědomky vyzradit při phishingovém útoku, nebo sami publikovat z nejrůznějších důvodů, do nichž spadá i kyberšikana.

Zvláště v případě sociálních sítí se jedná o stejný obsah dat, který může publikovat uživatel sám - označení lokace, sociální dynamika, textový obsah nebo fotografie.[3][1]

Data produkovaná soukromým sektorem

Soukromý sektor má mnoho legitimních důvodů pro shromažďování osobních údajů: zejména se jedná o náležitosti obchodního styku a jiné legislativní nároky. Tato sekce popisuje data, která je potřeba udržovat pro využívání zmíněných služeb.

Eshopy a platební brány

Eshopy v rámci správy objednávek uchovávají všechny tradiční osobní údaje - především jména a adresy. Data z objednávek lze také sledovat na základě sériových čísel výrobků v záručních listech, které jsou provázány s daty o objednávkách.[3]

Nadto stejná i rozsáhlejší data uchovává každý obchod, který používá systém typu Customer relationship management (CRM). Mezi ně patří především obchody s obchodním modelem na základě předplatného (jmenujme například posilovny, nebo digitální i fyzická periodika), nebo s věrnostním programem. Do těchto rozšířených dat obvykle patří zpracovaná data o nákupních zvycích konkrétních nakupujících, která jsou cenná pro marketingové aktivity.[1][3]

Data z obchodních styků figurují na seznamech největších datových úniků: Jejich cílem se staly například společnosti Heartland payment systems (unikly kreditní a debitní karty), Uber (jména, e-maily, telefony, čísla řidičských průkazů), eBay (jména, adresy, data narození, hesla), Playstation Network (jména, e-maily, hesla, adresy, kreditní karty), Yahoo (jména, e-maily, data narození, telefony, hesla), NetEase, Adobe, Home Depot, Target Stores, nebo TJX Companies (ve všech případech kreditní karty).[19]

Bankovní a jiné finanční instituce

Detailnější informace o finanční situaci uchovávají banky a společnosti zabývající se vyhodnocováním úvěrových zpráv. Opět ukládají všechny tradiční osobní údaje, ale nadto i informace o příjmech, úsporách a majetku osob: vše s cílem vyhodnotit finanční kredibilitu klienta.[3]

Mezi významné úniky těchto dat patří JP Morgan Chase (jména, adresy, telefony, e-mailové adresy), Equifax (čísla sociálního zabezpečení, datum narození, čísla řidičských průkazů, informace o kreditních kartách), nebo Anthem (čísla sociálního zabezpečení, data narození, historie zaměstnání).[19]

Hotely

Ubytovací zařízení od návštěvníků vyžadují vyplnění osobních dat z legislativních důvodů.[1][37]

V minulosti tato data unikla například ze sítě Marriott International (jména, adresy, čísla pasů, cestovní informace, čísla kreditních karet)[19]

Data brokers

Samostatnou kategorií soukromého sektoru jsou tzv. datoví makléři (data brokers) - společnosti, jejichž obchodním modelem je sběr, agregace a následné prodávání dat. Tyto společnosti spojují různé datasety a výsledný objem informací shromážděných o konkrétních uživatelích je značný: jména, adresy, data narození, kontaktní informace, demografie (etnicita, náboženské přesvědčení, věk, pohlaví, sexualita, vzdělání, historie zaměstnání), příjmy, majetek a mnoho dalších faktorů, včetně takových detailů jako jsou zájmy, nebo vlastnictví domácích zvířat.[1][2]

Přesně tato data se objevila při úniku z firmy Exactis - přes 400 proměnných pro 340 milionu záznamů.[38]

Data produkována veřejným sektorem

Ze zřejmých důvodů jsou to úřady a jiné prvky veřejné správy, nebo veřejných služeb (policie, zdravotnické instituce, veřejné školství), které drží největší objem neanonymních dat o občanech. Například:

- jména, jména spřízněných osob a vztahy (zejm. data o rodině)[2][6][5][7]
- adresy bydliště[2][5][7]
- telefonní čísla[2][5][7]
- e-mailové adresy[2][6][5][7]
- matrika - rodinný stav, narození, úmrtí[1]
- občanské, řidičské, zbrojní a další průkazy[1]
- záznamy v registru vozidel[3]
- záznamy kriminální historie (a informace obsažené v policejních hlášeních)[3]
- finanční data, udělené koncese, pracovní historie[1]
- sociální pojištění
- zdravotní pojištění a lékařská data (lékařské zprávy)[3][1]
- extenzivní data ze sčítání lidu (dosažené vzdělání)
- účast ve volbách

Dva významné případy, kdy byla tato mimořádně cenná data ukradena, se týkaly United States Office of Personnel Management (US OPM) - úřadu spravující informace o federálních zaměstnancích Spojených Států -, a databáze Čínských uchazečů o zaměstnání.

V prvním případě unikly identity federálních zaměstnanců, včetně dat o otiscích prstů, a další kritické bezpečnostní informace. V druhém případě šlo o všechny osobní údaje a navíc citlivá data obsahující biometrické údaje a nebo například politické smýšlení.[19][39]

Vynucené stopy

Mezi vynucené stopy autor práce počítá všechna data o uživateli, nebo jeho zařízení, vyprodukovaná a publikovaná uživatelovým zařízením, ale bez jeho vědomí, na základě funkce software, jehož cílem je právě sběr dat.

Především komerční sektor využívá několik technik sběru dat, které autor řadí mezi zdroje vynucených stop - zejména se jedná o monitorování aktivity uživatele v rámci webu komerčního subjektu i mimo něj.

Historie prohlížení webu - pokračování

Informace o virtuálním pohybu návštěvníků webu, a zvláště o zdrojích, z nichž na web přicházejí, je užitečná pro cílení propagačních aktivit a optimalizaci obsahu. V kapitole "Historie prohlížení webu" byla popisována možnost sledování HTTP referer hlavičky; bylo nicméně podotknuto, že tato technika je nespolehlivá a proto jsou využívány jiné způsoby.

V jednodušších případech přidává jeden systém při generování odchozího hypertextového odkazu tzv. GET parametr požadavku, nebo jiným způsobem mění URL odkazu, na který uživatel klikne: z odkazů typu `https://example.com/?origin=x.com`, nebo `https://example.com/#src=123` lze zdroj návštěvy webu `example.com` poznat podle řetězce za otazníkem, či mřížkou – zmíněných parametrů požadavku –, který pro funkci odkazu ale jinak nemá žádnou praktickou funkci.

Vytvoření a čtení takového odkazu je nutné synchronizovat mezi odchozím a příchozím systémem, což je ale snadné, pokud oba ovládá jeden majitel: tedy například pokud se jedná o odkazy uvnitř jednoho webu, nebo když autor webu propaguje článek na více platformách, kam sám vkládá odkaz (tedy na Facebook přidá k odkazu řetězec `?src=facebook`, na Twitter `?src=twitter` a podobně).

V jiném případě systém, z něhož uživatel odchází, mezi uživatele a odchozí odkaz vloží přesměrovávající skript, který nejprve zaregistruje uživatelovu akci a poté ho teprve přesměruje (jsou to odkazy typu `https://example.com/redirect.php?target=x.com`). Tato akce je ovšem zaznamenána jenom na straně původce odkazu a pokud ten není stejný jako cíl odkazu, nebo pokud není URL během přesměrování pozměněna, velmi pravděpodobně slouží takové sledování aktivity jenom původci.[35]

Odkazy prvního typu jsou důležité například pro affiliate programy, zatímco druhého typu obvykle prostě sledují uživatelovu aktivitu (sociální sítě je využívají k monitorování aktivity a zájmů uživatelů).[35][2]

HTTP cookie tracking

Protože HTTP referer, ani další dva uvedené způsoby neprozrazují více, než jednu poslední adresu, a jsou podvrhnutelné, uchylují se jiné služby k dalším, komplexnějším způsobům sledování, obvykle za nějakého použití HTTP cookies.

HTTP cookie je jednoduchý mechanismus, který umožňuje webové aplikaci uložit část dat do uživatelova prohlížeče, který je poté odesílá s každým požadavkem. Nejdůležitější

důvod, proč se cookies využívají, je ten, že HTTP protokol je standardně bezstavový, což znamená, že aplikace musí na každý požadavek odpovědět individuálně a nezávisle, pouze na základě jeho URL a parametrů.

S tímto paradigmatem nelze při vyhodnocování požadavku brát v úvahu předchozí akce uživatele, mezi které ale může patřit i přihlášení do systému. To se obvykle provádí tak, že webová aplikace zaregistruje požadavek o přihlášení, autorizuje ho, uloží si záznam o aktuálním přihlášení a jeho ID uloží právě do cookie. Prohlížeč se poté s každým dalším požadavkem identifikuje tímto ID a aplikace může rozhodovat, jak na něj zareaguje i s ohledem na předchozí stav.

Mezi další legitimní použití cookies patří nastavení jednodušších osobních preferencí při používání webové aplikace i bez přihlášení na straně serveru, tedy například nastavení stylu zobrazování produktů v eshopu.

Nicméně: identifikace uživatele (byť anonymizovaná) znamená narušení jeho soukromí, pokud si jí není vědom. Aplikace může vytvořit cookie, kterou identifikuje návštěvu a prohlížeč ji poté bude odesílat s každým požadavkem, umožňujíc tak rekonstruování historie návštěv každé stránky webu, na které se objevuje skript, který tuto cookie zaregistruje, což typicky obsahuje všechny stránky.[2][1][5]

Za normálních okolností jsou cookies z bezpečnostních důvodů prohlížečem poskytovány pouze pro doménu, na které byly vytvořeny (prohlížeč vybírá odesílané cookies podle požadavku), a tak takové sledování může probíhat pouze v rámci jednoho webu, ovšem existují způsoby, jak tuto bariéru obejít a sledovat přihlašovací cookie napříč doménami a navíc návštěvy na webech třetích stran spojit s aktivním přihlášením (to jest návštěvy de-anonymizovat).[2][8] Viz přílohu A pro jednoduchý příklad.

Evercookie

Zvláštní pozornost v této kapitole musí dostat projekt Evercookie Amerického vývojáře Samyho Kamkara, vyvinutý v roce 2010. Protože je cookies standardně možné z paměti prohlížeče jednoduše vymazat, tato knihovna si vzala za cíl experimentálně ověřit způsoby, jakými lze jejich obsah "schovat" do nejrůznějších webových prvků a po jejich vymazání ho rekonstruovat. Tyto prvky jsou obvykle cache nejrůznějšího obsahu: jeden z asi sedmnácti způsobů spočívá ve vygenerování obrázku, jehož barvy lze dekódovat zpět na řetězec znaků, a který je jako obrázek vynuceně uložen do cache paměti.[40]

Evercookie se mimo jiné objevuje v prezentaci "Tor Stinks", která podle Edwarda Snowdena unikla z NSA, a ve které se diskutují možnosti identifikace uživatelů anonymizační sítě Tor.[41]

Sledovací kódy na webu

Sledovací kódy na webu navazují na záznamy aktivit, které byly zařazeny mezi pasivní stopy. Rozdíl od nich autor spatřuje ve způsobu jejich implementace a typu a využití shromažďovaných dat: data shromážděna tímto způsobem nejsou potřebná pro běh, bezpečnost, nebo údržbu systému, nýbrž pro obchodní potřeby.[1][5]

Tyto kódy zpravidla využívají HTTP cookies. Jak bylo zmíněno, existují techniky, kterými lze s pomocí cookies spojit data o prohlížení z různých domén, což představuje problém především při spojení všech těchto dat s aktivním přihlášením v jedné službě.

Sledovací kódy Google Analytics, využívajících cookies, se podle statistik analytického nástroje BuiltWith v době psaní práce nachází na 67.97% z milionu a 89.11% z deseti tisíc nejnavštěvovanějších webů. Widgety sociálních médií vkládané do webů (zejména ty od Facebooku) jsou také běžné, ačkoliv se nepodařilo sehnat souhrnnou statistiku, která by vyjadřovala počet webů, na nichž se nachází alespoň jeden z nich. Nicméně: jen samotný jeden nejpopulárnější widget, Facebook Sharer, se nachází na zhruba 6.49% z milionu nejnavštěvovanějších webů a na 19.42% z prvních deseti tisíc.[42][43]

Podle zprávy Ghostery Team, "The Tracker Tax" z roku 2018 obsahovalo nějaký sledovací kód téměř 90% načtených stránek - přes 20% z nich navíc obsahovalo takových kódů přes 50. Podobná zpráva z roku 2017 uvádí číslo 77,4%.[44][45]

Právě Facebook a Google mají - mimo úřadů a datových makléřů - pravděpodobně nejlepší osobní data o konkrétních uživatelích; ti navíc zůstávají při prohlížení webu k daným službám přihlášeni (tzn mají aktivní přihlašovací cookie) a skripty, kterými lze toto přihlášení sledovat a potenciálně spojit s aktuálně prohlíženou stránkou, nebo přesnou aktivitou na otevřené stránce, se nachází na zhruba 70+% webů.[2][1]

Sledování ze strany Facebooku se věnovala například práce From social media service to advertising network z roku 2015.[46]

Behaviorální data

V předchozí sekci byl kladen důraz spíše na rekonstrukci historie prohlížení, nicméně kromě toho je možné ve spolupráci s dalšími skripty sledovat i samotnou aktivitu na otevřené stránce (pohyby kurzoru, gesta, frekvence a pozice kliků, stisky kláves, rychlost psaní, kombinace kláves, rychlost čtení - scrollování, interakce s elementy...) a přiřadit jí k již sledované návštěvě.[1][2][8]

Tuto skutečnost dobře ilustrují dva příklady. První z nich, web https://clickclickclick.click,

je zaměřen na laickou veřejnost a poněkud odlehčenou formou komentuje, co všechno o uživateli zaznamenal. Druhým příkladem jsou analytické nástroje Hotjar, nebo Smartlook. Jejich služby umožňují nahrávat každý pohyb myši, kliknutí a scrollování a na základě těchto dat vytvářet "tepelné mapy" ilustrující nejčastější interakce návštěvníků s webem: dokonce je možné si každou návštěvu přehrát.

Sledování aktivity při používání produktu nepředstavuje etický problém, pokud s ním uživatel aktivně vyjádří souhlas (tzv. opt-in). Méně eticky přijatelně lze pohlížet na metodu implicitního souhlasu, kdy člověk musí aktivně vyjadřovat nesouhlas (tzv. opt-out) – tím horší je to případ, čím hlouběji v uživatelských nastaveních je možnost vyjádření nesouhlasu ukryta. Nejhorším narušením soukromí je potom sledování uživateli úplně utajené, s jehož prostředky se snaží bojovat například legislativa známá jako "EU Cookie Law", nebo nejnovější GDPR.[2]

Dosah tohoto typu sledování je dalekosáhlejší, než k čemu je nejčastěji využíváno. Behaviorální data v podstatě obsahují vše, co uživatel vidí, slyší, sleduje, čte; co píše, říká; a v důsledku zprostředkovaně i co si myslí. Behaviorální data jsou lepší pro predikování chování, než demografická data.[1][2]

Potenciální narušitelé soukromí

Množina subjektů, které mohou mít zájem na využívání osobních dat uživatelů je rozsáhlá. Nejužší zásah do soukromí mohou provádět jednotlivci, ale s velikostí organizace se rozšiřují možnosti a důsledky hromadného sledování.

Mezi prvními potenciálními narušiteli jsou tedy civilní individuálové - amatéři. Jejich hrozba je limitována nižšími technickými možnostmi, ale přesto mají určité možnosti poškození nechráněných cílů.[4] Příkladem budiž kyberšikana, které je vystavena většina dospívajících.[47]

Nebezpečnější individuálové jsou vycvičení profesionálové, operující v šedé zóně ekonomiky i na černém trhu. Mezi ně spadají hledači lidí a tajných informací ("skip tracers"), a profesionální kyberzločinci na volné noze. Tito mají navíc více motivace pro sdružování se v menších skupinách.[3][4][2] Obě tyto uvedené skupiny narušitelů jsou pro další potřeby práce uváděny společně jako "civilní sektor", i když v případě skupin typu organizovaného kyberzločinu by bylo možné uvažovat nad jeho zařazením do obchodního sektoru v rámci šedé zóny ekonomiky, nebo černého trhu.

Většími organizačními jednotkami potom zpravidla jsou legální podniky, a to buďto ty zabývající se vlastním podnikáním (majitelé dat), nebo ty, jejichž obchodním modelem je přímo shromažďování a využívání osobních dat (datoví makléři a reklamní společnosti). U obou hrozí zneužívání uživatelských dat pro monetární zisky[1], nebo neautorizovaný přístup zaměstnanců. Obě varianty se běžně objevují v tisku.[48][2][1]

Největšími možnosti disponuje vládní sektor, a to především z důvodu rozsahu shromažďovaných dat, a z podstaty věci nižšího dozoru veřejnosti. Kromě správců jednotlivých datových registrů mají k jejich datům dobrý přístup bezpečnostní složky (policie, a zejména rozvědka a kontrarozvědka).[1]

Literatura je plná příkladů překračování pravomocí a morálních hranic ze strany vládního sektoru.[1][2] V USA byly například zneužity data ze sčítání občanů při každém větším vojenském konfliktu (v první světové válce pro lokalizaci branců, ve druhé pro lokalizaci osob s Japonskými kořeny, při válce v Afganistanu pro sledování Arabských občanů) a v nacistickém Německu byla data sčítání využita pro označení židovských občanů.[2] Konečně nejlepší příkladem z moderní doby je skandál globálního sledování NSA, které bylo odhaleno a medializováno E. Snowdenem a dalšími. Další případy odhalily, že zatímco jsou whistlebloweři sledováni, vládní sektor sám informace tají a vynucuje mlčení, včetně mlčení o vynucování mlčení.[1]

[1]Pokud je služba dostupná zdarma, produktem jsou reklamní plochy a/nebo uživatelská data.[1]

Důvody pro využití dat

Podobně jako v případě rozčlenění příkladů digitálních stop, se členění této kapitoly nezakládá striktně na literatuře, ale slouží k sémantické dekompozici textu. Jednotlivé cíle jsou rozděleny podle hlediska charakteru (monetární, nebo nemonetární zisk) a podle subjektu (potenciálního narušitele), který je s tímto cílem obvykle asociován. Samozřejmě, pro potřeby cílů, které nabývají znaků více hledisek, je vyčleněna kapitola "smíšené cíle".

Důvody zde popisované zároveň označují hrozby pro uživatele.

Monetární cíle

Civilní sektor

Nejpřímějším způsobem, kterým mohou civilní útočníci sledovat monetární zisk, je zfalšování plateb, a to skrze krádež identity, nebo-li zosobnění ("impersonation"), to jest zcizení údajů za jejichž pomoci lze autorizovat podvodnou platbu. Zejména jde o využití čísel kreditních karet.[4][9][3][6][5]

Podobným způsobem je vytvoření dluhu ve jménu oběti za statky, které převzal útočník. Specifický příklad byl popisován ve Spojených Státech, kde byly zaznamenány případy, kdy útočníci využili krádež identity pro získání lékařského ošetření. Pro tento typ podvodu jsou zpravidla potřeba informace jako je jméno, datum narození a bydliště.[3]

Dalším prostředkem pro tento typ cíle je vydírání na základě digitálního obsahu: buďto se jedná o navrácení násilně smazaného obsahu (modus operandi ransomware), nebo o vyhrožování publikováním obsahu (například citlivých fotografií, nebo záznamů konverzací, případně ilegálního obsahu - v některých zemích například pornografie).[4][2]

Obchodní sektor

Data jsou obchodní potenciál a prostředek pro nové obchodní modely. Nové typy služeb postavených na informačních technologiích (na příklad AirBnB, nebo Uber) úspěšně získávají podíl na trhu na úkor tradičních ekvivalentů, zatímco i preexistující konkurenti využívají data z vlastních služeb pro optimalizaci marketingových aktivit a obchodních modelů.[1][2]

Pravděpodobně nejčastějším prostředkem obchodního využití osobních dat je profilování, segmentace a cílený marketing. Profilováním lze odhalit zájmy, nákupní zvyky i osobnostní rysy uživatele, segmentací se tito uživatelé seskupují do příbuzných skupin, na než je potom reklama cílená. Cílená reklama je přesnější, má lepší konverze (poměr mezi zobrazením reklamy a akcí, kterou propaguje) a je tudíž cennější, než reklamní plochy s obecným účelem. Proto je sledování uživatelů za účelem přizpůsobování reklamy výnosný byznys.[20][4][2][9][1][8]

Přizpůsobit reklamu je možné nejen cílením, ale i obsahem, čehož lze s úspěchem využít například v politické kampani.[1]

Kromě toho, díky sjednocení reklamních ploch napříč doménami (skrze cookies, nebo jiný způsob device fingerprinting) je možné dále využít možnosti segmentace s ovlivňování cílové skupiny, nebo jednotlivců kdekoliv na webu. Navíc je technicky možné v reálném čase cílit reklamu například na blízké obchody, a to na základě fyzické lokalizace zařízení.[2][1]

Na základě těchto příkladů za zmínku stojí, že příliš dobře cílená reklama má tendence uživatele spíše vyděsit a odradit a proto se marketéři musí uchylovat k jejich polidštění, což buďto obsahuje vysvětlení, proč byla tato reklama takto zacílena, nebo skrytí cílených reklam mezi úmyslně nerelevantní.[1]

Jinou možností, jak zvyšovat objem příjmů obchodu, je přizpůsobení cenové nabídky a nabídky služeb konkrétnímu uživateli podle jeho lokality, demografie, nebo jiných dat. Cílem těchto technik je dynamicky nabídnout zákazníkovi nejvyšší cenu, kterou je ochoten zaplatit, čehož lze dosáhnout buďto návrhem a propagováním vybrané služby z portfolia, nebo přímo uváděním jiné ceny. V některých případech bylo také zaznamenáno odpírání určitých služeb (typicky bankovních, nebo pojistných) na základě lokality zákazníka, tedy praxe známá pod termíny “redlining”, nebo přeneseně, v případě odmítnutí na základě digitální stopy, “weblining”.[2][1]

Datová analýza je také samostatným segmentem ekonomiky, stejně jako obchod s daty. Prodávání dat z vlastních služeb je možnost dostupná pro jejich provozovatele, ale také existuje možnost skupování dat z více zdrojů, jejich spojování a další prodávání těchto agregovaných, rozšířených datasetů.[2][8][1] Riziko obchodů s daty záleží na kupujícím. Byly zaznamenány případy, kdy byl prodány seznamy zranitelných osob (seniorů, osob v dluzích apod.) společnostem zabývajícími se nebankovními půjčkami.[1][2]

V literatuře se objevují další příklady: Pro uchazeče o zaměstnání je užitečné zmínit, že sledování digitální stopy má své místo v personalistice.[49][9][8]

Automatické rozpoznávání státních poznávacích značek vozidel je využíváno pro exekuce a proběhly pokusy se rozpoznáváním obličeje pro potřeby předcházení opakovaným krádežím v obchodech.[1]

Nakonec posledním uvedeným monetárním cílem využití dat obchodním sektorem je průmyslová špionáž.[1]

Nemonetární cíle

Civilní sektor

Prvním nemonetárním cílem, kterého dosahuje civilní sektor skrze užití dat je vlastní sledování a synchronizace a to skrze služby jako fitness náramky, synchronizované kalendáře, úkolníčky a podobně.[2][6]

Problém s tímto sledováním nastává v momentě, kdy jsou jeho data neautorizovaně zneužita ke kyberstalkingu (pronásledování za použití informačních technologií). Jeho pachatelem je buďto známý oběti (typicky partner), nebo neznámá osoba (v tom případě hrozí nebezpečí groomingu). V obou případech je motivace útočníka různá, ale může ústit ve fyzické napadení.[3][4][5]

Jiným zneužitím kyberprostoru je kyberšikana, která má za cíl zdiskreditovat oběť. V extrémních případech může dojít k nastrčení kompromitujících záznamů (ilegálního materiálu, nebo falešných komunikací).[8][47]

Dalším cílem může být vydírání pro obsah nemonetárního charakteru, typicky využití citlivých fotografií k donucení oběti k produkování dalšího podobného obsahu.[1]

Častým cílem civilních útočníků je také získání vlastní reputace a posílení vlastního ega. To je typicky motivace destruktivních útoků, nebo trollingu.[4]

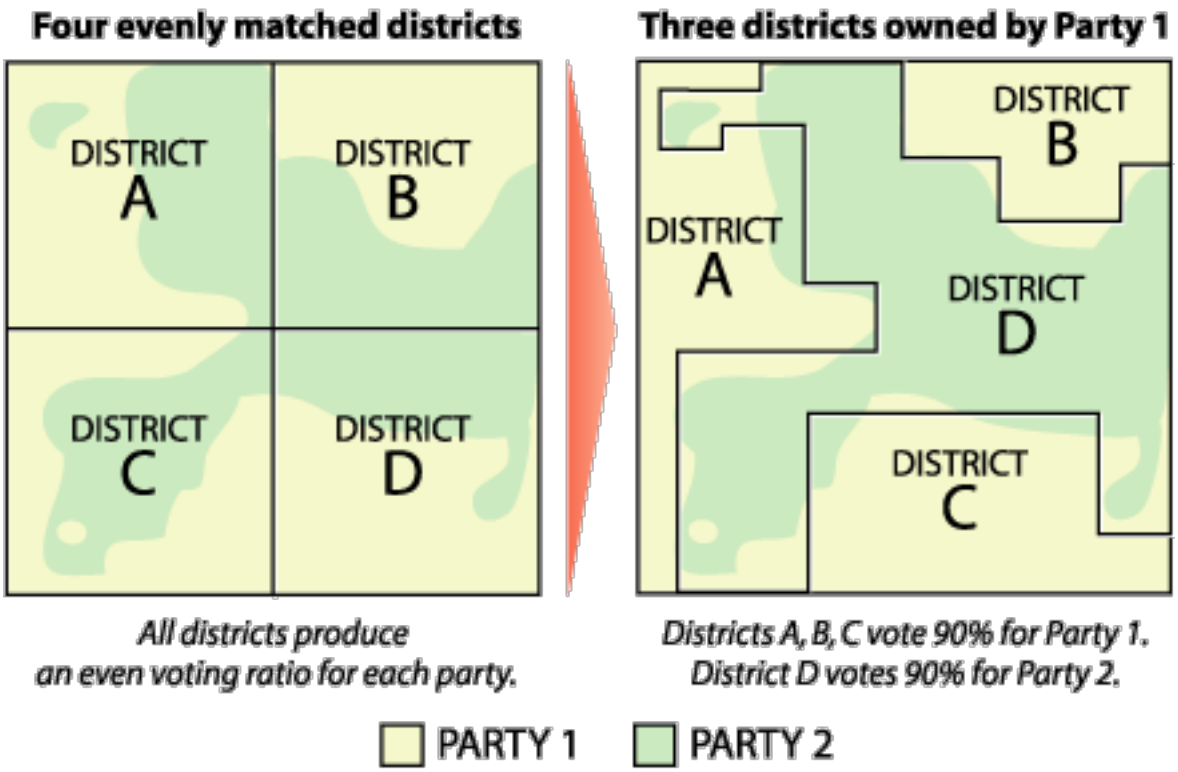
Vládní sektor

Ovlivnění voleb

Prvním způsobem, kterým vládní sektor s využitím dat sleduje nemonetární cíle, je stejně jako v případě obchodního sektoru cílení reklam. To se objevilo v tisku v návaznosti na kauzu Cambridge Analytica, ale faktem zůstává, že cílení reklam s úspěchem využívali i demokratičtí kandidáti a to ještě před touto kauzou.[1]

Druhý známý případ je praxe známá jako gerrymandering. Jedná se o způsob ovlivnění voleb skrze účelové změny volebních obvodů, který je obzvláště účinný ve většinovém volebním systému. Jeho princip ilustruje následující obrázek: hranice volebních obvodů jsou zkresleny tak, aby byla většina voličů jedné strany koncentrována do několika málo obvodů v nichž ona strana dosáhne jasného vítězství, a voliči strany druhé aby byli rozprostřeni do ostatních obvodů tak, aby zajistili méně přesvědčivé vítězství, ale zato ve více obvodech (proto tento systém funguje více ve většinovém volebním systému, ale méně v poměrném).

Tato technika se datuje až do roku 1812, ale stejné nástroje, které jsou využívané k cílení reklam, jí dodávají sílu.[1]



Obrázek 3.1: Gerrymandering: ovlivnění volebních výsledků skrze účelové překreslení volebních obvodů[50]

Prevence kriminality

Další oblastí zájmu vládního sektoru je bezpečnost.

V první řadě, data lze využívat v kriminálním vyšetřování a to buďto jako důkazní materiál, a nebo pro identifikaci a lokalizaci zločinců (na základě dat geolokace, nebo podle rozeznávání SPZ, či biometrických dat osob).[2][9][1][8]

I pokud se nejedná o kriminální činnost, lze s pomocí technologií sledovat páchání

přestupků: například nepovolené parkování, nebo absence STK podle rozeznané SPZ.

Stejné techniky sledování mohou mít své místo v prevenci, ačkoliv dopady na míru kriminality jsou stále nejasné. Podle některých studií kamerové sledování díky psychologickému efektu snižuje kriminalitu, podle jiných ne - stejně dobře jako kamery mohou fungovat například kvalitní osvětlení, nebo plakáty s motivy očí.[2][1]

Speciálním případem kriminality je terorismus. Zatímco v ostatních oblastech je využití je využití informačních technologií velice efektivní, prevence terorismu je jeden z úkolů, který datovou analýzou a umělou inteligencí řešit příliš dobře nelze. Na rozdíl od ostatních úloh vstupují do hry další faktory:

- akty terorismu jsou příliš vzácné a unikátní případy, mezi nimiž nelze s dostatečnou jistotou hledat společné prvky a na jejich základě vytvářet modely predikce
- v hromadném sběru dat nevadí, když data několika, nebo dokonce mnoha jednotlivců chybí, protože je možné pro tyto jednotlivce použít zobecněný model, navržený podle chování ostatních uživatelů. V prevenci terorismu jde právě o ty jednotlivce, kteří se sledování aktivně brání
- požadavek na absenci chyb je příliš vysoký; při cílení reklam nevadí chyby ani prvního a ni druhého typu ("false positive" a "false negative"), zatímco označení potenciálních teroristů je nutné prověřovat lidskými zdroji, které přemíra chyb prvního typu může zahltit a může odpoutat pozornost od osob unikajících díky chybě druhého typu[2][1]

Přes tyto problémy je to právě válka s terorismem, která je často používána pro ospravedlnění rozsahu kontrarozvědného sledování. V USA byl po útocích 11. září vydán ve stavu pohotovosti třiceti denní příkaz, který rozšiřoval pravomoci bezpečnostních složek, a ten byl v následujících letech prodlužován a rozšiřován.[1][2]

Kontrarozvědné sledování

Podmínky pro hromadné sledování občanů se zlepšují z několika důvodů; například:

- cena sledování klesá; úložná zařízení jsou výrazně levnější a samotná výpočetní i specializovaná sledovací technika zlevňuje a technologicky se vyvíjí
- data jsou dostupná: digitalizovaná a online
- všechna data kolují po stejné infrastruktuře, která využívá téměř stejnou technologii. Proto jsou špionážní řešení přenositelná[1]

Rozsah kontrarozvědného sledování je kritizován ochránci soukromí a to především z obav z jeho zneužívání.

V tomto ohledu je nejčastěji skloňována agentura NSA a to především díky odhalení E. Snowdena. Kromě toho mají NSA a další Americké tajné služby výraznou výhodu v tom, že největší technologické firmy sídlí v jejich jurisdikci a tudíž mají pravomoc vyžadovat jejich součinnost - vydání metadat a dat, nebo jinou výpomoc při vyšetřování. Navíc také velká část internetového provozu fyzicky prochází skrze území USA a tudíž skrze aktivní prvky pod kontrolou Amerických poskytovatelů internetu.[1]

Ve skutečnosti není důvod se domnívat, že by se kontrarozvědnému sledování nevěnovala jediná země s jakýmkoliv rozpočtem na informační bezpečnostní službu (rozpočet NSA se odhaduje v řádech desítek miliard dolarů). Podle dalších odhalení E. Snowdena NSA data sdílí s ostatními agenturami USA (CIA, FBI, DEA, DIA, DHS...; celkem 17, možná více) i se svými mezinárodními spojenci (hlavně anglicky mluvícími zeměmi a západní Evropou).[1]

Hlavním argumentem ochránců soukromí proti hromadnému sledování je ochrana principů demokracie a svobody; v případě nedemokratických systémů je zneužití sledování zcela reálné.[2][1]

Data vedou k člověku - k jeho fyzické lokaci. Toho může být zneužito k umlčování disidentů, zastrašování a perzekuci. Například v roce 2014 byli na Ukrajině telefonním operátorem označeni účastníci masových protestů, kteří následně obdrželi textovou zprávu s touto informací. [1][51]

Vytváření map známostí je další velice silný nástroj v rukou bezpečnostních složek. Co je na něm problematické je, že se osoba může stát terčem sledování jenom skrze známost s jinou osobou, která je sledována. Vzhledem k teorii, že všechny osoby jsou od sebe vzdálené maximálně na šest známostí - a bylo prokázáno, že na různých službách, jako Facebook, nebo Twitter je to výrazně méně - to znamená, že se do sledování dostává příliš velké množství osob.[52][2][1]

Další často zmiňovanou hrozbou je cenzura internetu, v souvislosti s čímž je nejčastěji popisován Čínský systém, přezdívaný "The Great Firewall of China".[53]

Dat jako důkazního materiálu lze také zneužít pro konstruování politických procesů.[1]

Z těchto důvodů ochránci soukromí argumentují, že kontrarozvědné sledování občanů musí být regulováno aktuálními, moderními zákony a přísně kontrolováno soudní mocí. Přesto se ale může pro omezení sledování nedostávat politické vůle, neboť prevence terorismu je silné téma a návrh redukce sledování může implikovat návrh redukce bezpečnosti, a to i navzdory tomu, že podstatně větší hrozbu, než terorismus, představuje organizovaný zločin.[54][1][55]

Kybernetická válka

Kybernetické divize se staly součástí moderních armád. Do jejich působnosti patří špionáž, kybernetické útoky a hybridní válka. Přirozeně, kvůli míře utajení lze pouze dohadovat původce, rozsah, nebo vůbec existenci útoků - existuje ovšem množství incidentů, které bezpečnostní experti přisuzují jednotlivým mocnostem.[2][1]

Jedním z příkladů byl malware Stuxnet, který podle všeho významně poškodil Íránský jaderný program. Usuzuje, se, že jeho původcem byly Izrael a USA.[56][1]

Další hrozby se přisuzují Číně. Aktuální je například kauza kolem firem Huawei a ZTE a otázky jejich napojení na Čínskou bezpečnost, a nebo obvinění výrobce serverů Supermicro z instalace infiltračního čipu.[57][58][59]

V České Republice byl také medializovaný spor o vydání J. Nikulina, který byl zadržen na území ČR na základě Amerického zatykače a o jeho vydání se přela Americká a Ruská diplomacie kvůli podezření, že se jedná o hackera ve službách Ruské Agentury pro výzkum internetu.[60]

Rusko je také často spojováno s metodami hybridní války, a to zejména šířením dezinformací. Cílení dezinformačních kampaní je další stinnou stránkou cílené reklamy - která byla také probírána na kauze Cambridge Analytica.[2] Novou zbraní dezinformátorů jsou podvržená videa manipulovaná za pomoci umělé inteligence, známá jako “deep fakes”.[61]

Smíšené cíle

Do následující kapitoly spadají cíle, které mohou buďto nabývat hlediska jak monetárního, tak nemonetárního, a nebo jsou vlastní více sektorům.

Útok pro získání prostředků

Prvním takovým případem je kybernetický útok, jehož cílem je zneužití dat získaných od jednotlivce pro přístup k dalším zdrojům. Typicky se jedná o útok na jednotlivce, jejichž přístupové údaje jsou využity pro přístup ke korporátním, nebo vládním systémům, nebo ovládnutí jejich zařízení a/nebo identity pro šíření spamu, nebo pro phishing.

Útok následující, provedený za pomoci dat získaných z prvního, potom může mít monetární i nemonetární cíle.[4]

Příkladem tohoto cíle mohou být zaznamenané útoky na bezpečnostní společnosti RSA Security (unikla data o zaměstnancích a autentizační tokeny) a Verisign (možná byly vydány falešné certifikáty a unikla zatím neznámá data).[19]

Kyberzločin jako služba

Stejně víceúčelový je segment nájemného kyberzločinu (Crime as a service). Do portfolia jeho aktivit patří provádění cílených útoků na zakázku (např. DDoS), vývoj a dodávání malware (např. ransomware), poskytování infrastruktury pro provádění útoků, praní špinavých bitcoinů, a nebo krádež, nebo naopak nastrčení kompromitujících záznamů - e-mailů a dalších komunikací, finančních a strategických dokumentů, grafických a video dat apod. - čímž je možné sledovat poškození pověsti, nebo uvěznění oběti.[62][63]

Manipulace obsahu

Manipulace obsahu je možná na stejném principu, jako cílení reklam. Je potřeba připomenout, že provozovatelé služeb mají k dispozici možnosti, jak skrytě propagovat, nebo naopak omezovat některé příspěvky (například umisťováním těchto příspěvků častěji a výše v uživatelských "feeds", nebo naopak), výsledky vyhledávání, nebo návrhy vyhledávání - v tzv. "našeptávačích".

Tyto možnosti mohou být využity pro sledování jakýchkoliv cílů, a nebo poskytnuty třetím stranám za úplatu.[1]

Spojení obchodního a vládního sektoru

Ochránci digitálního soukromí v souvislosti s kontrarozvědným sledováním upozorňují na propojení s obchodním sektorem.

V první řadě bezpečnostní složky mohou využívat korporátní data a to buďto dobrovolně vydaná (nebo koupená), případně vynuceně zpřístupněná (skrze soudní příkazy, hrozby, nebo v nejhorším případě skrze vlastní hacking). Velice nebezpečný je v tomto ohledu vliv na poskytovatele telekomunikačních služeb a internetu.[1]

Kromě toho stojí za zmínku, že technologie často přechází od tajných služeb do obchodního sektoru a poté do rukou nezávislým hackerům a naopak. Hackovací skupiny jako Gamma group dodávají sledovací software vládám: operují poměrně otevřeně (mají vlastní konferenci, ISS World) a jejich marketing se neliší od civilního. Neštítí se přitom dodávkám represivním režimům útočícím na disidenty a novináře. Mimochodem, Česká

Republika také figurovala na seznamu zákazníků takové organizace, a sice Hacking Team.[64][1][2]

Určitou roli hraje znovu i fakt, že se technologie standardizuje a je tedy snadno přenositelná. Filtrovací software, který má chránit lokální korporátní sítě tak lze využít k politické cenzuře na úrovni WAN sítí.[1]

Finanční cíle obchodního sektoru se potkávají s potřebami vládního dohledu.

Američtí autoři také uvádí případy, kdy datoví makléři obchodovali s veřejnými službami (jako kupující i jako prodávající).[2][1]

Funkce systému

Pro úplnost je vhodné doplnit, že každý systém potřebuje ukládat a využívat data nutná k jeho provozu - některá jsou nutná pro běh služby (lokace pro mobilní signál, číslo kreditní karty apod.), nebo pro identifikaci a autorizaci uživatele (jméno, e-mailová adresa apod.). Míra a způsob využití dat je silně individuální, ale vychází z principu služby.[1][2]

Vedlejší efekty

Využití dat pro popisované cíle s sebou nese řadu neplánovaných vedlejších účinků.

Poškození pověsti individuálů a organizací v důsledku úniku dat je jeden z nich.[4]

Skandál NSA poškodil trh i diplomacii USA (způsobil například přesun dat mimo území USA a pokles prodejů US technologií v důsledku ztráty důvěry v tamější firmy).[1]

Pro jednotlivce je potom nebezpečná ztráta soukromí a případné neplánované vyzrazení tajemství (např. sexuální orientace, nebo těhotenství), jehož důsledkem snadno může být emocionální i jiná újma. Byly zaznamenány případy, kdy takové tajemství prozradily cílené reklamy.[2][1][4]

Dalším důležitým důsledkem progresivní digitalizace je generování kompletní historie mladého člověka. Téměř všechna komunikace je přesunuta na média, kde uživatelé nekontrolují záznamy, což znamená permanentní záznam všeho. Všechno z mládí člověka se může kdykoliv snadno vrátit: Každý neuvážený příspěvek, všechny malé prohřešky. Pokud osoba zastává jakoukoliv veřejnou pozici (ať už pozici politickou, nebo status celebrity), takové informace, jakkoliv nevýznamné v době publikování, se vrací s účinkem devastujícím kariéru.[1][5]

Na to navazují psychologické důsledky sledování. Všudypřítomný dohled vytváří kulturu strachu, v níž jsou všichni podezřelí; strach z trestu, sledování samotné i jenom hrozba sledování odstraňuje důstojnost a podporuje psychické problémy, jako úzkost, nebo nízké sebevědomí. V jeho důsledku lidé mění, autocenzurují vlastní chování; přestávají sledovat a komentovat citlivá témata. Whistleblowing i společenská angažovanost občanů je proto v ohrožení.[2][1]

Častou paralelou k tomuto stavu je Panoptikon: návrh věznice, kde všichni vězni mohou být sledováni z jedné strážní věže, ale nemohou rozlišit, zda jsou aktuálně sledováni, nebo ne. To je nutí se chovat, jako by byli sledováni po celou dobu.[65]

Cílení reklamy a přizpůsobování obsahu může ovlivňovat uživatele ještě jedním nezmíněným způsobem: podporováním nezdravé zpětné vazby, označované termíny "filter bubble", "filter abuse", nebo "hall of mirrors". Ve zkratce jde o teorii, která varuje, že technologie, která se přizpůsobuje uživateli, mu může poskytovat jenom obsah, který potvrzuje jeho přesvědčení, i kdyby mělo jít o dezinformace.[2][1]

Závěrečnou myšlenkou je třeba poznamenat, že technologie i legislativa se mění, čímž je možné umožnit a legalizovat nové, neočekávané použití starých dat.[1]

Možnosti zneužití dat

Tato kapitola rozebírá prostředky, kterými mohou data sbírat provozovatelé, ukrást útočníci, a oba dva využít.

Zisk dat

Získání dat legitimním způsobem

Bylo již uváděno, že mnoho dat vzniká z principu služeb a jejich ukládání (byť dočasné) a zpracování je nutné pro jejich existenci (připomeňme metadata telefonních hovorů a digitální adresy). Kromě toho připomeňme, že shromažďování některých dat (opět metadata telefonních hovorů, a záznamy ubytovacích zařízení) ukládá provozovatelům zákon.[31][32][37]

Dalším způsobem sběru dat byly metody popisované v kapitolách "Záznamy aktivity", "Sledovací kódy na webu", "Historie prohlížení webu - pokračování": zabudované sledování aktivity, prováděné za účelem optimalizace systému, přizpůsobení obsahu, nebo cílení reklam.[1][5]

Legální metodou je také obchod s daty: služby mohou poskytovat kontrolovaný přístup k využití vlastních dat, nebo prodávat celé datasety. Prodejem a zpracováním datasetů se také věnují datoví makléři a v neposlední řadě přechází data do vlastnictví jiného subjektu při obchodních akvizicích.[2][1]

Získání dat nelegitimním způsobem

V oblasti elektronických útoků je rozeznáváno velké množství hrozeb; tato práce uvádí výčet těch relativně běžných.

Malware, exploits, backdoors

První oblastí je malware: červi, viry a trojští koně, zejména typu ransomware (který uživateli zašifruje data a vyžaduje platbu za jejich odemčení), nebo keylogger (který monitoruje stisky kláves s cílem zaregistrovat obsah konverzací, nebo přístupové údaje), či jiný druh spyware (jakýkoliv malware sledující uživatelovo zařízení), nebo adware (malware zobrazující nežádoucí reklamy).[66][4][6][7]

Aktivní malware má široké možnosti využití a také se objevily případy Malware-as-a-service, služby produkující zájemcům generický program z jehož výdělků získávají provize (HelloSpy, Satan).[1]

Jiný způsob narušení bezpečnosti představují zneužití slabých míst ochrany software ("vulnerability exploit"): kromě nově objevených zranitelností ("zero day"), odhalených při vyšetřování útoku, existuje také černý trh s návody na provedení těchto zneužití. Pokud navíc bezpečnostní složky hromadí know-how o zranitelnostech, které uchovávají pro vlastní užití, chyby nikdy nejsou opraveny a mohou být objeveny nepřítelem.[1][4]

Hledání zranitelností se věnují i bezpečnostní experti operující legálně a s etickými cíli: například organizace OWASP publikuje seznam nejčastějších zranitelností ve webových aplikacích a další experti se účastní Bug bounty programů (odměňování za hlášení technických chyb ve službách větších společností).[24][1]

Často popisované je úmyslné vytvoření zranitelností, a to pomocí backdoors ("zadních vrátek") zabudovaných v software i v hardware, nebo rootkits (malware, jehož cílem je vytvoření zranitelnosti).[1][66]

Ochránci soukromí upozorňují na lobbying bezpečnostních složek za legalizaci a pokusy o implementaci backdoors do komerčních produktů, jako například ve sporu FBI s Apple o prolomení zabezpečení iPhone teroristy ze San Bernardino, nebo v případě kauzy Huawei. Podle jejich argumentace, každá vyvinutá zbraň okamžitě znamená svojí vlastní zranitelnost; každý backdoor je vždy zneužitelný jinou stranou, než která ho implementovala.[2][1][67][57][58]

Hrozba backdoors a spyware je platná i pro oblast útoků typu man in the middle, spočívajících v odposlechu a manipulaci internetové komunikace na úrovni aktivních síťových prvků. Výsledky takových útoků lze využít mimo jiné ke krádeži identity a získání neautorizovaného přístupu do webových služeb, k nimž uživatel přistupoval skrze zasaženou síť. ("cookie hijacking").[3][5]

Brute force

Pro prolomení autorizace, nebo šifrování existuje několik variant útoku na bázi hrubé síly extrémního výpočetního výkonu. Jejich principem je zkoušení velkého množství klíčů, dokud dešifrování jedním z nich není úspěšné.[1]

Síla klíče (hesla) je určena především počtem kombinací, které lze z obsažených znaků vytvořit standardní kombinatorikou: pokud heslo o délce 6 znaků obsahuje pouze 26 znaků malých písmen anglické abecedy, existuje 26^6 = 308 milionu kombinací. Pokud

je to jenom o jeden znak více, jedná se již o 8 miliard kombinací, a pokud se jedná o 6 znaků obsahující malá a velká písmena, jde o téměř 20 miliard kombinací. Přidáním číslic se za stejné délky hesla číslo zvyšuje na bezmála 57 miliard, přidáním sady deseti speciálních znaků vznikne při šesti pozicích a 72 možných znacích 139 miliard kombinací. Přidáním dalších šesti znaků vzniká 19 triliard (19 * 10^21) kombinací.

Pro představu, projekt RC5-72 organizace Distributed.net, který pracuje na prolomení 72 bitové šifry hrubou silou za pomoci ohromného výkonu rozsáhlé decentralizované sítě počítačů, aktuálně generuje průměrně 1284 miliard klíčů za sekundu. S tímto výkonem by bylo poslední zmíněné heslo hrubou silou prolomeno nejpozději za 479 let. První, šestiznakové heslo obsahující malá písmena, by bylo prolomeno v řádu milisekund.[68][69]

Pokud jde o oblast kryptografie, pokud není klíč vyzrazen jinou cestou, tímto způsobem jsou možnosti prolomení šifrování vyčerpány.[1]

Pro hádání uživatelských hesel existují další techniky, které jsou založeny na dřívějších únicích databází obsahujících existující hesla. V první řadě, pokud je v jednom úniku zaregistrována kombinace e-mailové adresy a hesla, existuje šance, že je daný uživatel použil na jiných službách.[6][1]

Kromě toho jsou ta nejjednodušší hesla používána více uživateli a jiná slabá hesla lze prolomit slovníkovými útoky (“dictionary attack”).[6][7]

Pro ochranu hesel ze strany správce služby by měly být v databázi uloženy jenom jejich otisky (hash) a to navíc vytvořeny kryptograficky silným algoritmem (nikoliv md5, nebo SHA-1, které jsou příliš slabé) s použitím tzv. solení. Pro uživatele je důležité poznamenat, že pokud provozovatel služby dat dokáže uživateli zaslat zapomenuté heslo, znamená to, že je heslo u správce uloženo zcela nedostatečně.[2][1]

Úniky databází nejsou nijak vzácné a proto je důležité, aby provozovatel uplatňoval správné zásady ukládání hesel. Uniklá hesla hashována slabými algoritmy lze prolomit pomocí tzv. rainbow tables.[1]

Social engineering

Protože je matematická kryptografie dostatečně silná a zranitelnosti software skryté a nedostupné pro masové zneužití, dalšími častými prostředky jsou útoky na samotného uživatele, které mají za cíl jej přesvědčit obejít softwarovou ochranu. To jsou techniky nazývané social engineering.

Jejich principem je lhaní a zneužití důvěry, nebo neznalosti uživatelů. Nejtypičtějším

příkladem je phishing: navázání komunikace, kdy útočník předstírá, že je představitel služby, kterou uživatel může využívat (sociální média, banky, poskytovatel telekomunikací, úřady apod.) a vyžaduje od uživatele "ověření identity", a to zpravidla přesměrováním na falešný web (kopírující oficiální web), kde má uživatel zadat své osobní údaje (přístupové údaje, číslo kreditní karty apod.).[4][1][6][7]

Podobné metody lze postavit na krádeži identity a zosobnění uživatele pro zneužití jeho kontaktů (např. "drobnou přátelskou finanční výpomocí").[2]

Další podobnou metodou je scareware: malware zneužívající neznalost uživatele a strach z domnělé hrozby (vyvolané například sdělením, že uživatelův počítač je nakažen virem, kterým ve skutečnosti nakažen není[2]) s cílem přimět uživatele koupit produkt, který má domnělou hrozbu odstranit.[4]

Všechny uvedené metody lze vést jednak skrze elektronickou komunikaci (e-mail, instant messaging), tak také skrze telefon, nebo zcela osobně. V těchto případech útočník více odhaluje svou identitu, ale většina lidí je nečeká a ani zaměstnanci komerčních, či veřejných služeb, proti nimž může být stejný útok veden, proti němu není vyškolena.[3]

Dodatek

Občasně opomíjenou hrozbou je také offline útok: vloupání, krádež zařízení, průkazů, elektronických karet, či fyzických dokumentů, nebo jiné vynucení vydání dat, nebo přístupových údajů (např. na hranicích, nebo při vyšetřování).[3][6][2]

Nakonec lze pro cílené sledování osob využít podobných technik, které práce dále doporučuje jako metody kontroly vlastní digitální stopy (vyhledávače, digitální archivy apod.).[3][6]

Konečně, všechny tyto uvedené metody lze využít pro získání přístupu do informačních systémů a databází, z nichž lze neoprávněně získat data potřebná pro dále popisované metody zneužití velkých dat.

Metody využití

Způsoby zpracování dat jsou shodné a nezáleží na tom, jestli se data snaží zneužít majitel dat, nebo útočník, který je majiteli zcizil.

Prvotním využitím je jejich duplikace a záloha; data jsou subjekty, které zajímají,

[2]Ironicky, kromě scareware samotného.

ukládána na tak dlouho, jak je možné, nikoliv jak je potřeba: limitující faktor je jenom cena a velikost úložného prostoru. Nad databázemi je potom možné vyhledávat v kontextu (podle klíčových slov, zmínek o jménech apod.) a provádět další agregace. Především je ale možné šifrovaná data uložit do té doby, než bude k dispozici nástroj na jejich rozluštění.[1]

Agregace dat znamená jejich spojování a prolinkování, které lze provádět v podstatě dvěma způsoby: deterministicky, nebo korelací.

V prvním případě je potřeba mít v obou datasetech k dispozici unikátní identifikátor osoby, kterými jsou například: e-mailová adresa, číslo průkazu (zejm. občanského průkazu, nebo pasu), číslo zdravotního, nebo sociálního zabezpečení, číslo kreditní karty apod. Jednotlivá spojení je potom možné řetězit a shromáždit značné množství dat o konkrétních osobách. Tímto způsobem lze také spojovat pseudonymní data s offline identitou.[1]

Pokud nejsou k dispozici unikátní identifikátory, lze s určitou mírou spolehlivosti korelovat informace z více datasetů. Pokud se dva telefony pohybují po většinu času ve stejných místech, pravděpodobně patří stejnému člověku; pokud se v blízkosti dvou telefonních budek, z nichž byly uskutečněny hovory se stejným adresátem, v době, kdy byly hovory uskutečněny, nacházel nějaký mobilní telefon, byl to pravděpodobně jeho majitel, kdo telefonoval; pokud se k anonymnímu účtu přihlásil uživatel ze tří různých hotelů, kde byl ve stejnou dobu přihlášen, v množině osob, které se nacházely ve všech třech hotelech v udané časy, nezbude příliš prostoru pro pochybnosti.[1][2]

Na základě průniku dat vzniká jenom malá množina jedinců, čehož lze dobře využít pro deanonymizaci anonymizovaných dat. Pokud je v datasetu uvedeno, že osoba je muž, ve věku 18-26 let, obyvatel České Republiky a vysokoškolského vzdělání, všechny jednotlivé ukazatele označují velké množství osob, ale množina osob, které spadají do *všech* kategorií, bude nepoměrně užší. Výzkumníci v různých studiích byli na základě korelace schopni identifikovat až 97% anonymizovaných dat.[2]

Další škálu využití obsahují statistické metody. Z nich je velice důležitá segmentace, již zmiňovaná v kontextu cílení reklam.[4][1]

Škola umělé inteligence a strojového učení přináší další možnosti zpracování nedeterministických dat, zejména v oblasti vytváření vzorců. Na základě sledovaného obsahu na internetu, nebo dokonce obsahu chladničky je možné, po naučení algoritmu na základě ukázkových dat, s určitou pravděpodobností uživatele zařadit do kategorií podle etnicity, náboženství, ideologie, sexuality, rodinného stavu, věku, pohlaví, zdravotnické diagnózy, nebo osobnostních typů. Také je možné takto identifikovat osoby na základě jejich

behaviorálních dat.[1][20][70][71]

Díky těmto možnostem práce se stochastickým obsahem je možné rozeznávat a kategorizovat obsah grafických a video formátů: do toho spadá například optické rozeznávání znaků (tj. digitalizace textu), automatický přepis hlasového záznamu, rozeznávání SPZ a rozeznávání biometrických dat, tj. tvaru obličeje, způsobu pohybu (chůze). To vše lze provádět v reálném čase na základě obrazu z bezpečnostních kamer.[1][2][5][3][70][71]

Data mining a umělá inteligence mají velký potenciál a ještě nejsou zcela probádanými oblastmi; dostupná výpočetní síla roste a bude umožňovat jejich častější a rozsáhlejší využití.[1]

Nezmiňovány také zůstaly všechny možnosti manuálního využití informací.

Ochrana

V následující kapitole jsou rozebrány možnosti, které mají uživatelé k dispozici pro odstranění vlastní digitální stopy a ochrany digitálního soukromí.

Kontrola

Předpokladem pro úspěšnou ochranu soukromí je vytvoření modelu hrozeb (tj. identifikace možností zneužití dat, které jsou relevantní k uživateli) a analýza aktuálního stavu (tj. označení již existujících a přístupných dat).[2]

Nejlepší možnosti kontroly představují rozhraní služeb, které umožňují vyzvednutí archivu všech uživatelských dat: zejména jde o velké společnosti, jako Facebook, Google, Apple, Microsoft.[3][3][2]

Pokud služba, u níž klient vlastní účet, takové rozhraní neposkytuje, je možné se na jejich představitele obrátit s žádostí na základě státní, nebo mezinárodní legislativy o ochraně osobních údajů a soukromí - především GDPR. Obyvatelé západních zemí mají v tomto ohledu zpravidla dostačující možnosti.[2]

Oba způsoby vyžadují, aby si uživatel pamatoval, kde svá data publikoval, což není vždy možné. Naštěstí, vyhledávače nezapomínají. Vyhledávání vlastní osoby (zejm. podle jména, přezdívek, e-mailu, telefonu, zaměstnavatele, školy, státní příslušnosti a místa bydliště) obecnými internetovými vyhledávači (Google, Yahoo, Bing, Altavista, DuckDuckGo, Seznam, Centrum) je doporučováno už jen protože je to technicky nejméně náročný způsob kontroly vlastní stopy.[4][3][5][8]

Za zmínku stojí služba Google Alerts, která umožňuje uživateli nastavit upozornění pro případ, že je indexována nová URL obsahující zadaná klíčová slova.[4][3]

Podobným, ale specializovaným případem jsou zaměřené vyhledávače, například vyhledávače osob (Intelius, Zabasearch, Pipl), sociální média, vyhledávače telefonních čísel, nebo digitalizované rodokmeny (Ancestry). Významným příkladem je web `haveibeenpwned.com` provozovaný bezpečnostním expertem T. Huntem, který shromažďuje data uniklá při nejrůznějších útocích a umožňuje uživatelům ověřit, zda jejich e-mailová adresa mezi těmito kompromitovanými daty figuruje.[4][9][3][5][6]

[3]Autor, který si prováděl audity v únoru a následně v listopadu 2018, může potvrdit, že se u velkých poskytovatelů možnosti zobrazení a úpravy osobních dat podstatně zlepšily po implementaci GDPR v polovině roku 2018

Poslední možností je nákup informací od datových makléřů (Epsilon, Acxiom, Datalogix, Lexisnexis, TLO, Westlaw, Intelius).[2]

Možnosti ochrany

Následující kapitola si klade za cíl shrnout doporučované techniky odstranění a prevence digitální stopy. Je ale na místě znovu připomenout, že žádná opatření nemohou zaručit dokonalou ochranu. Například, proti pasivní stopě chránit nelze, neboť je často předpokladem pro fungování služby.[8]

Obecné zásady

Na základě rešerše literatury byly identifikovány určité obecné zásady, které je vhodné mít na paměti při uplatňování aktivní ochrany. Na základě těchto doporučení jsou popisovány další metody:

- Bezpečí a soukromí obvykle nejsou dostupné bez vkladu - peněžního, časového, nebo ve formě změny chování[2]
- Metody ochrany nesmí neporušovat zákon, ani být společensky nepřijatelné[2][3][1]
- Metody také nesmí dramaticky ovlivňovat život uživatele[2]
- Nesmí ani klást nároky na ostatní osoby - ani technické, nebo časové (obzvláště netechničtí uživatelé budou bezpečnostní opatření maximálně obcházet pro zjednodušení dosažení cíle), ani finanční[2]
- Nejlépe funguje ochrana na pozadí (už jen protože hrozí menší šance chyby uživatele)[1]
- Technologie musí být dostupná (většina uživatelů si nemůže vyvíjet vlastní software)[2]
- Opatření nesmí budit přílišné podezření. Tajemství plodí zvědavost[3][2]

Opt-out

Prvním krokem, který může uživatel učinit je apel na provozovatele služeb. Ti zpravidla umožňují jejich klientům oznámit jim, že o některé služby - zpravidla o odesílání reklamních sdělení, personalizaci reklam a obsahu, odesílání anonymních dat pro vyhodnocování chování software apod. - nemají zájem.

Toto oznámení může probíhat v režimu opt-in (vyjádření zájmu, přihlášení), nebo opt-out (vyjádření nezájmu, odhlášení). Především kvůli druhému způsobu je potřeba

projít všechna nastavení a odhlásit všechny druhy sledování, shromažďování dat, nebo přizpůsobení.[4][1][8]

To se týká všech zařízení (na úrovni operačního systému i software) a služeb. Jednotlivým návodům pro konkrétní produkty se hojně věnují internetové články, které laickým uživatelům s nastavením pomohou.

Některé služby (zejm. sociální média), nebo software (operační systém) také umožňují jiná nastavení soukromí (viditelnost příspěvků, vyhledatelnost profilu), která je vhodné kriticky zhodnotit.[4]

Fyzické zabezpečení

Správně provedené fyzické zabezpečení dat předchází škále offline útoků, které mohou získat přesah do kybersvěta. Typicky se jedná o instalaci odposlouchávacích zařízení, krádeže identity a podobně.

Silnou ochranou vlastních zařízení a sítě před neautorizovaným přístupem jsou běžné metody domácí bezpečnosti: bezpečnostní dveře, okna a zámky, domácí alarm, kamerová ochrana a trezor.[3][6] Kromě toho je doporučována silná poštovní schránka, či P.O. Box, nebo privátní alternativy této služby.[3]

Pro ochranu před odcizením mobilních zařízení byl navržen a jako de-facto standard úspěšně implementován zámek Kensington Security Slot (často označován jako Kensington lock).[4]

V neposlední řadě je v otázce fyzické ochrany dat zmiňována likvidace papírových dokumentů způsobem, který znemožňuje jejich rekonstrukci.[2][3]

Hardware ochrana

Kromě fyzického zabezpečení doporučují autoři několik dalších doplňků, které chrání únikům dat na úrovni hardware.

Produkty dohledatelné pod názvy “privátní filtr pro snížení úhlu pohledu”, nebo “viewing angle screen protector” vypadají jako folie, která se připevní na displej zařízení a snižuje pozorovací úhel, čímž předchází odpozorování informací okolostojícími (tzv. “shoulder surfing”).[4][2]

Pro nabíjení mobilních zařízení z neznámých USB slotů je doporučován USB adaptér bez datových pinů, nebo redukce z vysokonapěťové sítě.[4]

Pro ochranu před bezdrátovým signálem je využíván efekt Faradayovy klece: i když stejného efektu dosáhne obyčejná hliníková fólie, na trhu existují poněkud elegantnější varianty, známé například pod názvem Faraday bag.[2]

Nakonec i tak jednoduché opatření, jako je samolepka na kameře zařízení může plnit svůj účel ochrany před jejím neautorizovaným aktivováním.[2][1]

Software ochrana

Více možností existuje v oblasti software.

První oblast, na které se odborníci shodují, je zakódování zařízení a to jednak z hlediska přístupu do operačního systému (ten je obvykle zabezpečen přístupovým heslem, nebo biometrickým skenem), tak z hlediska šifrování pevného disku (to provádí operační systém, nebo software na základě předchozí autentizace). Pokud je disk šifrován, není možné jej fyzicky vyjmout ze zamčeného zařízení a přečíst externě.[4][2][1][6]

Stejným způsobem lze šifrovat USB disky a existují disky, které takové šifrování poskytují nezávisle na operačním systému.[4]

Do aktivní softwarové ochrany patří zejména antivirus a to nějaký od respektovaných výrobců: neznámý software nemá žádnou garanci úspěšnosti.[72][4][1][6]

Dalším krokem je ochrana webového prohlížeče. V této oblasti se doporučuje několik rozšíření:

- Aktivní detekce a blokování sledovacích kódů (Ghostery, Disconnect)[2][9][1][8]
- Blokování reklamy (AdBlock Plus)[4][8]
- Vyžadování HTTPS spojení, kdykoliv je možné (HTTPS Everywhere)[2][1]

Pro ochranu digitálních komunikací vzniklo několik technologií, které si kladou za cíl ochranu před jejich odposlechem.

V první řadě autoři doporučují - zejména při přístupu k internetu z nezabezpečených veřejných sítí - využívat virtuální privátní síť (VPN), která všechny síťové požadavky tuneluje skrze šifrované spojení do jiného zařízení, které tyto požadavky přeposílá a vrací odpověď. Tím se omezují možnosti odposlechu na neznámé aktivní infrastruktuře sítě a kromě toho je trochu lépe chráněna IP adresa a lokalita původního zařízení.[4][6][8]

Další úrovní jsou anonymizující služby, z nichž nejznámější je Tor, který metody VPN uplatňuje v širším měřítku pro obfuskaci a decentralizaci komunikace. Síť Tor tvoří řada nezávislých serverů, které si vyměňují protékající komunikaci, a to v několika

krocích proto, aby bylo velmi těžké identifikovat původního odesílatele konkrétních požadavků.[9][2][1][8]

Pro vyjasnění případných nejasností: "anonymní" režim prohlížečů po zavření odstraní dočasná data a nevytváří lokální historii vyhledávání, přihlašování a prohlížení, nijak ale neanonymizuje samotného uživatele. Právě tím je oproti službám typu Tor nedostatečným opatřením.[8]

Pro šifrování samotných zpráv také existují alternativní komunikační nástroje:

- pro instant messaging je často představován Off-the-Record Messaging (OTR), jako nadstavba nad Jabber[2][1]
- pro e-maily existuje šifrování pomocí šifrovacího programu PGP. Alternativně, některé služby, jako Protonmail, nabízejí implementaci end-to-end šifrování ve svých produktech[2][1]
- pro hlasový přenos existují služby jako například TorFone, nebo Silent Circle[1]

Stejně jako v oblasti odhlášení se z nepotřebných služeb je vhodné si provést úklid v software a uložených datech. Odstraněním nepotřebného SW a aplikací se snižuje šance využití chyb v nich se potenciálně vyskytujících. Stejně tak je potřeba odstranit všechna nepotřebná data, která pro uživatele nemají užitek, ale mohou představovat zcela neočekávané riziko. Pro odstranění těchto dat se doporučuje SW typu Ccleaner.[4][7][9]

Nakonec se autoři shodují na potřebě pravidelně aktualizovat užívaný software (v případě SW, který není autorem již podporován, hledat alternativy) a zálohovat data na bezpečném místě mimo zařízení (přinejmenším v šifrovaném cloudovém prostředí) a to ideálně offline a redundantně (dva externí disky zamčené na dvou různých lokacích). Zálohy v online prostředí je možné provádět automatizovaně, bez významného přičinění uživatele, zatímco druhá zmiňovaná možnost uložení představuje lepší ochranu za cenu nižšího komfortu.[4][1][6][7][2]

Některé z těchto metod softwarové ochrany soukromí poskytují integrovaná řešení, jako například Tor Browser, nebo Epic Browser jako internetové prohlížeče, nebo Tails jako linuxový operační systém.[2][1]

Vlastní chování

Nejdůležitějším aspektem ochrany soukromí je vlastní chování. Nejlepší technologická opatření nemohou zabezpečit uživatele, který se chová nezodpovědně a to z důvodů, které byly během práce několikrát popisovány. Proto je důležitá alespoň určitá úroveň technologické gramotnosti uživatelů.[1][3][8]

Zodpovědné chování

Cílem zodpovědné interakce s digitálním prostředím je minimalizace rozsahu a obsahu publikovaných vlastních dat. Kontrolu aktivní stopy může v první řadě provádět sám uživatel.

V první řadě do této kategorie spadá určitá zdravá dávka skepticismu a kritického myšlení, které jsou nezbytné pro detekci phishingu a jiných pokusů o zveřejnění uživatelových údajů. Je tedy doporučované velice kriticky sledovat zprávy, které člověk obdrží neočekávaně, a to obzvláště, pokud obsahují hypertextové odkazy a žádosti o autorizaci kontaktů, zkoumat možnosti nastavení služeb, množství a typ oprávnění, která instalované aplikace vyžadují (v krajním případě i číst licenční ujednání), vyhodnocovat autenticitu obsahu, který je návštěvníkům předkládán, a věrohodnost zdroje (např. díky správné URL a platnému HTTPS certifikátu a v případě telefonních hovorů podle telefonního čísla volajícího).[4][3][8][6][1]

Za druhé je potřeba připomínat, že žádný publikovaný obsah nelze bezpečně "odpublikovat". Proto je potřeba zmínit, že kritické myšlení má své opodstatnění nejen při přijímání informací, ale také při jejich sdílení: je nutné zvažovat obsah komunikace (nepublikovat kompromitující obsah, nezávisle na bezpečnostních opatřeních) a její dosah (okruh osob, jimž je informace směřována). Také se nedoporučuje nikde (ani v prohlížeči a tím méně u služeb) ukládat citlivá data, jako jsou hesla, nebo čísla platebních karet.[4][3][8][1]

Tématu účinků sociálních médií (a potenciální závislosti na nich) se seriózně věnuje psychologie; z pozice této práce je ovšem doporučeno opustit nepotřebné služby (zejm. sociální sítě) a/nebo limitovat objem informací na nich publikovaných a přesunout i soukromé konverzace mimo možnosti digitální záznamu.[2][1][3]

V případě soukromé komunikace se v literatuře objevuje informace, že komunikace probíhající v reálném čase (tj. verbální komunikace, přenos zvuku, videohovor) je legislativně složitější a také technologicky náročnější odposlouchávat: analýza hlasového záznamu je možná, ale výpočetně a datově složitější.[2]

Z důvodu popisovaného sledování napříč doménami je také doporučováno oddělení přenosu citlivých informací od běžného elektronického provozu, tj. přinejmenším využívat dva prohlížeče: jeden pro běžné procházení internetu, a druhý pro komunikace, bankovnictví apod.[4][8]

Nakonec je potřeba se staré výpočetní techniky zbavovat stejně obezřetně, jako zmiňovaných fyzických dokumentů: navrácením do továrního nastavení a formátováním, vyjmutím a fyzickou likvidací disku.[3]

Uživatel by měl čas od času udělat audit svojí bezpečnosti a případně přijmout nová opatření.[4]

Autorizace

Silná politika hesel je často doporučována, často nepochopena a často ignorována. Velká část uživatelů stále používá slabá hesla, nebo je znovu-používají.

Síla hesla je důležitou ochranou proti útokům hrubou silou. Vzhledem k jejich nízké efektivitě je ale častěji přistupováno ke slovníkovým útokům, nebo ke zkoušení již známých, uniklých hesel, jak bylo popisováno v kapitole "Získání dat nelegitimním způsobem".

Vzhledem k tomu, že úniky osobních informací, včetně párů e-mailových adres a nešifrovaných hesel, jsou časté a nebezpečí ztráty digitální identity je tudíž reálné, je nutné používat naprosto unikátní a náhodná hesla, a to proto aby nebylo možné data z jednoho úniku použít pro kompromitaci dalších účtů.[4][3][7][2][8]

Unikátnost hesel je tedy důležitější, než jejich samotná síla. Potřeba pamatovat si více unikátních hesel může vést uživatele ke tvorbě pseudo-unikátních hesel: `heslo1`, `heslo2`, `heslogmail`, `hesloseznam` apod., pročež je nutné v radách pro vytváření hesel chovat na paměti uživatelskou přívětivost.[1] Literatura popisuje nejméně tři způsoby, jak ukládat velké množství skutečně unikátních a silných hesel:

- Vytvořením mnemotechnických pomůcek (frází, jejichž zkratka tvoří heslo)[7][2]
- Využíváním správců hesel (1Password, LastPass apod.)[2]
- Bezpečným ukládáním hesel (ve formě střeženého poznámkového bloku, nebo USB klíčenky)[2][1]

Stejný problém pseudo-unikátních hesel představuje i politika pravidelných změn hesel. Pokud jsou hesla unikátní a silná, experti se kloní k názoru od této praxe upustit.[2][1]

Kde je k dispozici ta možnost, není na škodu pro přihlašování do služby nastavit vícefaktorové ověřování.[4][2][6]

Kromě toho je také nutné pamatovat na výchozí hesla nastavená pro některá zařízení, jak například IoT, nebo aktivní síťové prvky (routery).[4][6]

Výběr kvalitních dodavatelů a partnerů

V dosavadní rešerši bylo poznamenáno, že pečlivý výběr partnerů a dodavatelů technologií je dobrým opatřením pro minimalizaci šancí na vyzrazení uchovávaných údajů.

V případě internetových nákupů je vhodné preferovat certifikované obchodníky, pyšnící se například certifikací ISO 27001, PCI DSS a dalšími.[4]

Otevřenost partnerů je dalším indikátorem věrohodnosti: ti, kteří umožňují zobrazit a mazat nebo měnit uživatelova data jsou pokládáni za důvěryhodnější.[2]

Také existují alternativní nástroje, nebo komunitně vyvíjené, nebo s jinými obchodním modely, ohleduplnějšími k soukromí.[2] Jako příklady slouží vyhledávač DuckDuckGo[2][5], privátní e-mailové služby jako Protonmail, Countermail, Neomail apod.[5][2][1], nebo decentralizovaná sociální síť Mastodon.[73]

Některé soukromé e-mailové služby (jmenovitě Lavabit a Silent mail) ukončení svého podnikání zdůvodňovaly tím, že na ně byl vyvíjen nátlak pro sdílení dat se bezpečnostními složkami.[2]

Výběr partnerů podle původu a jurisdikce jejich sídla je také myšlenka hodná prozkoumání. Jedná se o vhodnou úlohu pro model vícekriteriálního rozhodování, do nějž mohou jako parametry vstupovat žebříčky, které hodnotí úroveň osobní svobody (potažmo ekonomické svobody, svobody tisku apod.), délka formální neutrality země, nebo prezence v mezinárodních vojenských paktech. Bez formálního ověřování, které autor pro vyvozování oficiálních důsledků navrhuje, je zde pouze naznačeno, že ve zmíněných žebříčcích se na vysokých příčkách často umisťují země jako Švýcarsko, Norsko, nebo Island.[1]

Chaotické chování

Jiným oborem technik změny chování je chování s cílem decentralizace a obfuskace dat. Popisované techniky, byť poskytují novou úroveň ochrany, mohou sice být náročnější na provedení, ale ještě stále nespadají mezi ty, které by nebyly dostupné veřejnosti, nebo které by výrazně ovlivňovaly život. Jejich zásadní myšlenkou je určitá nepředvídatelnost a stochasticita.[3]

Centralizace dat znamená big data, jejichž analýzu je potřeba ztížit fragmentací. Na decentralizaci je možné pohlížet ze dvou pohledů: buďto jsou data jednoho uživatele rozprostřena mezi více služeb, nebo jsou data jedné služby rozprostřena mezi vlastní uživatele (resp. jejich zařízení). Protože je druhá varianta technologicky náročnější a méně častá, operuje dále práce s prvním případem.[1][2]

Pozitivní je, že je digitální stopa obvykle přirozeně roztroušena - některé údaje má prohlížeč, některé vyhledávač (potažmo jakýkoliv koncový server), některé ISP - a udělat si celkový obraz není zcela jednoduché, i když prakticky možné. To, že tyto segmenty

zůstanou decentralizované, je nelepší podmínkou pro to, aby byli uživatelé v bezpečí před zneužitím svých stop.[1]

Chaotické chování je nelepším aktivním způsobem, jak chránit vlastní reálnou identitu. Generování nových (náhodných) uživatelských jmen, nebo dokonce e-mailových adres pro každou internetovou službu, která nepotřebuje znát skutečnou identitu (všechna fóra, registrace pro přístup k uzavřenému obsahu apod.), se může zdát náročné, ale účinně znemožňuje spojování jednotlivých dat v jeden obraz.[2]

Menší úroveň roztříštění aktivní stopy je oddělení profesionálního, oficiálního a soukromého, osobního digitálního života a to především skrze užívání více e-mailů, uživatelských jmen a hesel. V případě úniku identity v soukromém digitálním životě potom existuje šance, že kritické části oficiálního nebudou poškozeny.[2][8]

Někteří autoři doporučují používání unikátních e-mailových adres pro každou jednotlivou službu, což dává smysl, neboť e-mail je de facto unikátní identifikátor a umožňuje deterministické, strojové spojování datasetů[4], nicméně může to být velmi časově náročné. Jistým kompromisem je vytvoření několika tříd adres: kromě zmiňované kombinace profesionální a soukromé adresy, je možné udržovat například také adresy vyhrazené pro pracovní účely, vyřizování objednávek z internetových obchodů, newslettery a obecný nedůležitý spam.[8][3]

Další kategorií metod je účelová obfuskace (zkreslení) dat. Zatímco někteří autoři vyjadřují určitou počáteční nevoli k uvádění nepravdivých informací, jiní v podstatě konstatují, že "v existujících datech jsou už tak často chyby, a proto není na škodu je podporovat" a že některé údaje ve skutečnosti není potřeba uvádět, i když je služby vyžadují.[2][3][1]

Do metod obfuskace tak spadají falešné identity a pseudonymy a částečné dezinformace. Je nutné poznamenat, že některé způsoby vytváření falešných identit jsou ilegální (typicky padělání dokladů), jiné nikoliv.[2][3][5][1]

Je například zcela v pořádku tvrdit, že jméno uživatelova prvního domácího mazlíčka bylo `wAgSy65g!inH7Ert`: odpovědi na bezpečnostní otázky nemusí nikdy obsahovat skutečné údaje, které by byly za normálních okolností lépe odhadnutelné potenciálním útočníkem, než hesla.

Také, při internetové objednávce s odběrem na pobočce a platbou v hotovosti je jméno jediný údaj, který je pro transakci relevantní a to navíc pouze v případě, že obsluha

[4]E-mailové adresy představují pravděpodobně nejlepší způsob pro sjednocování více datových souborů i vzhledem k tomu, že často nelze uvádět fiktivní, neboť se na ně typicky odesílají potvrzující e-maily s každou registrací

vyžaduje prokázání identity.[5][2]

Oporou při generování náhodných identit jsou dostupné internetové generátory náhodných údajů, včetně dočasných e-mailových adres a telefonních čísel.[2][3]

Skutečné jméno, nebo adresu je tak nutné používat pouze v případech, kdy je na něm služba skutečně závislá (doručení zboží na fakturu, komunikace s úřady).[3]

Pro roztroušení finančních údajů a dat nákupních zvyků je také zcela legální nechat si k bankovnímu účtu vystavit více platebních karet, zřídit více bankovních účtů a nebo pro většinu menších transakcí vybírat peníze z bankomatů a platit hotovostí.[2][1][6][8]

Kolektivní imunita

Ostatní uživatelé

Především kvůli stopám vytvořeným třetími stranami a útokům vedeným k získání prostředků k pokračování jiných útoků je potřeba podobná opatření šířit mezi širší okruh uživatelů. Uživatel, který tuto problematiku bere vážně zlepšuje nejen svou ochranu, ale snižuje i šance na to, že by byly jeho údaje využity k poškození blízkých, známých, zaměstnavatele apod.

Prvním aspektem kolektivní ochrany je tedy přesvědčení ostatních uživatelů k dodržování podobných pravidel. Jejich data pronikají do uživatelova soukromí a mohou ho ovlivnit a proto uživatel sám má právo ochranu svého soukromí vyžadovat.[4][2]

Tento bod se týká především přátel a rodiny: zvláště pak dětí. Ochrana digitálního soukromí není zábavná, nebo srozumitelná často ani pro rodiče, natož pro děti; děti navíc chtějí sdílet obsah, napodobovat influencery a tvořit. Rodiče by je neměli omezovat, nebo plošně sledovat ze stejných důvodů, kterými bylo argumentováno proti vládnímu sledování, ale přesto je nutné digitální gramotnost a zásady bezpečnosti a soukromí zahrnout do výchovy i výuky.[2]

Kolektivní efekt

Dalším aspektem kolektivní imunity je fakt, že hromadné užívání určitých technologií a metod zlepšuje globální ochranu internetu.

Například, kdokoliv, kdo používá šifrovací a anonymizující technologii pro neškodné účely, vytváří kouřovou clonu pro menšinu, která je potřebuje použít nutně (disidenti a whistlebloweři). Pokud šifrování používají jenom osoby, které skutečně chtějí něco

skrývat, je to zaznamenáno a vztyčuje to nad uživatelem rudou značku; pokud je používají všichni, samotné používání šifrování na podezřelé chování neupozorňuje, nebo není v silách potenciálně represivního kontrolora kontrolovat všechny osoby.[1]

Čím více lidí se chrání, tím náročnější je hromadné sledování všech; masové šifrování komunikací by navýšilo náklady na hromadné sledování astronomicky a přimělo sledující subjekty hledat alternativní metody, což je vítězstvím pro všechny sledované. Stejně tak, pokud se uživatelé naučí skrývat data, efektivita cílení reklam a přizpůsobování obsahu bude klesat, což může vést provozovatele služeb k novým, obchodním modelům šetrným k soukromí. Nakonec by mělo by být cílem všech uživatelů zvyšovat cenu sledování a tím aspirovat na jeho limitování.[1]

Vyžadování vlastních práv

Posledním aspektem společné obrany je veřejná angažovanost jednotlivců a organizací. Je potřeba apelovat na dodržování a legální posilování osobních práv v oblasti soukromí a osobní bezpečnosti, a na ohleduplnost a zabezpečení zaváděných systémů.

Předpokladem pro legální akce je znát svá legální práva (GDPR, NISD, ePrivacy, zákon č. 101/2002 Sb. apod.) a vyžadovat jejich dodržování.[4] Dále je také potřeba sledovat nové bezpečnostní otázky a apelovat na zakotvení jejich řešení v legislativě.[1]

Kromě toho je dále potřeba od služeb - soukromých i veřejných -, které využívají uživatelská data, vyžadovat transparenci (dostupnost *srozumitelných* informací o tom, jak jsou data využívána, za jakým účelem, jakými algoritmy, a jaká práva uživatel má), možnosti individuální kontroly, legitimní a legální využití dat (přesně stanovená pravidla a jejich dodržování), nezávislý dohled, vyvozování zodpovědnosti a zabezpečení.[1][2]

Zastánci soukromí tvrdí, že občané musí získat více kontroly, pravomocí a ochrany. Prosazení těchto změn je přirozeně úkol vhodný spíše pro bezpečnostní a právní experty, ale jedná se o cíl, na který musí aspirovat společnost jako celek.

Další řešení

Jak je zmíněno v úvodu, přestože uvedené metody ochrany poskytují decentní úroveň soukromí, někteří autoři navrhují pokročilá řešení, která, i když poskytují další úroveň ochrany dat, nejsou vhodná pro každodenní použití a pro širší veřejnost. Obvykle jde případy, kdy je sledování prováděno cíleně a vykonavatelem je zvláště vyškolená osoba. To není předmětem práce a tak jsou jejich rady zmiňovány jen pro doplnění. Ve případě většiny dále zmíněných metod platí za soukromí uživatel časem a komfortem.

Prvním, očividným řešením je částečné opuštění digitálního prostoru. Přesun komunikací offline, používání papírového poznámkového bloku a diáře, fyzická organizace poznámek a dokumentů a další podobné metody mají v některých případech své opodstatnění a účinně znemožňují elektronický únik dat, ale jejich udržování je časově náročné a postrádají výrazné přednosti elektronických řešení, jakými jsou textové vyhledávání, automatické řazení, automatizované upomínkování a nebo synchronizace.[2][3][1]

Další možností jsou anonymní, předplacené služby: kryptoměny a předplacené platební karty pro další ochranu finančních údajů a předplacené SIM karty a virtuální telefoní čísla pro ochranu komunikací.[2][3][5]

Ochraně komunikace se také věnují metody steganografie, jejichž jednou variantou jsou fotografie rukopisu, které nejsou strojově velmi dobře čitelné (ačkoliv optické rozeznávání znaků je technologicky možné a rozvinuté).[1]

V oblasti ochrany před sledování na veřejném prostranství, kromě možnosti se aktivně vyhýbat kamerám, existují produkty, které slibují aktivní ochranu před aktivním sledováním (dohledatelné pod termínem “privacy wear”).[1][2]

Možnost vykoupení dat od datových makléřů může být inzerována, ačkoliv není jisté, do jaké míry je účinná.[2]

Do velice pokročilých metod obfuskace dat a šíření informačního šumu, které zmiňují někteří autoři, patří například vytvoření iluze celého života (včetně přihlášení odběru energií a telekomunikací, nebo předplatných) na jiném místě, cestování a výběry ze vzdálených bankomatů, nebo skutečná relokace na novou adresu (například v zahraničí), či legální změny jmen.[1][3]

Nakonec je opět nutné připomenout, že falšování některých dokladů je samozřejmě ilegální a stejně tak podnikání akcí pro rozbití mechanismu sledování.[1][2]

Vlastní práce

Experiment

Úvod

V rešerši bylo zdokumentováno značné množství potenciálních úniků více či méně soukromých dat. Tato potenciálně slabá místa ochrany soukromí obsahují nakonec v podstatě všechny interakce s internetem a proto není - z důvodu rozsahu - možné provést experiment, který by pokryl všechny možnosti úniků a pro každý z nich bezpečně potvrdil, nebo vyvrátil, zda právě tudy skutečně data unikají a jsou nějakým způsobem využita.

Proto byl výzkum omezen na data soukromých konverzací. Do této kategorie spadá bezpočet komunikačních kanálů, a proto byly pro měření a vyhodnocení vybrány kanály tří společností - Facebook Inc. (Messenger), Google LLC (Google Hangout, Gmail) a Microsoft Corporation (Skype, Outlook).

Jenom málo typů využití dat lze s úspěchem nezávisle potvrdit - jsou to jenom ty, které poskytují přímou zpětnou vazbu. Jako metrika indikující využití uniklých dat byl vybrán objem zobrazované reklamy relevantní k obsahu konverzací.

Principem pokusu je sledování objemu zobrazených reklam relevantních ke konkrétním zvoleným tématům: měření probíhá v kontrolním období a poté v pokusném, mezi nimiž byly do konverzací úmyslně podvržena klíčová slova vztahující se ke sledovanému tématu. Pro každou skupinu kanálů bylo vybráno jiné téma, díky čemuž je případně možné označit původce úniku dat.

Cílem experimentu tedy je na základě zvýšeného množství zobrazených reklam na určité téma prokázat využívání dat ze soukromých digitálních konverzací pro marketingové účely a to v případě kanálů pod kontrolou společností Facebook Inc., Google LLC a Microsoft Corporation.

Metodika experimentu

Předpoklady měření

Účty

Z různých důvodů není možné vytvořit kompletní novou identitu zcela odstíněnou od autorovy vlastní digitální stopy. Proto jsou pro potřeby měření použity autorovy vlastní účty u společností, které jsou předmětem zkoumání, a je sledována odchylka od předchozího stavu.

Všechny účty byly navráceny do výchozího nastavení soukromí. Pro tyto potřeby byly vytvořeny nové účty a podle nich byly nastaveny původní, které jsou použity při měření.

K těmto účtům zůstává během celého měření i ovlivňování autor přihlášen ve využitém prohlížeči.

Prostředí

Všechna měření probíhají v prohlížeči Google Chrome bez jakýchkoliv doplňků a s vymazanými historií prohlížení, cookies a cache.

Operační systém stroje, na němž probíhá měření (laptop řady Acer Aspire), je Microsoft Windows 10 a připojení k internetu je dostupné skrze domácí síť poskytovanou společností O2.

Metodika měření

Způsob podvrhnutí

Na vybraných médiích je započata konverzace, během níž jsou úmyslně zmíněna dále uvedená klíčová slova a to:

- v psané formě ve zprávách příchozích i odchozích
- ve formě souvisejících hypertextových odkazů
- formou obrazových formátů (fotografií)
- formou souvisejících emotikonů

Kromě klíčových slov konkrétních témat jsou dále zmiňována klíčová slova indikující zájem: “chci”, “potřebujeme”, “musíme mít”, “hodilo by se”, “líbí se mi”, “prohlížel jsem si”, “zajímá mě” a podobně.

V prohlížeči, v němž probíhá měření autor provádí minimum dalších akcí, jako např. vyhledávání zmíněných témat, nebo návštěvy vyměňovaných odkazů, což má za cíl snížit vliv jiných úniků dat. Výjimkou je pouze náhodné procházení zmíněných médií při měření.

Před měřením byl všem sledovaným webům udělen souhlas s využíváním cookies.

Témata

Jako podvržená témata byla záměrně vybrána ta, která autora nikdy nezajímala a je proto je menší šance, že by výsledky experimentu byly ovlivněny preexistujícím stavem.

Všechna témata mají charakter zboží.

Lyžařské vybavení

Podvrženo pro: Facebook Inc.

Klíčová slova: **lyže**, lyžování, lyžař, lyžařské vybavení, lyžařská sezóna, hory, sjezdové lyže, jezdit, vázání, sjezdovka, skiareál, lanovka, sníh, Atomic, Elan, Head, Rossignol, Völkl, Dynastar...

Odkazy:

- odkazy na weby zmiňovaných výrobců
  - https://www.atomic.com/
  - https://www.elanskis.com/us-us/cs_CZ/
  - http://www.head.cz/
- odkazy na obchody:
  - https://www.sportisimo.cz/sjezdove-lyze/
  - https://www.intersport.cz/sjezdove-lyze
  - https://www.alza.cz/sport/sjezdove-lyze/18856307.htm
- odkazy na konkrétní produkty:
  - https://www.harfasport.cz/head-supershape-i-rally-prd-12-2019/
  - https://www.mall.cz/lyze/elan-formula-red-qs-el45-17-100
  - https://www.levnelyze.cz/redster-doubledeck-3-0-gs-m-w.htm
  - http://www.happysport.cz/produkt/volkl-rtm-84-uvo-2016-2017/167

Obrazový formát: fotografie lyží

Emotikony: lyže, lyžař, snowboardista, sněhová vločka, lanovka, hory

Obrázek 4.1: Ukázka grafického obsahu pro téma lyže[74][75][76]

Automobily a příslušenství

Podvrženo pro: Microsoft Corporation

Klíčová slova: **auto**, automobil, pneu, pneumatiky, dezén, motor, náhradní díly, jezdit, autobazar, autoškola, "řidičák", Škoda, Fabia, Roomster, Scout, Karoq, Kodiaq, Volkswagen, Brouk, Van, Toyota, Chevrolet, Fiat, Range Rover...

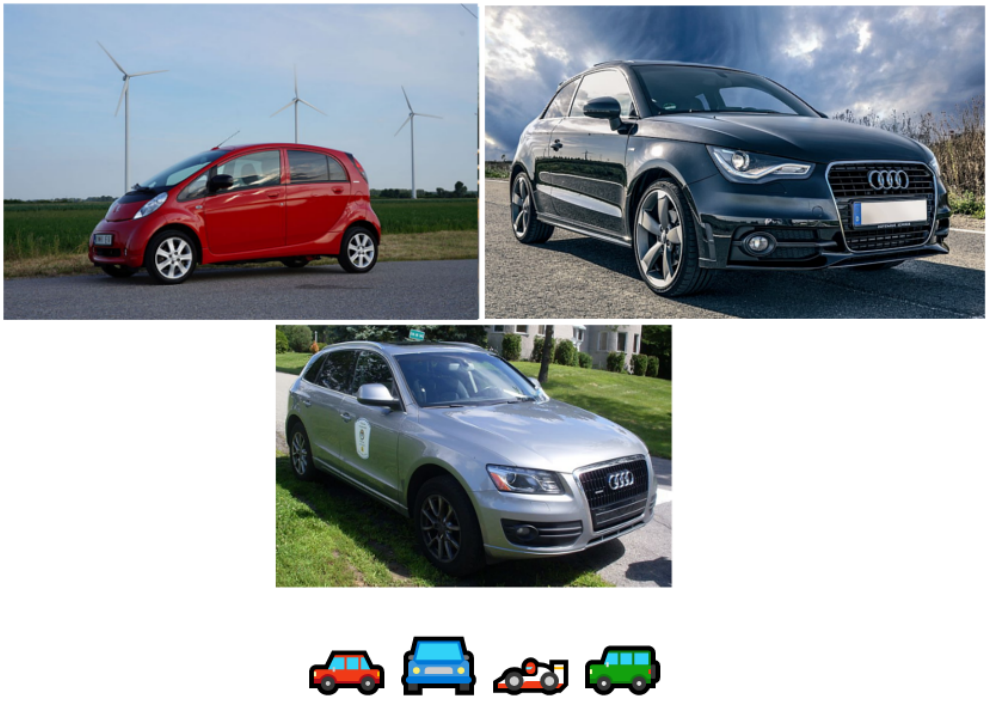
Odkazy:

- odkazy na weby zmiňovaných výrobců
  - https://www.skoda-auto.cz/
  - https://www.volkswagen.cz/
  - https://www.toyota.cz/
- odkazy na obchody:
  - https://www.sauto.cz/inzerce/osobni/skoda
  - https://www.autoesa.cz/volkswagen
  - https://www.aaaauto.cz/toyota/
- odkazy na konkrétní produkty:
  - https://www.aaaauto.cz/cz/skoda-octavia-scout/car.html?id=256484391

- https://www.autoesa.cz/volkswagen/passat/kombi/nafta/226164258
- https://www.sauto.cz/osobni/detail/toyota/corolla/18189720

Obrazový formát: fotografie automobilů

Emotikony: variace automobilů



Obrázek 4.2: Ukázka grafického obsahu pro téma automobily[77][78][79]

Hodinky

Podvrženo pro: Google LLC

Klíčová slova: **hodinky**, hodinářství, pánské hodinky, náramkové hodinky, analogové / "ručičkové" / digitální hodinky, , švýcarské hodinky, Swatch, TAG Heuer, Tissot, Festina, Casio...

Odkazy:

- odkazy na weby zmiňovaných výrobců
  - https://www.tagheuer.com
  - https://www.swatch.com/
  - https://www.tissotwatches.com/

- odkazy na obchody:
  - https://www.timestore.cz/Znacka/Swatch-299/
  - https://www.hodinky-koscom.cz/cz/hodinky-tag-heuer
  - https://www.hodinky-365.cz/panske-hodinky-tissot-x2s012433
- odkazy na konkrétní produkty:
  - https://www.hodinky-365.cz/swatch-blue-grid-yvs454-x1180071
  - https://www.helvetia-hodinky.cz/zbozi/tag-heuer-formula-1-waz1010-ba0842
  - https://www.timestore.cz/Znacka/Tissot-303/Bridgeport/Tissot-Bridgeport-42063
  - https://www.timestore.cz/Znacka/Swatch-299/Swatch-Metal-Knit

Obrazový formát: fotografie pánských náramkových hodinek

Emotikony: hodinky

Obrázek 4.3: Ukázka grafického obsahu pro téma hodinky[80][81][82]

Dovolená v asii

Toto téma bylo sledováno již v kontrolním měření, ale použito k podvrhování bylo až ve druhém ovlivňování.

Klíčová slova: **dovolená**, **zájezd**, Asie, Východ, exotika, Vietnam, Hanoi, Korea, Čína, Japonsko...

Média

Seznam sledovaných webů

Reklamní plochy byly sledovány na výběru z nejnavštěvovanějších webových stránek podle globálního i národního žebříčku Alexa rank, sestaveného v době psaní práce.[83][84]

Z těchto webů byly vybrány ty, které:

1. Obsahují běžné reklamní plochy, nebo viditelně označené promované příspěvky.
2. Pro přístup nevyžadují přihlášení.
3. Jsou anglickém, nebo českém jazyce.

Celý seznam sledovaných domén:

- youtube.com
- facebook.com
- yahoo.com
- reddit.com
- seznam.cz
- idnes.cz
- novinky.cz
- super.cz
- csfd.cz
- seznamzpravy.cz
- sport.cz
- centrum.cz
- aktualne.cz
- blesk.cz
- denik.cz
- stream.cz
- info.cz

Metodika vyhodnocení

K vyhodnocení je využito kontingenčních tabulek do nichž vstupují naměřené počty zobrazených reklam rozdělené podle znaků "ovlivněno / neovlivněno" (podvržením konverzace) a "relevantní / nerelevantní" (k danému tématu) .

Pro prvotní ověření existence vlivu je vytvořena jedna tabulka, která obsahuje informace sebrané ze všech témat a sledovaných zdrojů reklamy. Pokud je pozorovatelný zvýšený výskyt reklamy pro jedno téma, nebo na jednom médiu, pak jsou vytvořeny i specifické tabulky, do nichž nevstupují sumy reklam, ale pouze měření z toho daného média a tématu.

Tabulka 4.1: Příklad kontingenční tabulky využité k ověření ovlivnění relevance

	Nerelevantní	Relevantní
Neovlivněno	n1	n2
Ovlivněno	n3	n4

Nad těmito tabulkami je vyhodnocen Pearsonův chí-kvadrát test nezávislosti v kontingenční tabulce při hodnotě pravděpodobnosti alfa=0,05.

Nulová hypotéza tvrdí, že veličiny relevance a přítomnost ovlivnění jsou nezávislé, tj. že podsunutí tématu do konverzace neovlivňuje četnost zobrazení souvisejících reklam. Alternativní hypotéza tvrdí, že znaky jsou závislé, tj. že ovlivnění konverzace ovlivňuje četnost zobrazení relevantních reklam.

Nulová hypotéza je odmítnuta, pokud vypočtená hodnota Pearsonova chí-kvadrát testu překračuje kritickou hodnotu chí-kvadrát testu pro hladinu pravděpodobnosti alfa.

Pro ověření je využit software GNU PSPP.

Výsledky měření

Kontrolní měření

Při kontrolním měření bylo zaznamenáno celkem 451 reklam, z nichž nerelevantní ke kterémukoliv z vybraných témat bylo 409 (90,6%). Nejvíce relevantních reklam bylo zobrazeno pro téma automobily a to v 37 případech. Pro ostatní témata - lyže, hodinky a dovolenou v Asii se jednalo o 0, 4 a 1 reklamu.

Tabulka 4.2: Protokol o měření. Kontrolní měření před ovlivňováním konverzací

zdroj	celkem	nerelevantních	lyže	automobily	hodinky	dovolená v Asii
youtube.com	23	22	0	1	0	0
facebook.com	14	14	0	0	0	0
yahoo.com	23	22	0	0	1	0
reddit.com	16	16	0	0	0	0
seznam.cz	32	31	0	1	0	0
idnes.cz	32	28	0	4	0	0
novinky.cz	29	27	0	2	0	0
super.cz	33	27	0	5	0	1
csfd.cz	29	25	0	3	1	0
seznamzpravy.cz	30	27	0	3	0	0
sport.cz	27	26	0	1	0	0
centrum.cz	26	23	0	3	0	0
aktualne.cz	26	24	0	2	0	0
blesk.cz	25	20	0	3	2	0
denik.cz	30	24	0	6	0	0
stream.cz	30	29	0	1	0	0
info.cz	26	24	0	2	0	0
SUM	451	409	0	37	4	1

První ovlivňování

Při prvním ovlivňování skrze konverzace bylo během 2 hodin a 21 minut na všech kanálech vyměněno 60 slovních spojení vyjadřujících zájem, 182 klíčových slov, 47 hypertextových odkazů, 26 obrazových formátů a 39 emotikonů.

Tabulka 4.3: Protokol o ovlivňování. Ovlivňování konverzací

téma	společnost	zájem	kl. slova	odkazy	obrázky	emotikony	čas (min)
lyže	Facebook	20	61	9	5	19	42
automobily	Microsoft	20	78	18	12	13	56
hodinky	Google	20	43	20	9	7	43

Měření po prvním ovlivňování

Při měření po ovlivnění bylo zaregistrováno 427 reklam, z nichž nerelevantních bylo 384 (89,9%). Relevantních reklam bylo zaznamenáno téměř stejné množství, jako v měření prvním.

Tabulka 4.4: Protokol o měření. Měření po ovlivnění konverzací

zdroj	celkem	nerelevantních	lyže	automobily	hodinky	dovolená v Asii
youtube.com	32	31	0	1	0	0
facebook.com	19	19	0	0	0	1
yahoo.com	21	20	1	0	0	0
reddit.com	17	17	0	0	0	0
seznam.cz	35	34	0	1	0	0
idnes.cz	28	24	0	4	0	0
novinky.cz	28	26	0	2	0	0
super.cz	22	22	0	0	0	0
csfd.cz	19	17	0	2	0	0
seznamzpravy.cz	33	25	0	8	0	0
sport.cz	25	17	0	8	0	0
centrum.cz	23	19	0	4	0	0
aktualne.cz	20	18	0	2	0	0
blesk.cz	25	24	0	1	0	0
denik.cz	31	28	0	1	2	0
stream.cz	20	20	0	0	0	0
info.cz	28	23	0	3	2	0
SUM	427	384	1	37	4	1

První ověření

Tabulka 4.5: Kontingenční tabulka. Ověření napříč tématy po ovlivnění konverzací

	nerelevantní	relevantní
neovlivněno	409	42
ovlivněno	384	43

relevance * ovlivneni [count, total %, expected].

relevance	ovlivneni		Total
	ano	ne	
ano	42,00	43,00	85,00
	41,24	43,76	,00
	4,78%	4,90%	9,68%
ne	384,00	409,00	793,00
	384,76	408,24	,00
	43,74%	46,58%	90,32%
Total	426,00	452,00	878,00
	48,52%	51,48%	100,00%

Chi-square tests.

Statistic	Value	df	Asymp. Sig. (2-tailed)	Exact Sig. (2-tailed)	Exact Sig. (1-tailed)
Pearson Chi-Square	,03	1,0	,862		
Likelihood Ratio	,03	1,0	,862		
Fisher's Exact Test				,909	,476
Continuity Correction	,00	1,0	,953		
N of Valid Cases	878,0				

Obrázek 4.4: Výsledky prvního ověření v programu PSPP

Výsledná hodnota Pearsonova chí-kvadrát testu vyšla jako 0,03. Kritická hodnota chí-kvadrát testu pro 1 stupeň volnosti a alfa=0,05 je 3,841.

Výsledek testu je tedy je výrazně nižší, než zvolená kritická hodnota a proto nezamítáme nulovou hypotézu a nelze proto tvrdit, že by mezi znaky existovala závislost. To jest, že ovlivnění v tomto případě nemělo vliv na četnost zobrazených reklam.

Druhé ovlivňování

Protože první způsob ovlivnění nebyl úspěšný, rozhodl se autor jiným způsobem ovlivnění ověřit, že pokud by data ze soukromých konverzací byla využívána k cílení reklam, bylo by možné je úspěšně zpracovat v čase mezi prvním ovlivněním a následným měřením.

Proto bylo využito čtvrté sledované téma, které bylo podvrženo pomocí organického vyhledávání (7 výrazů) a návštěvy doporučených odkazů (42), a to během 17 minut.

Z následující tabulky je patrné, že míra ovlivnění byla výrazně nižší, než v případě prvního ovlivnění.

Tabulka 4.6: Protokol o ovlivňování. Ovlivňování vyhledáváním

médium	vyhledávání	otevřené odkazy	doba ovlivňování (min)
Google	3	11	5
Seznam	2	20	4
Bing	2	11	8

Měření po druhém ovlivňování

Doba mezi prvním ovlivňováním a jeho měřením a mezi druhým ovlivňování a měřením byla stejná: v obou případech měření následovalo bezprostředně po konci ovlivňování.

Při tomto druhém měření po ovlivnění bylo zaregistrováno 429 reklam, z nichž nerelevantních bylo pouze 293 (68,3%). Největší zásluhu na tomto poklesu mělo 90 zaregistrovaných reklam na téma "dovolená v Asii", které bylo předmětem druhého ovlivňování.

Tabulka 4.7: Protokol o měření. Měření po ovlivnění vyhledáváním

zdroj	celkem	nerelevantních	lyže	automobily	hodinky	dovolená v Asii
youtube.com	26	24	0	0	2	0
facebook.com	14	14	0	0	0	0
yahoo.com	21	6	0	0	0	15
reddit.com	21	10	0	0	0	11
seznam.cz	29	21	0	0	0	8
idnes.cz	30	15	0	6	0	9
novinky.cz	33	24	3	0	0	6
super.cz	27	14	1	1	0	11
csfd.cz	26	12	1	9	0	4
seznamzpravy.cz	33	16	0	4	0	13
sport.cz	25	20	1	0	0	4
centrum.cz	22	16	0	5	0	1
aktualne.cz	27	22	1	1	0	3
blesk.cz	21	21	0	0	0	0

zdroj	celkem	nerelevantních	lyže	automobily	hodinky	dovolená v Asii
denik.cz	28	23	2	1	0	2
stream.cz	16	14	0	0	0	2
info.cz	30	21	0	6	2	1
SUM	429	293	9	33	4	90

Druhé ověření

Tabulka 4.8: Kontingenční tabulka. Ověření napříč tématy po ovlivnění vyhledáváním

	nerelevantní	relevantní
neovlivněno	384	43
ovlivněno	293	136

relevance * ovlivneni [count, total %, expected].

relevance	*ovlivneni* ano	ne	Total
ano	136,00	43,00	179,00
	89,71	89,29	,00
	15,89%	5,02%	20,91%
ne	293,00	384,00	677,00
	339,29	337,71	,00
	34,23%	44,86%	79,09%
Total	429,00	427,00	856,00
	50,12%	49,88%	100,00%

Chi-square tests.

Statistic	*Value*	*df*	*Asymp. Sig. (2-tailed)*	*Exact Sig. (2-tailed)*	*Exact Sig. (1-tailed)*
Pearson Chi-Square	60,55	1,0	,000		
Likelihood Ratio	63,03	1,0	,000		
Fisher's Exact Test				,000	,000
Continuity Correction	59,25	1,0	,000		
N of Valid Cases	856,0				

Obrázek 4.5: Výsledky druhého ověření v programu PSPP

Výsledná hodnota Pearsonova chí-kvadrát testu vyšla jako 60,55. Kritická hodnota chí-kvadrát testu pro 1 stupeň volnosti a alfa=0,05 je 3,841, a pro alfa=0,001 je 10,828.

Testová hodnota je výrazně vyšší, než kritická hodnota a proto - i při hladině přípustné chyby alfa=0,001 - zamítáme nulovou hypotézu a lze tvrdit, že znaky v tomto případě jsou závislé. To jest, že druhé ovlivnění mělo vliv na četnost zobrazených reklam.

Ověření pro samotné téma

Dle metodiky je ověřováno i ovlivnění samotného tématu “dovolená v Asii”.

Tabulka 4.9: Kontingenční tabulka. Ověření pro téma dovolená v Asii po ovlivnění vyhledáváním

	nerelevantní	relevantní
neovlivněno	450	1
ovlivněno	339	90

relevance * ovlivneni [count, total %, expected].

relevance	*ovlivneni* ano	ne	Total
ano	90,00	1,00	91,00
	44,36	46,64	,00
	10,23%	,11%	10,34%
ne	339,00	450,00	789,00
	384,64	404,36	,00
	38,52%	51,14%	89,66%
Total	429,00	451,00	880,00
	48,75%	51,25%	100,00%

Chi-square tests.

Statistic	*Value*	*df*	*Asymp. Sig. (2-tailed)*	*Exact Sig. (2-tailed)*	*Exact Sig. (1-tailed)*
Pearson Chi-Square	102,17	1,0	,000		
Likelihood Ratio	130,26	1,0	,000		
Fisher's Exact Test				,000	,000
Continuity Correction	99,95	1,0	,000		
N of Valid Cases	880,0				

Obrázek 4.6: Výsledky druhého ověření pro samotné téma v programu PSPP

Výsledná hodnota Pearsonova chí-kvadrát testu vyšla jako 102,17. Kritická hodnota chí-kvadrát testu pro 1 stupeň volnosti a alfa=0,05 je 3,841, a pro alfa=0,001 je 10,828.

Testová hodnota je výrazně vyšší, než kritická hodnota a proto - i při hladině přípustné chyby alfa=0,001 - zamítáme nulovou hypotézu a lze tvrdit, že znaky v tomto případě jsou závislé. To dále potvrzuje hypotézu, že výsledek druhého ověření byl způsobem ovlivněním vyhledávání tématu “dovolená v Asii”.

Vyhodnocení

Přestože se nepodařilo prokázat ovlivnění cílené reklamy skrze soukromou konverzaci, bylo prokázáno, že v době mezi ovlivňováním a měřením byl dostatečný časový rozestup pro to, aby byly reklamy skutečně přizpůsobeny.

To prokázal druhý pokus, při němž byla data podvržena vyhledáváním a procházením webu. Při měření po tomto ovlivnění byl úspěšně dokázán objektivní nárůst relevantních reklam.

Tyto dva závěry indikují, že data soukromých konverzací na platformách, které byly předmětem experimentu, nejsou využívána pro cílení reklam. Je to ovšem méně významný výsledek, než jaký by znamenal opačný výsledek prvního pokusu: tento závěr bohužel neprokazuje, že by data nebyla využívána jinak.

Zjištění, že je tomu právě opačně publikovali žurnalisté The New York Times v prosinci 2018. Podle jejich vyšetřování měly společnosti Spotify, Netflix a Royal Bank of Canada přístup k soukromým zprávám uživatelů Facebooku.[85] Stejný deník v březnu 2019 zveřejnil, že tento případ podléhá federálnímu kriminálnímu vyšetřování.[86]

Představitelé Facebooku ostatně sami přiznali, že Facebook automaticky skenuje soukromé zprávy pro potřeby předcházení nežádoucímu a kriminálnímu chování.[87][88]

Podobná zjištění se již v minulosti týkala Google, který skutečně měl i pro cílení reklam sledovat obsah e-mailů.[89] Navíc podle jiných zjištění společnost poskytovala stejný přístup i třetím stranám.[90]

Zprávy o sledování komunikace na Skype se objevily v roce 2013.[91][92]

Z těchto důvodů je navržené řešení relevantní i přesto, že se zneužívání konverzací neprokázalo statisticky.

Návrh řešení

Základní motivace

Matematická kryptografie je silná. Síla šifrování roste exponenciálně s délkou klíče a proto je relativně snadné vytvořit šifru, u níž je matematicky prokázáno, že je současnou známou technologií neprolomitelná.[1]

Kvůli silnému šifrování na síti (TLS) je pro odposlech komunikace potřeba sledovat data tam, kde jsou uložena v nešifrovaném textu, nebo ve formě, která umožňuje provozovateli komunikační aplikace původní sdělení dešifrovat, tj. na databázových a webových serverech provozovatele.[1]

Proto je pro lepší zabezpečení soukromí osobních konverzací nutné implementovat metody end-to-end šifrování, tedy taková řešení, která zprávy šifrují a dešifrují na samých koncových zařízeních uživatelů. S takovým řešením potom přes veškerou infrastrukturu poskytovatele komunikačního prostředku proudí jenom data, která nemá možnost rozluštit žádným způsobem, kromě hrubé síly.

V tom případě zbývá pro prolomení šifrování útok na uživatelovo zařízení, tj. jeho hardware, operační systém, ovladače zařízení nebo software, nebo na uživatele samotného.

Stále tedy existuje prostor pro narušení soukromí komunikace, ale příklady zneužití ze strany provozovatele, uváděné v předchozí kapitole, nejsou možné. Ze stejného důvodu také není možné data zneužít ze strany bezpečnostních složek, nebo kriminálních pachatelů přistupujících k infrastruktuře provozovatele.

Existující řešení

V Open Source komunitě existují dva v literatuře často zmiňované protokoly: Off-The-Record (OTR) a Pretty Good Privacy (PGP).[93][94] První z nich byl navržen pro instant messaging a využívá symetrické šifrování, druhý je zmiňován v kontextu šifrování e-mailů a využívá asymetrickou kryptografii. Jejich implementace na existující služby je za určitých okolností možná, ale nikoliv uživatelsky přívětivá.[93][94][95]

Na trhu také existují komerční i Open Source aplikace, které tento typ šifrování inzerují: WhatsApp, Signal, Telegram, Viber, LINE a další. Někteří autoři je doporučují, někteří mají námitky.[96]

Zkoumané řešení

Autor zkoumal možnosti implementace šifrování podobného OTR a PGP na webovou aplikaci Facebook Messenger skrze rozšíření do webového prohlížeče.

Autor uvažoval použití symetrického šifrování (Advanced Encryption Standard (AES), nebo Vigenèrovy šifry, potažmo podobné Augustovy šifry), bez automatické výměny klíčů. Ustanovení počátku šifrování mělo proběhnout nastavením šifrovací fráze pro zadanou URL, na straně odesílajícího uživatele.

Základní návrh uvažoval následující případ užití:

- Oba uživatelé využívají stejný komunikátor skrze webové rozhraní
- Oba uživatelé nainstalují rozšíření
- Uživatelé si mimo médium dohodnou šifrovací klíč. Tento klíč si pro zvolenou konverzaci (URL) uloží do vlastního repozitáře klíčů
- V místním repozitáři klíčů jsou všechny fráze zašifrovány tajným heslem jednotlivého uživatele. Při zahájení relace uživatel tímto heslem dešifruje hesla, která následně rozšíření používá
- Rozhraní pro odesílání zpráv je skryto a nahrazeno vlastním rozhraním. Při odeslání zprávy z tohoto nového rozhraní je zpráva zašifrována a teprve poté předána originálnímu
- Všechny zprávy, které jsou prefixovány příznakem, že se jedná o zprávu šifrovanou zvoleným protokolem, jsou skriptem dešifrovány a vloženy na původní místo v HTML kódu aplikace

Generování a ukládání kódů mělo být zjednodušeno takovým způsobem, aby bylo pro uživatele snadné klíče vyměnit. Buďto mělo být možné uložit vlastní frázi (na níž bylo možné se domluvit jakéhokoliv bez přenosu její reprezentace), nebo vygenerovat kryptograficky silný klíč, který by si mohli uživatelé vyměnit formou textu, nebo QR kódu.

Hodnocení návrhu

Po prozkoumání možností se autor rozhodl od návrhu opustit z důvodu následujících omezení.

Zabezpečení klíčů

Hlavním problémem rozšíření webových prohlížečů je fakt, že jejich kód je vykonáván ve stejném rámci, jako načtená stránka. Bylo by tedy možné z kódu stránky přistoupit k vykonávanému kódu rozšíření a odhalit proměnné udržující aktuálně používané klíče.

Kromě toho, i kdyby byl repozitář klíčů chráněn před skripty načtené stránky, musel by stále být z principu dostupný prohlížeči, potažmo jeho výrobci.

Bylo by tedy pro zabezpečení repozitáře klíčů nejspíše nutné kromě rozšíření instalovat na klientské zařízení další aplikaci.

Vyzrazení a změny klíčů

Při vyzrazení klíče je kompromitována celá historie. Tento nedostatek by bylo možné řešit pravidelnou změnou klíčů, která ale z důvodů absence automatické výměny klíčů zůstává na zodpovědnosti uživatelů.

Naopak při změně klíče je historie zpětně nečitelná. Tento nedostatek bylo možné vyřešit ukládáním historie klíčů podle časového rozlišení, tj. v repozitáři klíčů udržovat informace o datu jejich změny.

Omezené možnosti šifrování jiného, než textového obsahu

Obrazové formáty by teoreticky bylo možné před textovým šifrováním kódovat v Base 64, ale objem dat snadno přerůstá maximální povolenou velikost zprávy. Například PNG o velikosti 369 KB (rozměr 1200x1600px) znamenalo v Base 64 493 KB dat. Jiné možnosti stenografie by mohly znamenat ztrátu kvality obrazu a byly by technicky a výpočetně náročnější.

Emotikony jsou součástí znakové sady Unicode a jako takové by je teoreticky bylo možné šifrovat jako text. To by ale vyžadovalo další ověření.

Funkcionalitu zobrazování náhledu hypertextových odkazů může suplovat skript za použití AJAX požadavku, který ale může narazit na problémy s direktivou CSP.

Kompatibilita

Mobilní webové prohlížeče, na rozdíl od desktopových, neumožňují instalaci rozšíření. Z toho důvodu by nebylo možné uvedené řešení na mobilních zařízeních implementovat bez vydání vlastního prohlížeče (např. na bázi Chromium).

Doporučení

Z důvodu popisovaných problémů se autor rozhodl toto řešení dále nerozvíjet a prozkoumat možnosti dostupných alternativ.

Na základě dodatečného průzkumu internetových zdrojů upoutala autorovu pozornost zejména aplikace Signal výrobce Signal Messenger LLC (dříve RedPhone a TextSecure společnosti Open Whisper Systems). Signal Protocol, na němž je aplikace postavena, vychází z OTR a Signal je Open Source s klienty pro OS iOS, Android, Linux, Windows a macOS. Je pozitivně hodnocen a doporučován řadou bezpečnostních expertů.[97][98][99][100][101][102][103][104]

Jak bylo zmíněno výše, nachází se na trhu alternativní nástroje a také Open Source řešení, vhodná spíše pro technicky zdatnější uživatele. Pro jejich porovnání autor navrhuje bližší šetření.

Výsledky a diskuse

Výsledky experimentu byly popisovány již v kapitole “Vyhodnocení”, neboť byly předpokladem pro další část práce.

Experimentem nebylo prokázáno ovlivnění cílené reklamy skrze data privátních konverzací. Bylo prokázáno, že při jiném typu ovlivnění (vyhledávání a procházení webu) byly i při menším objemu zmiňovaných klíčových slov reklamní plochy silně ovlivněny a tudíž, že bylo možné v době mezi ovlivněním a měřením data o uživatelových preferencích na straně poskytovatele reklamy přizpůsobit.

Tyto závěry indikovaly, že tato data tímto způsobem využívána nejsou, ale dodatečná rešerše odhalila případy, které naznačovaly jiný typ užití.

Proto autor zkoumal možnosti implementace end-to-end šifrování. Poté, co autor prozkoumal možnosti šifrování vybraného kanálu a identifikoval problémy tohoto návrhu, rozhodl se na základě doporučení expertů doporučit vybrané existující řešení, nebo jeho zmiňované alternativy.

Závěr

Primárním cílem práce bylo navržení a provedení experimentu, který na měl na základě ovlivňování obsahu konverzace a následného zvýšeného počtu relevantních reklam prokázat využívání dat ze soukromých konverzací, vedených v aplikacích poskytovaných vybranými třemi společnostmi, pro účely cílení reklam.

Pro tento experiment byla vypracována metodika a scénář (kapitola "Metodika experimentu"), podle nichž byl následně pokus proveden (kapitola "Výsledky měření").

Protože výsledky prvního měření podezření neprokázaly, byl proveden dodatečný pokus s ovlivňováním reklamních preferencí podle vyhledávání informací a návštěvy tématických webů, který prokázal, že provozovatelé reklamních ploch na získané informace o uživatelových preferencích reagují v reálném čase, a tudíž, že kdyby byla data ze soukromých konverzací využívána stejným způsobem, bylo by se provedené ovlivnění projevilo (kapitola "Druhé ovlivňování").

Na základě dodatečné rešerše byly přesto objeveny informace naznačující jiné komerční využití těchto dat (kapitola "Vyhodnocení"), na základě čehož autor přistoupil k hledání možnosti zabezpečení vybraného komunikačního kanálu. (kapitola "Návrh řešení").

Autor navrhl zavedení end-to-end šifrování, představil existující řešení a zkoumal možnost jeho implementace na vybranou webovou aplikaci skrze rozšíření pro webový prohlížeč. Během návrhu identifikoval několik výrazných nedostatků zkoumaného přístupu a rozhodl se toto řešení zamítnout jako nefunkční. Na místo toho na základě kladných hodnocení expertů doporučil jedno z existujících řešení jako vhodné pro řešení studovaného problému (kapitoly "Zkoumané řešení" a "Doporučení").

Kromě praktické části také autor detailně prozkoumal problematiku digitálních stop a ochrany soukromí v digitálním prostředí, o které podal obsáhlý přehled, v němž rozlišil jaké typy dat tvoří digitální stopu, jaké subjekty mají zájem osobní data uživatelů využívat, jaké cíle tím sledují, jakými metodami ke sběru a zpracování dat přistupují, a konečně souhrn metod, které literatura doporučuje uživatelům pro ochranu soukromí.

Přílohy

Příloha A - Sledování cookie napříč doménami

Při ověřování možnosti sledování HTTP cookie napříč doménami udělal autor několik pokusů, kterými se snažil na doméně A (dále označované jako "Sociální síť") vytvořit cookie a do stránek na doméně B ("Nezávislý web") vložit skript, načítaný ze Sociální sítě, který by odeslal informace o návštěvě Nezávislého webu zpět na Sociální síť takovým způsobem, aby bylo možné spojit přihlašovací cookie na Sociální síti s URL Nezávislého webu. To znamená, že autor zkusil aktivně sledovat sám sebe napříč doménami.

Pokusy se čtením cookie JavaScriptovým kódem přes hranice objektů `iframe`, nebo `img` byly zpočátku neúspěšné – prohlížeč držel data obou domén oddělená –, a ani při pokusu zaslat aktuální adresu AJAXem nebyly cookies přeneseny (ani po benevolentním nastavení Cross Origin Request Scripting na straně Sociální sítě). Napsání proxy na straně Nezávislého webu nepřipadalo v úvahu (v běžném provozu by nebylo možné vložit svůj skript na server-side) a ani moc nedávalo smysl.

Nakonec se pokus podařil za použití JavaScript metody `Window.postMessage()`.

Autor napsal skript, který byl načten ze Sociální sítě a vložen do Nezávislého webu HTML tagem `script` s odkazem na doménu A. Ten po spuštění vytvořil neviditelný element `iframe` s načtenou stránkou Sociální sítě a pomocí `Window.postMessage()` elementu zaslal aktuální adresu URL (možné bylo zaslat jakoukoliv informaci, kterou měl JavaScript dostupnou na Nezávislém webu; klidně to mohly být všechny informace). V elementu `iframe` jiný spuštěný skript zprávu zaregistroval. V jednu chvíli se tak v `iframe` objevily vedle sebe informace přijaté z Nezávislého webu a přihlašovací cookie Sociální sítě, připravené k odeslání na servery Sociální sítě.

Toto řešení nebylo ani nijak příliš důmyslné a bylo by snadno odhalitelné, nicméně fungovalo, bylo navrženo a zprovozněno během pár hodin, a slouží jako proof of concept.

Seznam použitých zdrojů

[1] Schneier, Bruce. *Data and Goliath*. B.m.: W. W. Norton & Company, 2015. ISBN 978-0393244816.

[2] Angwin, Julia. *Dragnet Nation*. B.m.: Times Books, 2014. ISBN 978-0805098075.

[3] Ahearn, Frank a Eileen Horan. *How to Disappear*. B.m.: Globe Pequot Press, 2010. ISBN 978-1599219776.

[4] Day, Graham. *Security in the Digital World*. B.m.: IT Governance Publishing, 2017. ISBN 978-1849289610.

[5] Bazzell, Michael. *Hiding from the Internet*. B.m.: CreateSpace Independent Publishing Platform, 2016. ISBN 978-1522914907.

[6] Bazzell, Michael. *Personal Digital Security*. B.m.: CreateSpace Independent Publishing Platform, 2013. ISBN 978-1491081976.

[7] Krebs, Brian. *Spam nation*. B.m.: Sourcebooks, 2014. ISBN 978-1402295614.

[8] Brechlerová, Dagmar. *Digitální stopy a jejich odstraňování*. B.m.: Computer World 6/2015, 2015.

[9] Skoček, Jakub. *Digitální stopy – možnosti kontroly a eliminace pomocí vybraných volně dostupných nástrojů*. B.m.: Univerzita Karlova v Praze, 2012.

[10] Cambridge Dictionary. *data* [online]. 2019 [viděno. 2019-03-10]. Dostupné z: https://dictionary.cambridge.org/dictionary/english/data

[11] Cambridge Dictionary. *metadata* [online]. 2019 [viděno. 2019-03-12]. Dostupné z: https://dictionary.cambridge.org/dictionary/english/metadata

[12] *Zákon č. 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů*. 2000

[13] University of North Carolina at Chapel Hill. *sensitive data* [online]. 2019 [viděno. 2019-03-10]. Dostupné z: https://its.unc.edu/security/sensitive-data/

[14] *Nařízení Evropského parlamentu a Rady (EU) č. 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů)*

[15] Business Dictionary. *behavioral data* [online]. 2019 [viděno. 2019-03-12]. Dostupné z: http://www.businessdictionary.com/definition/behavioral-data.html

[16] 2040 Digital. *What Is Behavioral Data?* [online]. 2019 [viděno. 2019-03-12]. Dostupné z: https://www.2040digital.com/the-impact-of-digital/what-is-behavioral-data/

[17] Dictionary.com. *sensitive data* [online]. 2019 [viděno. 2019-03-10]. Dostupné z: https://www.dictionary.com/browse/digital-footprint

[18] Technopedia. *Digital Footprint* [online]. 2019 [viděno. 2019-03-14]. Dostupné z: https://www.techopedia.com/definition/2396/digital-footprint

[19] CSO. *The 18 biggest data breaches of the 21st century* [online]. 2019 [viděno. 2019-03-18]. Dostupné z: https://www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html

[20] Youyou, Wu, Michael Kosinski a David Stillwell. *Computer-based personality judgments are more accurate than those made by humans* [online]. 2015 [viděno. 2019-03-12]. Dostupné z: https://www.pnas.org/content/112/4/1036

[21] *Zákon č. 40/2009 Sb., trestní zákoník*

[22] Ministerstvo vnitra ČR. *Analýza odposlechů a sledování 2017* [online]. 2017 [viděno. 2019-03-20]. Dostupné z: https://www.mvcr.cz/clanek/odposlechy-zaznamy-telekomunikacniho-provozu-a-sledovani-osob.aspx

[23] wpmudev. *A History of WordPress Security Exploits and What They Mean* [online]. 2017 [viděno. 2019-03-19]. Dostupné z: https://premium.wpmudev.org/blog/wordpress-security-exploits/

[24] OWASP. *OWASP Top 10 Application Security Risks - 2017* [online]. 2017 [viděno. 2019-03-10]. Dostupné z: https://www.owasp.org/index.php/Top_10-2017_Top_10

[25] Rhodan, Maya a Time . *Privacy Advocates Warn About Airport Security Book Searches* [online]. 2017 [viděno. 2019-03-14]. Dostupné z: http://time.com/4832616/airport-security-books-tsa/

[26] Hern, Alex a The Guardian . *Fitness tracking app Strava gives away location of secret US army bases* [online]. 2018 [viděno. 2019-03-12]. Dostupné z: https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

[27] Kaspersky lab. *What is IP spoofing?* [online]. 2019 [viděno. 2019-03-21]. Dostupné z: https://usa.kaspersky.com/resource-center/threats/ip-spoofing

[28] Mimoso, Michael a threat post . *Internet Traffic Following Malicious De-*

tours Via Route Injection Attacks [online]. 2013 [viděno. 2019-03-10]. Dostupné z: https://threatpost.com/internet-traffic-following-malicious-detours-via-route-injection-attacks/102981/

[29] Kaspersky lab. *What is DNS Cache Poisoning or Spoofing?* [online]. 2019 [viděno. 2019-03-11]. Dostupné z: https://www.kaspersky.com/resource-center/definitions/dns

[30] Electronic Frontier Foundation. *Cell-Site Simulators/IMSI Catchers* [online]. 2019 [viděno. 2019-03-18]. Dostupné z: https://www.eff.org/pages/cell-site-simulatorsimsi-catchers

[31] *zákon č. 127/2005 Sb., o elektronických komunikacích v § 97, odst. 3*

[32] *vyhláška 357/2012 Sb., o uchovávání, předávání a likvidaci provozních a lokalizačních údajů*

[33] PixelPrivacy. *Browser Fingerprinting* [online]. 2019 [viděno. 2019-03-12]. Dostupné z: https://pixelprivacy.com/resources/browser-fingerprinting/

[34] Eckersley, Peter a Electronic Frontier Foundation . *How Unique Is Your Web Browser?* [online]. 2014 [viděno. 2019-03-09]. Dostupné z: https://panopticlick.eff.org/static/browser-uniqueness.pdfg/

[35] O'Donald, Shannon a Bl.ink . *8 Reasons Your Referrer Data Doesn't Tell the Whole Story* [online]. 2018 [viděno. 2019-03-14]. Dostupné z: https://www.bl.ink/blog/referrer-data/

[36] Electronic Frontier Foundation. *Investigating Machine Identification Code Technology in Color Laser Printers* [online]. 2005 [viděno. 2019-03-10]. Dostupné z: https://www.eff.org/wp/investigating-machine-identification-code-technology-color-laser-printers

[37] *zákon č. 565/1990 Sb., o místních poplatcích v § 3, odst. 4*

[38] Greenberg, Andy a Wired. *Marketing firm Exactis leaked a personal info database with 340 million records* [online]. 2018 [viděno. 2019-03-14]. Dostupné z: https://www.wired.com/story/exactis-database-leak-340-million-records/

[39] Maynard a The Independent. *China data breaches: 33-mil unprotected job applicant profiles leaked* [online]. 2019 [viděno. 2019-03-26]. Dostupné z: http://theindependent.sg/china-data-breaches-33-mil-unprotected-job-applicant-profiles-leaked/

[40] Kamkar, Samy. *evercookie* [online]. 2010 [viděno. 2019-03-18]. Dostupné z: https://samy.pl/evercookie/

[41] NSA. *Tor Stinks* [online]. 2013 [viděno. 2019-03-21]. Dostupné z: https://edwardsnowden.com/docs/doc/tor-stinks-presentation.pdf

[42] BuiltWith. *Google Analytics Usage Statistics* [online]. 2019 [viděno. 2019-03-10]. Dostupné z: https://trends.builtwith.com/analytics/Google-Analytics

[43] BuiltWith. *Facebook Sharer Usage Statistics* [online]. 2019 [viděno. 2019-03-26]. Dostupné z: https://trends.builtwith.com/widgets/Facebook-Sharer

[44] Ghostery Team. *The Tracker Tax* [online]. 2018 [viděno. 2019-03-19]. Dostupné z: https://www.ghostery.com/lp/trackertax/

[45] Ghostery Team. *Tracking the Trackers* [online]. 2017 [viděno. 2019-03-12]. Dostupné z: https://www.ghostery.com/lp/study/

[46] Alsenoy, Brendan Van, Valerie Verdoodt, Bob Heyman, Jef Ausloos, Ellen Wuaters a Gunes Acar. *From social media service to advertising network* [online]. 2015 [viděno. 2019-03-11]. Dostupné z: https://www.researchgate.net/publication/291147719_From_social_media_service_to_advertising_network_-_A_critical_analysis_of_Facebook's_Revised_Policies_and_Terms

[47] Anderson, Monica a Pew Reserach Center. *A Majority of Teens Have Experienced Some Form of Cyberbullying* [online]. 2018 [viděno. 2019-03-20]. Dostupné z: https://www.pewinternet.org/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/

[48] Roberts, Jeff John a Fortune. *Facebook, Uber and the Trouble With "God View"* [online]. 2018 [viděno. 2019-03-12]. Dostupné z: http://fortune.com/2018/05/12/facebook-employee-peeking/

[49] Sýkora, Filip a Ekonom. *Náborář tě vidí. O přijetí uchazeče rozhoduje i digitální stopa* [online]. 2017 [viděno. 2019-03-11]. Dostupné z: https://ekonom.ihned.cz/c1-65871270-pred-prvnim-pohovorem-zamette-digitalni-stopy

[50] Wikimedia Commons contributors. *File:Gerrymander diagram for four sample districts.gif* [online]. 2005 [viděno. 2019-03-20]. Dostupné z: https://commons.wikimedia.org/wiki/File:Gerrymander_diagram_for_four_sample_districts.gif

[51] Walker, Shaun, Oksana Grytsenko a The Guardian. *Text messages warn Ukraine protesters they are 'participants in mass riot'* [online]. 2014 [viděno. 2019-03-09]. Dostupné z: https://www.theguardian.com/world/2014/jan/21/ukraine-unrest-text-messages-protesters-mass-riot

[52] Bhagat, Smirti, Moira Burke, Carlos Diuk, Ismail Onur Filiz, Sergey Edunov a Facebook. *Three and a half degrees of separation* [online]. 2016 [viděno. 2019-03-10]. Dostupné z: https://research.fb.com/three-and-a-half-degrees-of-separation/

[53] Clark, Grant, Edwin Chan a Bloomberg. *The Great Firewall of China* [online]. 2018 [viděno. 2019-03-18]. Dostupné z: https://www.washingtonpost.com/business/the-great-firewall-of-china/2018/11/05/5dc0f85a-e16d-11e8-ba30-a7ded04d8fac_story.html?utm_term=.499f6a30c011

[54] BBC. *Organised crime threat greater than terrorism - National Crime Agency* [online]. 2018 [viděno. 2019-03-23]. Dostupné z: https://www.bbc.co.uk/news/uk-46050428

[55] Dearden, Lizzie a Independent. *Serious and organised crime kills more people than terrorism in the UK, warns NCA* [online]. 2018 [viděno. 2019-03-14]. Dostupné z: https://www.independent.co.uk/news/uk/home-news/serious-organised-crime-uk-economy-child-abuse-trafficking-drug-dealing-nca-a8611886.html

[56] Erben, Lukáš a Root. *Příchod hackerů: příběh Stuxnetu* [online]. 2014 [viděno. 2019-03-12]. Dostupné z: https://www.root.cz/clanky/prichod-hackeru-pribeh-stuxnetu/

[57] ČTK, iDNES.cz. *PŘEHLEDNĚ: Jak se vyvíjela kauza okolo čínské společnosti Huawei* [online]. 2019 [viděno. 2019-03-21]. Dostupné z: https://www.idnes.cz/zpravy/domaci/kauza-telefon-huawei-prehledne-babis-zeman-bis-nukib.A190111_114049_domaci_onkr

[58] Krčmář, Petr a Root. *Číňané infikovali při výrobě servery Supermicro miniaturní štěnicí* [online]. 2018 [viděno. 2019-03-24]. Dostupné z: https://www.root.cz/clanky/cinane-infikovali-pri-vyrobe-servery-supermicro-miniaturni-stenici/

[59] Khandelwal, Swati a The Hacker News. *Built-in Keylogger Found in MantisTek GK2 Keyboards—Sends Data to China* [online]. 2017 [viděno. 2019-03-15]. Dostupné z: https://thehackernews.com/2017/11/mantistek-keyboard-keylogger.html

[60] Reflex. *Kauza Nikulin přehledně: Od zadržení v Praze po vydání do Spojených států a reakci Ruska* [online]. 2018 [viděno. 2019-03-20]. Dostupné z: https://www.reflex.cz/nikulin-jevgenij-hacker

[61] Gollis, Ondřej a Český rozhlas. *Nová zbraň informační války. ‚Deep fakes mohou napáchat obrovské škody,' varuje odbornice* [online]. 2018 [viděno. 2019-03-21]. Dostupné z: https://www.irozhlas.cz/zpravy-domov/dezinformace-fake-news-deep-fakes-petra-vejvodova_1903060633_ogo

[62] Schwartz, Matthew J. a Bank Info Security. *Cybercrime-as-a-Service Economy:*

Stronger Than Ever [online]. 2016 [viděno. 2019-03-15]. Dostupné z: https://www.bankinfosecurity.com/cybercrime-as-a-service-economy-stronger-than-ever-a-9396

[63] Johnson, Larry a Enterpreteur. *Crime-as-a-Service Could Be the Next Big Threat to Your Business* [online]. 2017 [viděno. 2019-03-18]. Dostupné z: https://www.entrepreneur.com/article/298727

[64] Dočekal, Daniel a Lupa.cz. *Hack společnosti Hacking Team: na seznamu klientů je i česká policie* [online]. 2015 [viděno. 2019-03-26]. Dostupné z: https://www.lupa.cz/clanky/hack-spolecnosti-hacking-team-na-seznamu-klientu-je-i-ceska-policie/

[65] McMullan, Thomas. *What does the panopticon mean in the age of digital surveillance?* [online]. 2015 [viděno. 2019-03-19]. Dostupné z: https://www.theguardian.com/technology/2015/jul/23/panopticon-digital-surveillance-jeremy-bentham

[66] Abraham, Shawn a MalwareFox. *List of Types of Malware* [online]. 2017 [viděno. 2019-03-27]. Dostupné z: https://www.malwarefox.com/malware-types/

[67] Zetter, Kim, Brian Barrett a Wired. *Apple to FBI: You can't force us to hack the San Bernardino iPhone* [online]. 2016 [viděno. 2019-03-27]. Dostupné z: https://www.wired.com/2016/02/apple-brief-fbi-response-iphone/

[68] distributed.net. *Project RC5* [online]. 2019 [viděno. 2019-03-09]. Dostupné z: https://www.distributed.net/RC5

[69] Password Depot. *Brute-force attacks* [online]. 2019 [viděno. 2019-03-26]. Dostupné z: https://www.password-depot.de/en/know-how/brute-force-attacks.htm

[70] Veselý, Arnošt. *Metody umělé inteligence*. B.m.: Česká zemědělská univerzita v Praze, 2012. ISBN 978-80-213-2295-0.

[71] Veselý, Arnošt. *Úvod do umělé inteligence*. B.m.: Česká zemědělská univerzita v Praze, 2005. ISBN 80-213-1361-7.

[72] Jadrný, Petr a Český rozhlas. *Testem antivirů pro Android neprošly dvě třetiny programů. Uspěly jen velké firmy* [online]. 2019 [viděno. 2019-03-12]. Dostupné z: https://www.irozhlas.cz/veda-technologie/technologie/android-antivirove-programy-test_1903191409_pj

[73] Hogan, Patrick a The Outline. *Mastodon makes the internet feel like home again* [online]. 2017 [viděno. 2019-03-15]. Dostupné z: https://theoutline.com/post/2689/mastodon-makes-the-internet-feel-like-home-again

[74] Wikimedia Commons contributors. *File:Recreational skiers in the mountains take*

breaks and enjoy skiing in the regions.jpg [online]. 2013 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Recreational_skiers_in_the_mountains_take_breaks_and_enjoy_skiing_in_the_regions.jpg

[75] Wikimedia Commons contributors. *File:Arizona Snowbowl Grand Canyon Express Ski Lift Opening Celebration (30795518093).jpg* [online]. 2016 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Arizona_Snowbowl_Grand_Canyon_Express_Ski_Lift_Opening_Celebration_(30795518093).jpg

[76] Wikimedia Commons contributors. *File:Olympische Spelen Grenoble, ski-slalom dames. Nancy Green (Canada 2e), Bestanddeelnr 921-0892.jpg* [online]. 1968 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Olympische_Spelen_Grenoble,_ski-slalom_dames._Nancy_Green_(Canada_2e),_Bestanddeelnr_921-0892.jpg

[77] Wikimedia Commons contributors. *File:Lights-clouds-dark-car (23698576464).jpg* [online]. 2016 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Lights-clouds-dark-car_(23698576464).jpg

[78] Wikimedia Commons contributors. *File:Public domain image - Peugeot iOn electric car in front of wind turbines.JPG* [online]. 2015 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Public_domain_image_-_Peugeot_iOn_electric_car_in_front_of_wind_turbines.JPG

[79] Wikimedia Commons contributors. *File:Audi Q5 Baie D'Urfé Public Security (Auto classique VAQ Baie-D'Urfé '13).JPG* [online]. 2013 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Audi_Q5_Baie_D%27Urfé_Public_Security_(Auto_classique_VAQ_Baie-D%27Urfé_%2713).JPG

[80] Wikimedia Commons contributors. *File:Invicta Mens Automatic Pro Diver Watch 8926.jpg* [online]. 2015 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Invicta_Mens_Automatic_Pro_Diver_Watch_8926.jpg

[81] Wikimedia Commons contributors. *File:Fashion-wristwatch-time-watch (24217032812).jpg* [online]. 2014 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Fashion-wristwatch-time-watch_(24217032812).jpg

[82] Wikimedia Commons contributors. *File:Watches, Roamer Stingray and Kienzle - Karl Gebhardt Horological Collection - Gewerbemuseum - Nuremberg, Germany - DSC01839.jpg* [online]. 2016 [viděno. 2019-03-16]. Dostupné z: https://commons.wikimedia.org/wiki/File:Watches,_Roamer_Stingray_and_Kienzle_-_Karl_Gebhardt_Horological_Collection_-_Gewerbemuseum_-

_Nuremberg,_Germany_-_DSC01839.jpg

[83] Alexa. *The top 500 sites on the web* [online]. 2019 [viděno. 2019-03-18]. Dostupné z: https://www.alexa.com/topsites

[84] Alexa. *Top Sites in Czech Republic* [online]. 2019 [viděno. 2019-03-18]. Dostupné z: https://www.alexa.com/topsites/countries/CZ

[85] Dance, Gabriel J. X., Michael LaForgia, Nicholas Confessore a The New York Times. *As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants* [online]. 2018 [viděno. 2019-03-26]. Dostupné z: https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html

[86] LaForgia, Michael, Gabriel J. X. Dance Matthew Rosenberg a The New York Times. *Facebook's Data Deals Are Under Criminal Investigation* [online]. 2019 [viděno. 2019-03-19]. Dostupné z: https://www.nytimes.com/2019/03/13/technology/facebook-data-deals-investigation.html

[87] Chang, Lulu a Digital Trends. *Yes, Facebook is reading the messages you send through Messenger* [online]. 2018 [viděno. 2019-03-12]. Dostupné z: https://www.digitaltrends.com/social-media/facebook-reads-messenger-messages/

[88] Frier, Sarah a Bloomberg. *Facebook Scans the Photos and Links You Send on Messenger* [online]. 2018 [viděno. 2019-03-12]. Dostupné z: https://www.bloomberg.com/news/articles/2018-04-04/facebook-scans-what-you-send-to-other-people-on-messenger-app

[89] Hern, Alex a The Guardian. *Google will stop scanning content of personal emails* [online]. 2017 [viděno. 2019-03-09]. Dostupné z: https://www.theguardian.com/technology/2017/jun/26/google-will-stop-scanning-content-of-personal-emails

[90] Dodds, Laurence, Margi Murphy a The Telegraph. *Google admits it lets hundreds of other companies access your Gmail inbox* [online]. 2018 [viděno. 2019-03-21]. Dostupné z: https://www.telegraph.co.uk/technology/2018/09/20/google-admits-hundreds-companies-read-gmail-inbox/

[91] Mello, John P. a PC World. *Microsoft may be scanning your Skype messages* [online]. 2013 [viděno. 2019-03-15]. Dostupné z: https://www.pcworld.com/article/2039410/microsoft-may-be-scanning-your-skype-messages.html

[92] Goodin, Dan a ArsTechnica. *Think your Skype messages get end-to-end encryption? Think again* [online]. 2013 [viděno. 2019-03-12]. Dostupné z: https://arstechnica.com/information-technology/2013/05/think-your-skype-messages-get-end-to-end-encryption-think-again/

[93] Goldberg, Ian a The OTR Development Team. *Off-the-Record Messaging Protocol version 3* [online]. [viděno. 2019-03-28]. Dostupné z: https://otr.cypherpunks.ca/Protocol-v3-4.1.1.html

[94] Brocklehurst, George. *A pretty good introduction to Pretty Good Privacy* [online]. [viděno. 2019-03-28]. Dostupné z: https://georgebrock.github.io/talks/pretty-good-introduction/

[95] Rusendić, Stanko Krtalić. *Hacking privacy into Facebook's Messenger in 24 hours* [online]. [viděno. 2019-03-28]. Dostupné z: https://blog.stanko.io/hacking-privacy-into-facebook-s-messenger-in-24-hours-3da239baf6c5

[96] DeVault, Drew. *I don't trust Signal* [online]. 2018 [viděno. 2019-03-28]. Dostupné z: https://drewdevault.com/2018/08/08/Signal.html

[97] McMahon, Jordan a Wired . *Ditch all those other messaging apps: Here's why you should use Signal* [online]. 2017 [viděno. 2019-03-28]. Dostupné z: https://www.wired.com/story/ditch-all-those-other-messaging-apps-heres-why-you-should-use-signal/

[98] Kahle, Brewster. *Upgraded Secure Communications Applications I am Now Using* [online]. 2017 [viděno. 2019-03-28]. Dostupné z: https://blog.archive.org/2017/02/03/upgraded-secure-communications-applications-i-am-now-using/

[99] Lee, Micah a The Intercept . *Battle of the secure messaging apps: How Signal beats WhatsApp* [online]. 2016 [viděno. 2019-03-28]. Dostupné z: https://theintercept.com/2016/06/22/battle-of-the-secure-messaging-apps-how-signal-beats-whatsapp/

[100] O'Neill, Patrick Howell a CyberScoop . *Signal's protocol gets glowing reviews in first security audit* [online]. 2016 [viděno. 2019-03-28]. Dostupné z: https://www.cyberscoop.com/signal-security-audit-encryption-facebook-messenger-whatsapp/

[101] Cohn-Gordon, Katriel, Cas Cremers, Benjamin Dowling, Luke Garratt a Douglas Stebila. *A Formal Security Analysis of the Signal Messaging Protocol* [online]. 2017 [viděno. 2019-03-28]. Dostupné z: https://eprint.iacr.org/2016/1013.pdf

[102] Schneier, Bruce. *How Signal Is Evading Censorship* [online]. 2016 [viděno. 2019-03-28]. Dostupné z: https://www.schneier.com/blog/archives/2016/12/how_signal_is_e.html

[103] Electronic Frontier Foundation. *Surveillance Self-Deffense: Communicating with Others* [online]. 2018 [viděno. 2019-03-28]. Dostupné z: https://ssd.eff.org/en/module/communicating-others

[104] Yin, Dave a Channel Daily News. *Here's the encrypted call and messenger app that Edward Snowden himself uses* [online]. 2015 [viděno. 2019-03-28]. Dostupné z: https://channeldailynews.com/news/heres-the-encrypted-call-and-messenger-app-that-edward-snowden-himself-uses/45311