

Univerzita Karlova

Pedagogická fakulta

Katedra informačních technologií a technické výchovy (41-KITTV)

BAKALÁŘSKÁ PRÁCE

**Rozsah a dopady sběru osobních dat v digitálním
prostředí s přesahem do vzdělávání**

**The scope and impact of the collection of personal
data in the digital environment, extending to
education**

Jan Peterka

Vedoucí bakalářské práce: PhDr. Josef Procházka, Ph.D.

Studijní program: Specializace v pedagogice (B7507)

Studijní obor: Informační technologie se zaměřením na vzdělávání

Praha 2021

Odevzdáním této bakalářské práce na téma Rozsah a dopady sběru osobních dat v digitálním prostředí s přesahem do vzdělávání potvrzuji, že jsem ji vypracoval pod vedením vedoucího práce samostatně za použití v práci uvedených pramenů a literatury. Dále potvrzuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

Praha, 17.4. 2021

Podpis autora

Děkuji Emilovi Millerovi za poskytnutí TeX šablony a pomoc s formátováním práce.

Abstrakt

Cílem bakalářské práce je zmapovat téma vytváření digitální stopy a osobních dat v digitálním prostředí, a ukázat tuto problematiku zejména z pohledu jednotlivce, možného vlivu na něj a možností jeho vlivu na vytváření vlastní digitální stopy.

Dále se práce zabývá napojením tohoto tématu do stávajících i vznikajících rámcových vzdělávacích plánů a obecně do aktuálního školského vzdělávání.

Ze zjištěných poznatků pak vychází praktická část práce, která si dává za cíl vytvořit prototyp online prostředí pro seznamování lidí (a zejména žáků, studentů) s tímto tématem popularizační a interaktivní formou.

Klíčová slova

Sběr digitálních dat, Digitální stopa, Bezpečnost

Abstract

The aim of the bachelor thesis is to map the topic of creating a digital footprint and personal data in a digital environment, and to show this issue especially from the perspective of the individual, the possible influence on him and the possibility of his influence on creating his own digital footprint.

Furthermore, the work deals with the connection of this topic to existing and emerging educational plans and in general to the current school education.

The practical part of the work, based on these findings, aims to create a prototype online environment for acquainting people (and especially pupils, students) with this topic in a popularizing and interactive form.

Keywords

Collection of personal data, digital environment, security

Obsah

Úvod	8
1 Problematika ochrany digitálních osobních dat uživatelů	9
1.1 Úvod	9
1.2 Použité pojmy a koncepty	9
1.3 Motivace pro sběr dat	9
1.4 Technické možnosti sběru dat	10
1.5 Možnosti analýzy dat	10
1.6 Povědomí o problematice	11
1.7 Shrnutí	11
2 Zařazení problematiky digitální stopy do školních výukových materiálů	12
2.1 Úvod a vymezení pojmů	12
2.2 Vymezení dokumentů	13
2.3 Rámcové vzdělávací plány	14
2.3.1 RVP základní školy	14
2.3.2 RVP-G	14
2.3.3 RVP odborné školy – Informační technologie	15
2.3.4 RVP odborné školy – Informační služby	15
2.4 Návrh revizí RVP v oblasti Informatiky a informačních a komunikačních technologií .	15
2.5 Shrnutí	16
3 Rizikové dopady sběru osobních dat	17
3.1 Úvod	17
3.2 Změna chování na základě vědomí vytváření stopy	17

3.3	Možnosti zneužití dat	18
3.3.1	S dopadem na jednotlivce	18
3.3.2	S dopadem na společnost	20
3.4	Shrnutí	21
4	Hlavní zdroje dat digitální stopy uživatele	22
4.1	Úvod	22
4.2	Kategorie dat	22
4.2.1	Aktivní, pasivní a vědomě nevědomá digitální stopa	22
4.2.2	Identifikovatelná a anonymní data	23
4.3	Konkrétní typy dat	24
4.3.1	Historie prohlížení	24
4.3.2	Lokační data	25
4.3.3	Data vkládaná na sociální sítě a další platformy	26
4.3.4	Další data sledovaná sociálními sítěmi a dalšími platformami	26
4.3.5	E-mail	28
4.3.6	Finanční záznamy	28
4.3.7	Zdravotní data	29
5	Klíčové zásady uživatelské ochrany osobních dat	30
5.1	Obecné náhledy	31
5.1.1	Co je na internetu, to je veřejné	31
5.1.2	Budoucnost	31
5.2	Zásady a nástroje	31
5.2.1	Omezení vědomého sdílení informací	31
5.2.2	Anonymní prohlížení webu	31
5.2.3	Pravidla pro aplikace a služby	31
5.2.4	Šifrování dat	31
5.2.5	Phishing	31
5.2.6	Ochrana zařízení	31

5.2.7	Hesla a zabezpečení	31
6	Praktická část	32
6.1	Východiska	32
6.1.1	Návrh aplikace z pohledu uživatele	32
6.1.2	Prototyp	33
6.1.3	Scénáře	33
6.2	Uživatelské workflow a use-cases	35
6.3	Technické řešení	35
6.3.1	Výběr technologií	35
6.3.2	Základní funkcionality	36
6.3.3	Objektová vrstva	36
6.3.4	Ověření a otestování aplikace	36
6.3.5	Vyhodnocení provedeného ověření a doporučení pro další rozvoj aplikace	36
	Závěr	37

Úvod

Cambridge Analytica, Čínský kreditový systém nebo předvolání ředitelů firem jako je Facebook a Google před americký Kongres. To je několik mediálně známých případů, které v posledních letech vynášejí téma osobních dat, jejich směru a s tím spojených rizik, do povědomí širší veřejnosti. Jsou však uživatelé (tedy my všichni) o tématu dostatečně informováni? Jsou tyto informace takové, abychom se zvládali v tématu orientovat, a zároveň dělat zodpovědná, informovaná rozhodnutí o našem vlastním chování v digitálním světě?

Tato práce si dává za cíl téma osobních údajů a digitální stopy zmapovat, a to primárně z pohledu jednotlivce – uživatele. Pokouší se shrnout základní témata a pojmy, které pomáhají se v tématu zorientovat a mluvit o něm, prozkoumat typy dat a digitální stopy, která naším každodenním chováním vznikají, upozornit na rizika, a zároveň nabídnout konkrétní přístupy, kterými já jako uživatel můžu tuto realitu ovlivňovat.

Tyto teoretické poznatky jsou pak kromě jejich samotného přínosu i podkladem pro praktickou část práce - vytvoření prostředí, ve kterém je uživatel tímto tématem provázen.

Kapitola 1

Problematika ochrany digitálních osobních dat uživatelů

1.1 Úvod

Rozvoj digitální části života s sebou mimo jiné přinesl čím dál větší objem digitálních (a tedy relativně snadno uchovatelných, duplikovatelných a strojově či částečně strojově zpracovatelných) dat. V online světě trávíme čím dál větší množství času – v roce 2020 to bylo kolem 7 hodin denně (Kemp, 2019)) – a spolu s tím jsou čím dál pokročilejší technologie pro sběr a vyhodnocování dat.

V této kapitole se podíváme na důvody, které vedou k aktuálnímu stavu v oblasti sběru osobních dat, jak z pohledu ekonomické a jiné motivace, tak s pohledu technologického. Dále rozebereme rizika tohoto stavu.

1.2 Použité pojmy a koncepty

Networked privacy

Model soukromí, který oproti klasickému pohledu individuálnímu přidává rozměr propojenosti – zohledňuje tedy fakt, že lidé sdílejí (ať vědomě, nebo nevědomě) i informace o ostatních, a tedy moje soukromí není oddělitelné od soukromí dalších (Marwick – danah boyd, 2014)

Do češtiny by šlo přeložit jako *propojené soukromí*, v práci však používám originální název.

1.3 Motivace pro sběr dat

V diskuzích o osobních datech a jejich sběru se nejčastěji objevují jména dvou digitálních firem – Google a Facebook (případně ještě Apple). Důvodem je samozřejmě to, že jsou obě velkou součástí nabídky digitálních služeb.

Facebook se blíží ke třem miliardám aktivních uživatelů (Tankovska, 2020), u Google je výpočet složitější (protože jde o různé služby), ale jen mezi vyhledávači (tedy esenciální služba pro většinu uživatelů internetu) má podíl přes 90% (Statcounter, 2021). S celkovým odhadem počtu 4.6 miliardy digitálních uživatelů (Kemp, 2019) tedy jde o více než 4 miliardy uživatelů.

Obě tyto firmy mají svůj obchodní model založený na nabízení reklamy – v případě obou firem jde o primární zdroj příjmů. Jejich dominance na trhu (v USA má Google 30% trhu reklam (eMarketer Editors, 2019) v digitálním prostředí, Facebook následuje s 20%) je umožněna dobrým zacílením reklamy. Tato personalizace je umožněna právě sběrem dat a jejich analýzou.

1.4 Technické možnosti sběru dat

Pro tvorbu profilů jednotlivých uživatelů webu a nabízení reklamy je potřeba sbírat co nejvíce informací o chování uživatele v online světě. Vytvoření profilu člověka nebylo technicky téměř možné před začátkem používání *cookies*.

Cookies byly poprvé implementovány v roce 1994 do prohlížeče Netscape jako řešení pro ukládání stavu nákupního košíku (Penland, 2020). Cookies umožňují uchovávat informace o uživateli na dané stránce, s rozvojem webu a zejména vkládání skriptů a iframů do webů to umožnilo sledovat uživatele skrze cookies napříč weby, což umožnilo sledování historie prohlížení části webů a remarketing – nabízení produktů z obchodu, který uživatel na webu navštívil (Scott, 2020). To byl dlouho hlavní způsob profilování uživatelů pro reklamní účely, postupně byly možnosti cookies právně limitovány, a začaly se používat další techniky – *device a browser fingerprinting*, *tracking pixels* nebo *behavioral profiling*.

Jednotlivé technologie jsou více popsány v kapitole Hlavní zdroje dat digitální stopy uživatele.

1.5 Možnosti analýzy dat

Se zvyšujícím se množstvím veřejně dostupných osobních dat je snazší vytvářet statistické modely či samoučící systémy, které na základě dat dokáží identifikovat osobu, odhadovat její pohlaví, věk, náboženské a politické přesvědčení, charakterové rysy nebo sexualitu, jak ukázalo mnoho výzkumů (Azucar et al., 2018), a prakticky navázala například Cambridge Analytica (kterou popisují víc v pozdější kapitole). Je očividné, že to jsou často velmi citlivá data, která mohou být potenciálně zneužita mnoha způsoby a uživatel si nemusí být vůbec vědom, že si o něm takovouto představu někdo dělá a používá ji pro (například) nastavení zobrazovaného obsahu a reklam.

1.6 Povědomí o problematice

Důležitou otázkou je, jaké je aktuálně povědomí lidí o této realitě. To není snadná otázka, i z toho důvodu, že se oblast rychle mění a vyvíjí. Ovšem nějakou představu si udělat můžeme.

Podle PEW Research je 59% dotázaných občanů USA podle svého mínění nedostatečně informováno o tom, jak s daty společnosti nakládají, a zároveň má 81% pocit nedostatečné možnosti kontroly, a shodné procento vnímá rizika s tím spojená z vyšší než benefits (Research, 2019).

Studie (Hinds et al., 2020) zkoumající povědomí a změny ve vnímání digitálního soukromí mezi studenty po medializované kauze firmy Cambridge Analytica, dochází k závěrům, že:

- lidé málo vnímají dopady související s *networked privacy* – tedy uvědomění možností získávání a agregace dat o osobně z dat jiných osob.
- lidé si myslí, že jsou imunní vůči cílené reklamě, a tedy nemají obavu z toho, že budou data o nich využita k ovlivnění jich samotných
- z toho plynoucí závěr, že lidé nemohou dělat opravdu uvědomělé závěry, pokud mají takto nepřesnou představu

Dalším důležitým faktorem může být *privacy fatigue* (do češtiny lze zhruba přeložit jako "únava z tématu soukromí") (Choi et al., 2018). Ta plyne z pocitu ztráty kontroly, nepřehlednosti celého tématu a pocitu marnosti z dalších a dalších úniků dat a prolomení soukromí. Studie ukazuje, že tato únava vede k menšímu zájmu o téma (*behavioral disengagement*) a menší ochotě až úplné rezignaci na změnu vlastního chování.

1.7 Shrnutí

V této kapitole jsme si udělali základní přehled o problematice osobních dat a pojmech s ní spojených. Jedním z důležitých poznání je fakt, že pro mnoho lidí je tato problematika složitá (až do stavu, kdy rezignují na pokus se v ní orientovat), nemají pocit informovanosti, zároveň si nejsou vědomi všech aspektů spojených se soukromím a jejího narušení v souvislosti se sdílením osobních dat. To vede k přesvědčení, že dává smysl vytvořit nástroje pro větší informovanost a pochopení problematiky.

Kapitola 2

Zařazení problematiky digitální stopy do školních výukových materiálů

2.1 Úvod a vymezení pojmů

Cílem této kapitoly je dát téma osobních dat a digitální stopy do kontextu vzdělávacích materiálů a východisek, na to, jakou roli a prostor v nich aktuálně (v kontextu narůstající důležitosti tématu ve společnosti a společenských debatách) zaujímá.

Pro snazší orientaci v kapitole zde shrnuji použité zkratky a kódy a jejich význam:

RVP – Rámcový vzdělávací plán, definován (NÚV, b) je následovně:

Rámcové vzdělávací programy (RVP) tvoří obecně závazný rámec pro tvorbu školních vzdělávacích programů škol všech oborů vzdělání v předškolním, základním, základním uměleckém, jazykovém a středním vzdělávání. Do vzdělávání v České republice byly zavedeny zákonem č. 561/2004 Sb., o předškolním, základním, středním, vyšším odborném a jiném vzdělávání (školský zákon).

Klasifikace SŠ (dle (NÚV, c)):

Obory s maturitou

SŠ (M) – úplné střední odborné vzdělání s maturitou (obory kategorie M)

příprava má profesní charakter a délka studia je 4 roky. Po maturitě lze pokračovat ve vzdělávání na vysoké nebo vyšší odborné škole.

SŠ (L) – úplné střední odborné vzdělání s odborným výcvikem a maturitou (obory kategorie L)

studium připravuje pro náročná dělnická povolání a nižší řídicí funkce. V denní formě je 4leté a jeho významnou součástí je odborný výcvik (obory vznikly z dřívějších 3letých učeb-

ních oborů). Absolventi získávají maturitní vysvědčení a mohou pokračovat ve vzdělávání na vysoké nebo vyšší odborné škole.

SŠ (K) – úplné střední všeobecné vzdělání (obory kategorie K)

všeobecná příprava ve 4letých a víceletých gymnáziích je neprofesní a připravuje především pro vysokoškolské nebo vyšší odborné vzdělávání.

Obory s výučním listem

SŠ (H) – střední odborné vzdělání s výučním listem (obory kategorie H)

tradiční učební obory s tříletou přípravou ve středních odborných učilištích. Po získání výučního listu lze pokračovat navazujícím nástavbovým studiem a získat i maturitu.

SŠ (E) – nižší střední odborné vzdělání (obory kategorie E)

studium je tříleté nebo dvouleté, výstupem je výuční list. Obory mají nižší nároky v oblasti všeobecného i obecně odborného vzdělání a jsou určeny především pro žáky se speciálními vzdělávacími potřebami, např. pro absolventy dřívějších speciálních základních škol a žáky, kteří ukončili povinnou školní docházku v nižším než 9. ročníku základní školy. Obory připravují pro výkon jednoduchých prací v rámci dělnických povolání a ve službách.

2.2 Vymezení dokumentů

Pro prozkoumání, jak jsou témata zařazena ve vzdělání, pro nás budou primárním zdrojem zejména Rámcové vzdělávací plány (dále RVP). Je nutné se dívat na jejich návaznost, podobnosti a rozdíly na různých stupních a zaměřeních vzdělávání.

Konkrétně se tedy podíváme na RVP pro základní vzdělávání, které definují vzdělávací oblast Informatiky a její cíle, a dále budeme zkoumat RVP pro gymnázia a vybrané odborné školy se zaměřením na Informatiku a Informační technologie.

Kromě toho rozebereme aktuální plány na revizi oblasti Informatiky a ICT, která může přinášet změny i v této oblasti, a zároveň ukazovat na možné trendy v pojetí vzdělávání v oblasti digitálního světa a technologií a s tím souvisejících témat.

Konkrétně tedy v kapitole prozkoumáme propojení na následující dokumenty:

- Rámcové vzdělávací plány
 - RVP základní školy (NÚV, 2021b)

- RVP-G (kategorie L) (NÚV, 2021a)
- RVP pro odborné školy – Informační technologie (MŠMT, b) (kategorie M)
- RVP pro odborné školy – Informační služby (MŠMT, a) (kategorie M)
- Návrh revizí rámcových vzdělávacích programů v oblasti informatiky a informačních a komunikačních technologií (NÚV, a)

2.3 Rámcové vzdělávací plány

2.3.1 RVP základní školy

Vzdělávací oblast Informatika

Ve vymezení Cílového zaměření vzdělávací oblasti je v kontextu tématu důležitý tento bod

uvědomění si, respektování a zmírnění negativních vlivů moderních informačních a komunikačních technologií na společnost a na zdraví člověka, ke znalosti způsobů prevence a ochrany před zneužitím a omezováním osobní svobody člověka

konkrétně však tento bod není vymezen cíli ani učivem, které by s oblastí osobních dat a digitální stopy přímo souviselo.

2.3.2 RVP-G

Vzdělávací oblast Informatika a informační a komunikační technologie

Vzdělávací oblast Informatika a informační a komunikační technologie v RVP pro vyšší stupně vzdělávání navazuje na oblast Informatika v RVP pro základní školy. V cílovém zaměření vzdělávací oblasti se tedy nachází totožný bod:

uvědomění si, respektování a zmírnění negativních vlivů moderních informačních a komunikačních technologií na společnost a na zdraví člověka, ke znalosti způsobů prevence a ochrany před zneužitím a omezováním osobní svobody člověka

V oblasti Zdroje a vyhledávání informací, Komunikace je v učivu bod

informační etika, legislativa – ochrana autorských práv a osobních údajů

navazující na výstup

využívá informační a komunikační služby v souladu s etickými, bezpečnostními a legislativními požadavky

Alespoň částečně tedy jde o problematiku osobních údajů, ačkoli více z pohledu legislativního a etického než z pohledu soukromí.

V dalších oblastech se pak téma neobjevuje.

2.3.3 RVP odborné školy – Informační technologie

V cílech vzdělávání můžeme najít následující body (zvýraznění vlastní)

neohrožovali svým chováním v digitálním prostředí sebe, druhé, ani technologie samotné

uvědomovali si, že technologie ovlivňují společnost, a naopak chápali svou odpovědnost při používání technologií

Ve výsledcích vzdělávání pak najdeme konkrétní body

chrání digitální zařízení, digitální obsah i **osobní údaje** v digitálním prostředí před poškozením, přepisem/změnou či **zneužitím**; reaguje na změny v technologiích ovlivňujících bezpečnost

s vědomím souvislostí fyzického a digitálního světa **vytváří a spravuje jednu či více digitálních identit**; **kontroluje svou digitální stopu**, ať už ji vytváří sám nebo někdo jiný, v případě potřeby dokáže používat služby internetu anonymně

2.3.4 RVP odborné školy – Informační služby

V tomto RVP se žádné body přímo související s tematikou nenachází.

2.4 Návrh revizí RVP v oblasti Informatiky a informačních a komunikačních technologií

Aktuálně se pracuje na revizi RVP v oblasti Informatiky a informačních a komunikačních technologií, je tedy vhodné se podívat, zda tato revize nějak mění zakotvení tématu digitální stopy a osobních dat ve vzdělávání.

Jedno ze základních východisek návrhu revize je rozvoj digitální gramotnosti, v dokumentu definované jako

Digitální gramotností rozumíme soubor digitálních kompetencí (vědomostí, dovedností, postojů, hodnot), které jedinec potřebuje k bezpečnému, sebejistému, kritickému a tvořivému využívání digitálních technologií při práci, při učení, ve volném čase i při svém zapojení do společenského života.

V oblastech digitální gramotnosti může být pak pro naše téma relevantní bod

Vnímá a hodnotí potenciál i rizika zapojení digitálních technologií do různých procesů a v různých situacích a podle toho zodpovědně jedná.

Jak se to promítá přímo do očekávaných výstupů můžeme vidět v tabulce níže:

Na všechny typech SŠ nacházíme výstup

chrání digitální zřízení, digitální obsah i **osobní údaje** v digitálním prostředí před poškozením či zneužitím

A pro školy v kategoriích K, L, M, H dále

kontroluje svou **digitální stopu**, ať už ji vytváří sám nebo někdo jiný, dokáže používat služby internetu anonymně

a pro SŠ (E) podobný bod

buduje svou digitální identitu a zajímá se, jak k ní přispívají ostatní

2.5 Shrnutí

Téma osobních dat a digitální stopy můžeme v aktuálních vzdělávacích materiálech najít v obecném vymezení v cíli vzdělávací oblasti Informatika (respektive Informatika a informační a komunikační technologie) v bodě

uvědomění si, respektování a zmírnění negativních vlivů moderních informačních a komunikačních technologií na společnost a na zdraví člověka, ke znalosti způsobů prevence a ochrany před zneužitím a omezováním osobní svobody člověka

Konkrétní vzdělávací výsledky s tímto tématem jsme našli pouze u RVP odborných škol oboru Informační technologie

Zároveň se však toto téma výrazněji objevuje v návrhu revize, a tedy můžeme předpokládat, že se do RVP (a následně ŠVP jednotlivých škol) bude více promítat.

Vznik materiálů a prostředí pro zařazení této problematiky do výuky tedy považuji za užitečný a do budoucna nutný.

Kapitola 3

Rizikové dopady sběru osobních dat

3.1 Úvod

Tato kapitola se pokouší zodpovědět na otázku, proč je téma osobních dat pro jednotlivce důležité, tedy „Proč by mě to mělo zajímat?“. Poukazuje na obecná i konkrétní rizika a problémy, které se sběrem osobních dat v dříve neuskutečnitelném měřítku souvisí.

3.2 Změna chování na základě vědomí vytváření stopy

Jedním z diskutovaných témat je riziko změny chování na základě vědomí toho, že naše chování je sledováno, ukládáno a potenciálně analyzováno (Bryce – Klang, 2009).

Aktuálně ještě tato oblast není příliš prozkoumána, je tedy těžké říct, jak velké změny chování může nastávat. Pokusím se v této sekci načrtnout úvahy o těchto dopadech.

Jeden ze směrů úvah o změnách chování vychází z modelu *Panopticonu* (Foucault, 1980). Panopticon je popsán jako vězení, kde může být každý vězeň kdykoli sledován strážným. Protože vězeň neví, jestli zrovna je nebo není sledován, chová se, jako by byl sledován stále. Je tedy vytvořeno prostředí, kde se člověk stále cítí pod dohledem, a reguluje na základě toho svoje chování.

Tento teoretický model se v současnosti stává realitou skrze Čínský kreditový systém – systém, který každému Čínskému občanovi určuje *sociální skóre* na základě chování. Toto je umožněno enormním propojením mnoha zdrojů dat společností i státu, a právě cílenou snahou o co největší možnosti monitorování občanů. Dokonce můžeme říct, že tento model je zesílenou verzí Foucaultova Panopticonu – na rozdíl od něj tu má opravdu docházet ke stálému sledování a zpracování těchto dat (například pomocí systémů umělé inteligence).

Dalším zajímavým aspektem, vyplývajícím z aktuálního množství dat a networked privacy, je fakt, že i cílené snahy o nevytváření digitální stopy (například vypnutím telefonu nebo používáním anonymizačních nástrojů) může být samo o sobě také součástí stopy – například používání prohlížeče zaměřeného na soukromí zároveň zvyšuje rozpoznatelnost uživatele skrze *browser fingerprinting*.

Zároveň se zdá, že (alespoň) mladí lidé kontrolu nad jimi vytvářeným digitálním obsahem více řeší a přizpůsobují tomu své chování (Livingstone, 2008).

3.3 Možnosti zneužití dat

Jedním z problémů týkajících se osobních dat a jejich sběru je možnost jejich zneužití. Pokusíme se na téma nahlédnout ze dvou pohledů – možného zneužití dat zaměřeného na jednotlivce, a rizika zneužití s dopadem na společnost. Je však třeba mít na vědomí, že už jen z pohledu *networked privacy* toto rozdělení nemůže mít ostrou hranici, a je pouze orientační. *Pojem zneužití používám v práci ve smyslu takého využití, jež lze vnímat jako negativní pro daného jedince či skupinu.*

3.3.1 S dopadem na jednotlivce

Přemýšlení o osobních datech primárně vede k uvažování, jestli a jak mohou být zneužita proti mé osobě. Zároveň je častý i pocit, že "nemám co skrývat", a tedy není třeba svou digitální stopu hlídat.

Podíváme se tedy na několik obecných rizik i konkrétních případů zneužití dat.

Vloupání

S příchodem a rozmachem sociálních sítí se začaly objevovat případy vloupání, kdy zloději využívají právě data ze sociálních sítí k jejich naplánování. 75% dopadených pachatelů se domnívá, že jiní pachatelé tato data využívají (Foster, 2015). To není překvapivé ve chvíli, kdy 50% respondentů průzkumu na sociálních sítích sdílí informaci o tom, že jsou na dovolené. (Foster, 2015)

Krádeže identity

Kromě vloupání také narostl počet krádeží identity (ide, 2016), kdy se použitím dat ze sociálních sítí výrazně zjednodušilo používání cizí identity oproti vytváření falešné například pro pojišťovací podvody.

Social engineering

Další možností zneužití dat je jejich použití pro další útoky, jako je *spear phishing* a obecně v oblasti *social engineering*. Jednou z fází social engineering útoků je *information gathering* (Mouton et al., 2016), které sociální sítě výrazně zjednodušily. Spear phishing útoky jsou často zaměřené na firmy, nástrojem pro sběr informací pak může být například profesní sociální síť LinkedIn (Beckers et al., 2017), nebo pomocí nalezení specifických zájmů osoby, skrze které dojde k navázání kontaktu a získání důvěry (Hahnagy, 2011)

Neoprávněné sledování

Objevily se případy, kdy byly lokační data služby použity ke sledování osoby. V roce 2014 se ukázalo, že zaměstnanci společnosti Uber měli možnost (a tuto možnost využívali) sledovat lokace pasažérů – například známé osobnosti, novináře nebo svou bývalou přítelkyni (Morgan, 2017)

Neoprávněné zjišťování informací

Ve Spojených státech vyšetřování ukázalo, že zaměstnanci policie ve stovkách případů během dvou let nahlížely do osobních záznamů, které nesouvisely s výkonem práce. Šlo například o data bývalých partnerů nebo novináře, který vydal kritický článek o místním policejním oddělení. (Services, 2016)

Šíření fotografií dětí

Výrazně odlišným tématem, který však také ukazuje na rizika sdílení různých osobních dat, jsou opakované případy, kdy se fotografie dětí, sdílené jejich rodiči na sociálních sítích nebo sdílecích službách (ulozto.cz, rajce.cz), dostanou do oběhu ve skupinách pedofilů (Juna – Burýšek, 2020).

Úniky dat

Už běžnou součástí digitálního světa jsou úniky dat – jen v roce 2020 bylo zveřejněno téměř 4000 případů (Security, 2021). Tato data mohou být riziková z mnoha pohledů, výraznou ukázkou je únik dat služby Ashley Madison – seznamovací aplikace pro zadané. Tato data byla využita pro vydírání uživatelů skrze e-maily obsahující citlivá data a vyhrožující jejich zveřejněním rodině a na sociálních sítích. (Doffman, 2020)

Lokační data

Jedním z cenných druhů dat jsou lokační data. Při používání smartphonů v podstatě není možné si být jistý, že tato data nejsou sbírána. V čem může být tento sběr rizikový, upozorňuje například článek The New York Times (Thompson – Warzel, 2019), který ukazuje, jak je z lokačních dat a částečné znalosti o pohybu (například veřejně známé osoby) možné dosledovat celou její historii pohybu – tedy potenciálně například to, s kým a kde se setkávala, nebo kde bydlí (například u celebrit, které to cíleně tají).

Další případy – ve spojení s konkrétními postupy pro jejich zamezení – jsou rozebrány v kapitole Klíčové zásady uživatelské ochrany osobních dat.

Je možné si všimnout, že v několika z těchto případů se objevuje zmiňovaný princip *networked privacy* – tedy osoba, která data sdílí nemusí být stejná jako ta, která je terčem zneužití.

3.3.2 S dopadem na společnost

Henrik Skaug Sætra tvrdí, že nahlížení na soukromí pouze z pohledu jednotlivce a jeho svobodné volby se svými daty a soukromím naložit podle svého uvážení, není dostatečný (Sætra, 2020). Argumentuje tím, že volba jednotlivce v má tomto případě dopady i na ostatní – například na jejich možnosti uchování soukromí:

(...) it is impossible for me to be fully unknown in a world where everyone else is fully known.

Rozhodování jednotlivců pak vede k suboptimálním výsledkům pro společnost jako celek.

Soukromí pak nazývá *agregovaným veřejným statkem*, tedy něčím, co je poskytováno členům společnosti, a zároveň vyplývá ze společné aktivity většiny členů.

Tento pohled se dá stručně ukázat na příkladech konkrétních momentů, kdy se téma objevilo ve společenském povědomí.

Cambridge Analytica

Případ firmy Cambridge Analytica, která získala a zpracovala data z více než 50 milionů profilů na Facebooku (Cadwalladr – Graham-Harrison, 2018) pro použití v profilované politické reklamě, je pravděpodobně dostud největší kauza, otevírající téma soukromí a osobních dat. Ač se po letech vyšetřování ukazuje, že byl pravděpodobně vliv menší, než se zdálo (BBC, 2020) (Rathi, 2019), vyvolala mnoho otázek mimojiné o tom, jak mohou být osobní data využita pro politickou kampaň v dosud neviděném měřítku.

Čínský sběr dat

Česku bližší případ práce s osobními daty s potenciálním vlivem na celou společnost, je firma Zhenhua Data Technology, která vytvářela na základě získaných veřejných a komerčně dostupných dat profily strategicky významných osob české společnosti. Šlo o politiky, pracovníky bezpečnostních složek či podnikatele (Valášek – Horák, 2020). Česko bylo samozřejmě jenom jednou z mnoha zemí, o kterých tento sběr a analýza probíhaly, ukázal ale jasněji, jak může být s daty nakládáno. Zajímavé je, že i v tomto případě byla část dat z automatizovaného sběru ze sociální sítě Facebook, která tuto praxi zakazuje. Firma Zhenhua byl pak dle mluvčí Facebooku ze sítě zablokována, ovšem ukazuje to na fakt, že ani po kauze Cambridge Analytica se reálná ochrana nakládání osobními daty u této firmy příliš nezměnila.

Citlivá lokační data

Příkladem ohrožení veřejných zájmů může být i případ, kdy fitness aplikace Strava zveřejnila mapu agregovaného pohybu uživatelů. Na mapě se tím objevilo umístění vojenských základen USA, například v Afghánistánu. (Hern, 2018).

3.4 Shrnutí

Kapitola 4

Hlavní zdroje dat digitální stopy uživatele

4.1 Úvod

V této kapitole se podíváme na kategorizace dat podle toho, jak se uživatel podílel na jejich vytvoření, a podle jejich anonymity.

Dále si rozebereme konkrétní typy dat pro úplnější představu o tom, co všechno je ukládáno a zpracováváno.

Tato data budou užitečná v praktické části práce pro vytváření scénářů a modelových dat podle reálných typů dat a možnostech přístupu k nim.

4.2 Kategorie dat

Data, která vytváříme používáním technologií a digitálních služeb, můžeme kategorizovat několik způsoby. Dělení může být například

- **aktivní** vs **pasivní** digitální stopu (Madden et al., 2007)
- případně doplněna o **vědomě nevědomou** (Fish, 2009)
- **identifikovatelné** vs **anonymní**

4.2.1 Aktivní, pasivní a vědomě nevědomá digitální stopa

Aktivní stopa

Aktivní stopou se myslí takový typ dat, který uživatel vědomě publikuje. Může jít o příspěvky, komentáře a reakce na sociálních sítích či osobních webech, fotky a jiné soubory nahrané na cloudová úložiště, nebo vytvořené uživatelské účty.(Madden et al., 2007)

V tomto pohledu nerozlišujeme, kdo je majitelem dat a jak s nimi kdo může nakládat, to závisí na podmínkách konkrétní platformy (omezené legislativou dané země).

Pasivní

Pasivní stopa se skládá z typu dat, které uživatel vytváří svým používáním digitální platformy či služby, bez přímého sdílení či nahrávání nějakých svých dat. Obvykle jde o analytická data, používána pro lepší technický, bezpečnostní nebo marketingo-ekonomický efekt. Může jít o informace o zobrazení stránky či příspěvku, IP adresu a další technické parametry připojeného uživatele (například lokaci). Může jít i o kombinovaná data, například určení zájmů či demografické skupiny, vytvořené na základě jednotlivých dat.

Vědomě nevědomá

Tony Fish přidává kategorii vědomě nevědomé digitální stopy, která se skládá z dat aktivně vložených jinými uživateli (Fish, 2009). Může jít o fotografie – na sociálních sítích, ale například i z různých akcí (kde člověk zveřejnění dat musí v České Republice odsouhlasit) nebo o data zveřejněná úřady

4.2.2 Identifikovatelná a anonymní data

Identifikovatelná

Sem můžeme řadit typ dat, které jsou přímo propojitelné s naší osobou. Jde pak o osobní údaje podle definice GDPR:

Personal data means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. (Council of European Union, 2016)

Anonymní

Může jít o data anonymizovaná službou či mezivrstvou, nebo také o data vytvořená uživatelem s použitím anonymizačních nástrojů, jako je například VPN (virtuální privátní síť, používána pro skrytí IP adresy uživatele).

4.3 Konkrétní typy dat

4.3.1 Historie prohlížení

Jednou z významných součástí pasivní digitální stopy je historie prohlížení – tedy záznam všech námi navštívených webových stránek (případně všech webových requestů).

Samostatně jde o data anonymní (nejsou přímo spojená s naší osobou), ovšem v realitě to tak nemusí být.

Tato data může uchovávat internetová prohlížeč, a to buď lokálně nebo na cloudu. Například v případě prohlížeče Google Chrome se zapnutou synchronizací jsou tyto informace ukládány jako součást dat Google profilu.

Data o webové aktivitě má také poskytovatel internetového připojení (Internet Service Provider – ISP), který je zároveň má spojená s naší IP adresou, která je vázaná na konkrétní smlouvu o poskytování internetu. V České Republice si tato data může vyžádat policie, a prokazatelně to dělá (Vokuš, 2019).

Kromě toho je možné aktivitu uživatelů napříč weby sledovat pomocí cookies třetích stran a různých typů fingerprintingu.

Cookies

Jak bylo zmíněno v kapitole 2, cookies jsou soubory, které si stránka ukládá do počítače uživatele, aby ho mohla identifikovat při dalších požadavcích. V současnosti mnoho webů obsahuje takzvané *cookies třetích stran*, které umožňují službám (například Google Analytics a Google Ads) sledovat aktivitu napříč webem.

Device a browser fingerprinting

S postupným legislativním tlakem na omezení rozsahu cookies se začaly služby přesouvat k používání *fingerprintingu*, tedy požívání jakéhosi otisku zařízení nebo prohlížeče, ze kterého uživatel přistupuje. Používané techniky jsou rozmanité, od získávání informací o prohlížeči a operačním systému (verze, jazyk, instalované doplňky), po *canvas fingerprint* využívající specifika v renderování webového prvku *canvas*, které se liší podle GPU nebo grafických ovladačů v daném zařízení.

Zjistit svůj browser fingerprint lze například na stránce <https://amiunique.org/fp>. Služba ukazuje, kolik informací je prohlížeč schopen získat, a to i bez schválení uživatelem.

Behavioral profiling

Relativně nově používanou technikou je *Behavioral profiling*, který se snaží rozlišit uživatele podle jejich chování – primárně charakteristik pohybu myši nebo způsobu psaní na klávesnici (Yang, 2010) (Jorgensen – Yu, 2011) (Ikuesan – Venter, 2019).

Tento způsob je stále ještě méně spolehlivý než dříve zmíněné, neboť využívá výrazně rozmanitější typ dat, ale v kombinaci s jinými nástroji, a s rostoucími možnostmi vyhodnocování těchto dat (například využitím strojového učení) se stává nezanedbatelnou možností v repertoáru nástrojů pro profilování a identifikaci uživatele.

4.3.2 Lokační data

Jak bylo zmíněno v kapitole Rizikové dopady sběru osobních dat, jedním z typů vyhledávaných dat jsou data lokační.

Tato data mohou například poskytnout vhled, do jakých obchodů uživatel chodí, a nabídnout tak *remarketing* napříč fyzickým a digitálním světem.

Lokační data je v současné době, kdy většina lidí vlastní smartphone, snadné získat. I aplikace, které reálně data pro svoje funkce nepotřebují, o ně mohou uživatele požádat. Prodejem těchto dat pak mohou získat finance na samotný vývoj aplikace „zdarma“.

Zde je na místě říct, že možnosti prodeje dat se mohou výrazně lišit podle toho, v jaké zemi – a tedy pod jakou legislativou – se uživatel pohybuje.

Lokační data dále potenciálně sdílejí aplikace, které reálně data používají pro svoje služby – mapy (*Google Maps*, *Mapy.cz* a podobné), dopravy (*IDOS*, *Pubtran*, *PID Lítačka* a podobné) a dovozy (*DámeJídlo*, *Uber* a podobné).

Například *Google Maps* uchovává lokaci a přiřazuje ji ke konkrétním místům (obchody, úřady a podobně). Zároveň uchovává informaci o předpokládané formě aktivity (např. chůze, běh, kolo, auto, hromadná doprava) včetně důvěry (pravděpodobnosti) v predikci.

Mohlo by se zdát, že lokační data jsou anonymní (typicky obsahují nějaké náhodné ID zařízení, časovou značku a lokaci), v realitě jsou však velmi snadno deanonymizována:

Describing location data as anonymous is “a completely false claim” that has been debunked in multiple studies, Paul Ohm, a law professor and privacy researcher at the Georgetown University Law Center, told us. “Really precise, longitudinal geolocation information is absolutely impossible to anonymize.”

“D.N.A.,” he added, “is probably the only thing that’s harder to anonymize than precise geolocation information” (Thompson – Warzel, 2019).

4.3.3 Data vkládaná na sociální sítě a další platformy

Nejvýznamnější složkou aktivní (veřejné) digitální stopy jsou pravděpodobně v současné době sociální sítě a obsah na nich. Jde o příspěvky, ale i komentáře, reakce (emotikony), projevení zájmu o událost a další vědomé interakce, jejichž výstupem je informace sdílená s dalšími uživateli.

Kromě toho, že jsou součástí celkového „balíku“ dat, které služby mají, může být jejich riziko i v nastavení veřejnosti, ať už uživatelem, nebo úmyslem či chybou dané služby – například v roce 2018 byly příspěvky přibližně 14 milionů uživatelů veřejné místo soukromých (Frenkel, 2018).

A zároveň, jak jsem uvedl již v kapitole *Problematika*, jsou to data dobře využitelná do modelů, které dokáží o uživateli odvodit i informace, které vědomě nepíše, a která jsou mnohdy velmi citlivá.

Fotografie

Specifickou částí zveřejňovaných dat jsou fotografie. U nich může často docházet ke vzniku *vědomě nevědomé* digitální stopy, když se objevujeme na fotkách, které sdílí někdo jiný (a naopak, pokud my sdílíme fotografie, na kterých nejsme jen my). Pokrok ve strojové analýze obrazu už umožňuje, aby byly osoby na fotografii rozpoznány automaticky (Facebook nám díky tomu může u nahrané fotky navrhnout, koho bychom mohli označit, Google Photos umožňuje zobrazit všechny naše fotky dané osoby s velmi vysokou přesností). Zároveň je často možné z fotografie získat další informace z *metadat*. To jsou data, která vytváří fotoaparát (dnes obvykle smartphone), a můžou obsahovat informace jako čas a datum, lokaci, název zařízení nebo technické parametry jako délku expozice či ISO. Některé služby (například Facebook) při nahrávání fotografie citlivá metadata smaže, aby nebylo možné uložením fotografie získat i je, ale uživatel s tím při nahrávání fotografie obecně počítat nemůže. Kromě toho se s rozvojem strojového učení ukazuje, že bude nejspíš možné lokaci často velice přesně určit i bez lokačních metadat (Neil, 2016), a to až s přesností ulice či domu.

4.3.4 Další data sledovaná sociálními sítěmi a dalšími platformami

Kromě dat, které na sociální sítě uživatel vědomě vkládá, tyto služby sbírají a vytváří další data, napojená na náš profil. Alespoň část z nich můžeme vidět, když si od daných služeb svoje data vyžádáme (na což máme dle GDPR nárok). V datech z Facebooku je možné vidět například tyto typy dat, které jako uživatel typicky nevím, že Facebook uchovává (nejsou nikde zobrazována) a nejsou třeba pro uživatelskou funkcionalitu služby:

- **friend_peer_group** – *Popis životní etapy vašich přátel na Facebooku*

Facebook si na základě dat o uživateli jeho profil zařadí do některé z předem definovaných skupin, například *Začínající dospělý život*, a tu pak využívá pro cílení reklamy a celkové nastavení toho, co uživatel vidí.

- **viewed** Facebook sleduje ve speciální kategorii
 - jaká videa, jakou část z nich, a kolik času celkově jsem strávil u videí z Facebook View
 - jaké zboží na Facebook Marketplace jsem si prohlížel
 - na jaké zobrazené reklamy jsem reagoval

Celkově zde jde o data, která jsou velmi cenná z pohledu reklams, monetizace a udržení uživatele ve službě.

- **ads_interests** Facebook k uživateli přiřazuje zájmy, které jsou pak využívány na reklamních aukcích.
- **unfollowed_pages** Facebook zachovává historii toho, jaké stránky uživatel přestal sledovat.
- **removed_friends** Stejně tak uchovává informaci o ukončených „přátelstvích“.
- **group_interactions** Počet interakcí (příspěvků, komentářů, reakcí) ve skupinách, jejichž je uživatel členem.
- **people_interactions** Počet interakcí s osobními profily.
- **your_topics** – *Sbírka témat, které určuje vaše aktivita na Facebooku a na jejímž základě se vytvářejí doporučení pro vás v různých částech Facebooku, jako je třeba kanál vybraých příspěvků, Zpravodajství nebo Watch.* Podobně jak ads_topics – seznam témat, jež se Facebook domnívá, že uživatele zajímají, a podle nich přizpůsobuje zobrazovaný obsah.

Podobně je možné se podívat na některá data, která vytváří a uchovává Google (samozřejmě podle toho, které služby uživatel používá). Opět zde uvedu některá překvapivější, která se nezdaří nutná pro základní funkcionalitu (v této sekci uvádím data, která nelze zařadit do jiné sekce). U každého bodu pak uvedu (pokud to bude možné), skrze jakou službu tato data vznikají .

- **názvy tisknutých souborů**

Služba *Google Cloud Print*, umožňující tisk přes internet, uchovává seznam s názvem tisknutých souborů.

- **informace o hrách**

O uživatelských službách *Google Play Games* jsou uchovávána například data o tom, kdy poprvé a kdy naposledy hráli danou hru, a souhrnné herní statistiky. Kromě toho jsou v jiné sekci uloženy všechny záznamy spuštění her.

- **hudba**

Služba *Google Play Music* uchovává mimojiné informaci o počtu přehrání jednotlivých skladeb, a seznam všech spuštění nahrávek s jejich časovou značkou. To platí i u služby *Google Podcasts*.

- **aplikace**

Služba *Google Store* – tedy nejpoužívanější služba pro správu a instalaci aplikací na zařízeních s operačním systémem *Android* – kromě seznamu někdy instalovaných aplikací zaznamenává údaje o zařízení (konkrétní model telefonu), na kterých byla instalována a telefonního operátora. Dále je uložena informace o každé instalaci a odinstalaci aplikací.

- **reklamy**

Google uchovává informaci o všech reakcích (kliknutích) uživatele na zobrazené reklamy

- **Android**

Seznam všech otevření aplikací. Vztahuje se na uživatele používající operační systém *Android* s přihlášeným profilem Google.

- **hlasové pokyny**

Hlasové pokyny pro službu *Google Assistant* jsou ukládány jako originální zvukové nahrávky. To se děje i v případech, kdy začalo zařízení poslouchat „omylem“ – kdy uživatel neřekl klíčové slovo pro začátek poslechu, ale jiný zvuk tak byl mylně interpretován.

- **vyhledávání**

Služba *Google Search* uchovává informaci o každém vyhledávání a o tom, kam z něj uživatel pokračoval

4.3.5 E-mail

Velká část online komunikace se stále děje přes e-mailové služby, zejména v případě komunikace pracovní. Ta může obsahovat mnoho (zejména pro firmu) citlivých informací, a jejich únik může mít velké následky (jak je vidět na příkladu uniklých e-mailů Hillary Clinton v prezidentské volbě v USA v roce 2016).

Kromě samotného obsahu e-mailu se dnes používají technologie, které umožňují sledovat uživatelskou interakci s přijatým e-mailem. Samotná technologie e-mailu toto nenabízí (oproti různým chatovacím službám, nebo protokolu RCS, postupně nahrazujícímu SMS). Často používanou technikou je *pixel tracking*, který funguje pomocí vložení „neviditelného“ obrázku o velikosti jednoho pixelu, který se při otevření nahrává z unikátní adresy, a tento dotaz na danou adresu je možné zaznamenat. U tom, že je otevření e-mailu sledováno, obvykle uživatel vůbec neví.

4.3.6 Finanční záznamy

V době, kdy velká většina finančních transakcí neprobíhá s fyzickými penězi (zejména v České Republice, která je k adopci nových technologií v oblasti převodu peněz velmi otevřená), se vytváří další část digitální stopy právě v této oblasti. Primárně mají veškeré záznamy k dispozici banky, u kterých transakce probíhají. Tyto informace podléhají bankovnímu tajemství, může si je však v oprávněných případech (dle § 8 zákona č. 141/1961 Sb., trestní řád) vyžádat policie, a to bez vědomí osoby, jejíž informace jsou nahlíženy. Výpisy z účtu mohou být vyžadovány i v jiných případech, třeba při bezpečnostní prověrce.

Znatelná část finančních transakcí se rovněž přesouvá do plateb telefonem, které probíhají přes dalšího prostředníka – data o plabách tedy nemá již jen banka, ale i další soukromá firma. Často používanými nástroji pro tyto platby jsou Alipay a WeChat (obojí primárně v Číně), Apple Pay, PayPal, Samsung

Pay, Amazon Pay nebo Google Pay. Většinou tedy jde o firmy, které poskytují i další služby a historie transakcí pak může být dalším zdrojem dat o uživateli.

4.3.7 Zdravotní data

Další oblastí velmi citlivých dat jsou data zdravotní. S postupnou digitalizací zdravotnictví bude možné tato data čím dál více propojovat, což nese jak nesporné výhody ve snížení zdravotnické administrativy a urychlení péče, tak ale potenciální rizika v případě úniku či zneužití těchto dat.

Kromě dat ve zdravotnictví se čím dál rozmanitější typy dat sbírají skrze *wearables* – nositelnou elektroniku, často nabízející různé fitness a zdravotní funkce. Aktuálně se tedy můžeme setkat se sběrem dat o pohybu, měřením tepu a dechu, až po měření odporu kůže či množství kyslíku v krvi. Tato data jsou pak automaticky zpracovávána pro rozpoznávání typu a intenzity pohybu či sportu, doby a kvality spánku, množství stresu a další.

Dá se očekávat, že se budou objevovat další typy sledovaných dat, sensory budou přesnější a z dat (a to i historických) bude možné vyčíst více informací – probíhají mnohé výzkumy, které se snaží v datech najít návaznost na různé zdravotní jevy, příkladem může být studie propojující informace o variabilitě tepu s psychickým i celkovým zdravím (Coutts et al., 2020).

Tato data typicky opět sbírají soukromé firmy, často nabízející mnoho dalších služeb – uvést můžeme Google (který v roce 2019 koupil firmu Fitbit, výrazného výrobce fitness zařízení), Apple či Samsung.

Ještě problematičtější částí může být v posledních letech velmi populární analýza DNA soukromými firmami. DNA je ze své podstaty extrémě osobní informace, a ani tyto firmy nejsou odolné vůči únikům dat, jak je vidět na případu firmy MyHeritage, které v roce 2018 unikly data 92 milionů účtů. V tomto případě nedošlo k úniku genetických dat, to je ovšem pravděpodobně jen otázka času. Aktuálně jsou genetická data užitečná zejména pro výzkum v oblasti medicíny a léčiv, s rostoucím pochopením genomu je těžké odhadnout, jakými způsoby by mohla být tato data využita či zneužita.

Kapitola 5

Klíčové zásady uživatelské ochrany osobních dat

Úvod

5.1 Obecné náhledy

5.1.1 Co je na internetu, to je veřejné

Úniky dat

5.1.2 Budoucnost

5.2 Zásady a nástroje

5.2.1 Omezení vědomého sdílení informací

5.2.2 Anonymní prohlížení webu

Anonymní režim

Prohlížeče

VPN

Doplňky

Do Not Track

Výběr nástrojů

5.2.3 Pravidla pro aplikace a služby³¹

Souhlasy cookies

Kapitola 6

Praktická část

Úvod

Praktická část této bakalářské práce se zabývá návrhem a vytvořením prostředí (dále *Aplikace*) pro vzdělávání této tematiky.

6.1 Východiska

Jak vyplývá z kapitoly *Zařazení problematiky digitální stopy do školních výukových materiálů*, téma digitální stopy a osobních dat se nejspíše bude po revizích více objevovat v RVP středních škol. Na tuto skupinu studentů tedy bude Aplikace cílit.

Prostředí může sloužit jako součást většího vzdělávacího oblouku, ovšem tato práce si nedává za cíl vytvoření metodik pro učitele, a tedy musí obsah Aplikace fungovat i sám o sobě.

Cílem je uživatele (studenty) v Aplikaci seznámit s tématem interaktivní, lákavou formou.

Zároveň však součástí Aplikace musí být kvalitní napojení na teorii i reálné příklady, včetně informování uživatele o tom, jak může svou digitální stopu spravovat.

6.1.1 Návrh aplikace z pohledu uživatele

Návrh aplikace je následující: Uživatel má možnost odehrát jednotlivé „Mise“, které představují fiktivní příběh. V každé Misi má uživatel ze simulovaných osobních dat různého typu vyřešit na začátku danou otázku. Zjednodušený use-case tedy je:

- uživatel si vybírá Misi
- uživatel je seznámen s příběhem Mise (symbolickým rámcem) a cílem - co je potřeba zjistit pro splnění Mise
- uživatel využívá dané datasety pro nalezení řešení

- po správném zadání odpovědi jsou uživateli zobrazeny doplňující informace - napojení na teorii, příklady z reálného světa, možnosti zabezpečení se v podobné situaci.

V sekci *Scénáře* budou popsány konkrétní možné scénáře, z nich část bude použita v prototypu.

6.1.2 Prototyp

Jak bylo řečeno, cílem práce je vytvoření prototypu. Ten tedy nemusí mít všechny funkce či grafické řešení aplikace, která by byla reálně veřejně použita ve vzdělávání, má za cíl pouze najít vhodnou podobu, otestovat její technickou proveditelnost a náročnost, a získat zpětnou vazbu od vybraných testerů.

Prototyp bude obsahovat několik scénářů s různými typy dat.

6.1.3 Scénáře

Jak bylo popsáno, jádrem Aplikace jsou Mise - tedy jednotlivé příběhy, ve kterých se uživatel seznamuje s různými situacemi, týkajícími se osobních dat.

Aplikace bude navržena tak, aby bylo možné snadno další scénáře přidávat, a tím rozšiřovat její vzdělávací potenciál (například v zacílení na jiné věkové skupiny).

Při tvorbě scénářů vycházím z dat a rizik popsaných v předchozích kapitolách a navrhuji následující testovací scénáře, které představují různé typy a kategorie dat.

Uvědomuji si, že některé scénáře ukazují činnost, jež je ilegální. Domnívám se, že možnost prožít si situaci z pohledu útočníka může vést k lepšímu prožití a přenesení do uvažování o vlastní ochraně. Zároveň u každého takového scénáře bude upozorněno, že jde o simulaci a jaké trestněprávní dopady by taková činnost měla v reálném světě.

Odhadnutí hesla

Cíl

Uhodnout heslo blízké osoby na sociální síť.

Používané zdroje dat

Příspěvky na sociální síti.

Typ dat

Veřejné / veřejné pro okruh lidí.

Dodatečné informace

Informace o tom, jak lidé tvoří hesla

Případy toho, kdy lidé měli veřejně informace, jež vedly k prolomení hesla.

Obecná doporučení, jak chránit svoje hesla

Upozornění na nelegálnost takové činnosti v reálním životě

V tomto scénáři je úkolem uhodnout heslo do sociální sítě. Scénář se odkazuje na témata vhodného zabezpečení svých účtů a situací, kdy člověkem sdílené informace napomáhají k prolomení jeho obran.

Jde o jednoduchý, začáteční scénář – pracuje s veřejnými daty z jednoho zdroje. Je cíleně zjednodušený oproti realitě.

Scénář se nezabývá hesly z pohledu kryptografického a obecně bezpečnostního, neboť to je již mimo rozsah naší práce. Bylo by však pravděpodobně možné takový scénář do Aplikace přidat pro vzdělávání v této oblasti.

Plánování vloupání

Cíl

Naplánovat vloupání dané osoby – udělat si představu o jejím bydlišti a době, kdy bude dům prázdný.

Používané zdroje dat

Příspěvky na sociální síti Příspěvky z fitness sociální sítě

Stránky firmy

Typ dat

Veřejná data

Dodatečné informace

Případy z praxe

Obecná doporučení, jak se tomuto typu útoku bránit

Upozornění na nelegálnost takové činnosti v reálním životě

Tento scénář již ukazuje práci s více zdroji dat a klade na uživatele větší nároky v hledání částí informací.

Vyšetřování korupce

Cíl

U podezřelé osoby vyšetřit možné korupční vazby na jiné osoby.

Používané zdroje dat

Lokační data od operátorů

Výpisy z bankovního účtu

Výpisy hovorů a SMS zpráv

Typ dat

Soukromé – dostupné pouze poskytovatelům a v oprávněných případech policii

Dodatečné informace

Informace o sběru a uchovávání dat operátory

Informace o přesnosti lokačních dat

Informace o právních možnostech policie si tato data vyžádat

Výběr vhodné reklamy

Cíl

Vybrat, jaké reklamy zobrazit jakým uživatelům.

Používané zdroje dat

Příspěvky na sociálních sítích

Chování na sociálních sítích

Historie prohlížení

Lokační data

Typ dat

Soukromé – data sbírají aplikace.

Dodatečné informace

Napojení na teorii *attention economy*

Informace o možnostech nastavení ochrany soukromí v různých aplikacích.

Tento scénář simuluje fungování vyhodnocování dat algoritmy firem jako je Facebook nebo Google, a následnou reklamní aukci. Má potenciál, aby na něj bylo navázáno šířeji tématem personalizované reklamy a *attention economy*.

6.2 Uživatelské workflow a use-cases

6.3 Technické řešení

6.3.1 Výběr technologií

Prototyp jsem se rozhodl vytvářet ve formě webové aplikace. Hlavními důvody pro toto rozhodnutí bylo:

- dostupnost pro uživatele i testery
Není potřeba nic instalovat, stačí jakýkoli moderní webový prohlížeč.
- snadné vytvoření prototypu

- snadná rozšiřitelnost a kooperace na projektu

Jako open-source projekt předpokládá možnosti další spolupráce, například s designéry. Úprava vizuálu je v případě webové aplikace čistě HTML+CSS(+JS), což je rozšířená dovednost, oproti jiným grafickým prostředím (jako Qt či herní enginy).

Konkrétně jsem se rozhodl pro využití webového frameworku **Flask** (založeného na jazyce Python). Motivací byla osobní zkušenost s tímto frameworkem, takže jsem měl dostatečnou představu o tom, že je v něm možné prototyp bez problémů vytvořit.

Na frontendové straně jsem se pro potřeby prototypu rozhodl nezavádět žádný rozsáhlý framework a pro někde nutné části Javascriptu jsem použil pouze microframework **stimulus.js** a **jQuery**.

Vývoj je verzován systémem **git** za použití veřejného repozitáře na službě GitHub.

6.3.2 Základní funkcionality

Nahrát předem definovaná data Mise

Zobrazit misi s jednotlivými záložkami

Zobrazovat data různých typů

- Příspěvky na sociálních sítích
- Výměna zpráv mezi dvěma osobami
- Výpis dat
seznam hovorů, výpis z bankovního účtu,...
- Datové body na mapě
- Souhrnné informace
například informace, které uchovávají sociální sítě o jednotlivých uživateli
- webové stránky

6.3.3 Objektová vrstva

6.3.4 Ověření a otestování aplikace

6.3.5 Vyhodnocení provedeného ověření a doporučení pro další rozvoj aplikace

Závěr

Lorem Ipsum

Literatura

- Identity theft rises sharply as fraudsters target social media. *Computer Fraud & Security*. 2016, 2016, 7, s. 1–3. ISSN 1361-3723. doi: [https://doi.org/10.1016/S1361-3723\(16\)30048-3](https://doi.org/10.1016/S1361-3723(16)30048-3).
Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1361372316300483>.
- AZUCAR, D. – MARENGO, D. – SETTANNI, M. Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*. 2018, 124, s. 150–159. ISSN 0191-8869. doi: <https://doi.org/10.1016/j.paid.2017.12.018>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0191886917307328>.
- BBC. *Cambridge Analytica 'not involved' in Brexit referendum, says watchdog* [online]. 2020. [cit. 2021-04-02]. Dostupné z: <https://www.bbc.com/news/uk-politics-54457407>.
- BECKERS, K. et al. A Structured Comparison of Social Engineering Intelligence Gathering Tools. s. 232–246, 08 2017. doi: 10.1007/978-3-319-64483-7_15. ISBN 978-3-319-64482-0.
- BRYCE, J. – KLANG, M. Young people, disclosure of personal information and online privacy: Control, choice and consequences. *Information Security Technical Report*. 2009, 14, 3, s. 160–166. ISSN 1363-4127. doi: <https://doi.org/10.1016/j.istr.2009.10.007>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1363412709000429>. The Changing Shape of Privacy and Consent.
- CADWALLADR, C. – GRAHAM-HARRISON, E. *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach* [online]. 2018. [cit. 2021-03-15]. Dostupné z: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.
- CHOI, H. – PARK, J. – JUNG, Y. The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*. 2018, 81, s. 42–51. ISSN 0747-5632. doi: <https://doi.org/10.1016/j.chb.2017.12.001>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0747563217306817>.
- COUNCIL OF EUROPEAN UNION. Council regulation (EU) no 2016/679, 2016.
<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>.
- COUTTS, L. V. et al. Deep learning with wearable based heart rate variability for prediction of mental and general health. *Journal of Biomedical Informatics*. 2020, 112, s. 103610. ISSN

- 1532-0464. doi: <https://doi.org/10.1016/j.jbi.2020.103610>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1532046420302380>.
- DOFFMAN, Z. *Ashley Madison Hack Returns To ‘Haunt’ Its Victims: 32 Million Users Now Watch And Wait* [online]. 2020. Dostupné z: <https://www.forbes.com/sites/zakdoffman/2020/02/01/ashley-madison-hack-returns-to-haunt-its-victims-32-million-users-now-have-to-watch-and-wait/>.
- EDITORS. *US Digital Ad Spending Will Surpass Traditional in 2019* [online]. 2019. [cit. 2021-03-15]. Dostupné z: <https://www.emarketer.com/content/us-digital-ad-spending-will-surpass-traditional-in-2019>.
- FISH, T. *My digital footprint: a two sided digital business model where your privacy will be someone else’s business*. Futuretext, 2009. OCLC: 639933238. ISBN 9780955606984.
- FOSTER, M. *Is The Way We Use Social Media Leaving Us Vulnerable To Burglary?* [online]. 2015. [cit. 2021-04-03]. Dostupné z: <https://socialmediaweek.org/blog/2015/05/social-media-leaving-us-vulnerable-burglary/>.
- FOUCAULT, M. *Power/knowledge: Selected interviews and other writings, 1972-1977*. Vintage, 1980.
- FRENKEL, S. *Facebook Bug Changed Privacy Settings of Up to 14 Million Users* [online]. 2018. [cit. 2021-04-02]. Dostupné z: <https://www.nytimes.com/2018/06/07/technology/facebook-privacy-bug.html>.
- HADNAGY, C. *Social engineering: the art of human hacking*. Wiley, 2011. OCLC: ocn635494717. ISBN 9780470639535 9781118028018 9781118029718 9781118029749.
- HERN, A. *Fitness tracking app Strava gives away location of secret US army bases* [online]. 2018. [cit. 2021-04-02]. Dostupné z: <https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases>.
- HINDS, J. – WILLIAMS, E. J. – JOINSON, A. N. “It wouldn’t happen to me”: Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*. 2020, 143, s. 102498. ISSN 1071-5819. doi: <https://doi.org/10.1016/j.ijhcs.2020.102498>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1071581920301002>.
- IKUESAN, A. R. – VENTER, H. S. Digital behavioral-fingerprint for user attribution in digital forensics: Are we there yet? *Digital Investigation*. 2019, 30, s. 73–89. ISSN 1742-2876. doi: <https://doi.org/10.1016/j.diin.2019.07.003>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1742287619300945>.
- JORGENSEN, Z. – YU, T. On Mouse Dynamics as a Behavioral Biometric for Authentication. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*,

- ASIACCS '11, s. 476–482, New York, NY, USA, 2011. Association for Computing Machinery. doi: 10.1145/1966913.1966983. Dostupné z: <https://doi.org/10.1145/1966913.1966983>. ISBN 9781450305648.
- JUNA, P. – BURÝŠEK, J. *Ruský web pro pedofily vystavuje kradené fotky českých dětí* [online]. 2020. [cit. 2021-04-02]. Dostupné z: <https://www.seznamzpravy.cz/clanek/rusky-web-pro-pedofily-vystavuje-kradene-fotky-ceskych-deti-89197>.
- KEMP, S. *Digital 2021: Global overview report* [online]. 2019. [cit. 2021-03-15]. Dostupné z: <https://datareportal.com/reports/digital-2021-global-overview-report>.
- LIVINGSTONE, S. Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression. *New Media & Society*. 2008, 10, 3, s. 393–411. doi: 10.1177/1461444808089415. Dostupné z: <https://doi.org/10.1177/1461444808089415>.
- MADDEN, M. et al. *Digital Footprints - Online identity management and search in the age of transparency* [online]. 2007. [cit. 2021-04-02]. Dostupné z: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2007/PIP_Digital_Footprints.pdf.pdf.
- MARWICK, A. E. – BOYD. Networked privacy: How teenagers negotiate context in social media. *New Media & Society*. 2014, 16, 7, s. 1051–1067. doi: 10.1177/1461444814543995. Dostupné z: <https://doi.org/10.1177/1461444814543995>.
- MORGAN, R. *Uber settles federal probe over 'God View' spy software* [online]. 2017. [cit. 2021-04-02]. Dostupné z: <https://nypost.com/2017/08/15/uber-settles-federal-probe-over-god-view-spy-software/>.
- MOUTON, F. – LEENEN, L. – VENTER, H. Social engineering attack examples, templates and scenarios. *Computers & Security*. 2016, 59, s. 186–209. ISSN 0167-4048. doi: <https://doi.org/10.1016/j.cose.2016.03.004>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0167404816300268>.
- MŠMT. *Publicistika, knihovnictví a informatika* [online]. a. [cit. 2021-03-15]. Dostupné z: <https://www.edu.cz/rvp/ramcove-vzdelavaci-programy-stredniho-odborneho-vzdelavani-rvp-sov/obory-l-a-m/72-publicistika-knihovnictvi-a-informatika/>.
- MŠMT. *Informatické obory* [online]. b. [cit. 2021-03-15]. Dostupné z: <https://www.edu.cz/rvp/ramcove-vzdelavaci-programy-stredniho-odborneho-vzdelavani-rvp-sov/obory-l-a-m/18-informaticke-obory/>.

- NEIL, A. *Google AI Can Work Out Photo Locations Without Geotags* [online]. 2016. [cit. 2021-04-02]. Dostupné z: <https://www.eteknix.com/google-ai-can-work-photo-locations-without-geotags/>.
- NÚV. *Návrh revízií ICT* [online]. a. [cit. 2021-03-15]. Dostupné z: <http://www.nuv.cz/file/3362/>.
- NÚV. *Rámcové programy* [online]. b. [cit. 2021-03-15]. Dostupné z: <http://www.nuv.cz/t/rvp>.
- NÚV. *RVP-G* [online]. 2021a. [cit. 2021-03-15]. Dostupné z: <http://www.nuv.cz/file/159>.
- NÚV. *RVP ZŠ* [online]. 2021b. [cit. 2021-03-15]. Dostupné z: <http://www.nuv.cz/file/4983/>.
- NÚV. *Střední vzdělávání* [online]. c. [cit. 2021-03-15]. Dostupné z: <http://www.nuv.cz/t/stredni-vzdelavani>.
- PENLAND, J. *Browser Cookies: What Are They & Why Should You Care?* [online]. 2020. [cit. 2021-03-15]. Dostupné z: <https://www.whoishostingthis.com/resources/cookies-guide/#:~:text=Cookies>.
- RATHI, R. *Effect of Cambridge Analytica's Facebook ads on the 2016 US Presidential Election* [online]. 2019. [cit. 2021-03-15]. Dostupné z: <https://towardsdatascience.com/effect-of-cambridge-analyticas-facebook-ads-on-the-2016-us-presidential-election-dacb5462155d>.
- RESEARCH, P. *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information* [online]. 2019. [cit. 2021-03-15]. Dostupné z: <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>.
- SCOTT, T. *Why is Internet such a mess* [online]. 2020. [cit. 2021-03-15]. Dostupné z: <https://www.youtube.com/watch?v=0FRjZtYs3wY>.
- SECURITY, R. B. *New Research: No. of Records Exposed Increased 141% in 2020* [online]. 2021. [cit. 2021-04-02]. Dostupné z: <https://www.riskbasedsecurity.com/2021/01/21/new-research-no-of-records-exposed-increased-141-in-2020/#download>.
- SERVICES, T. N. *AP investigation: Across U.S., police officers abuse confidential databases* [online]. 2016. [cit. 2021-04-02]. Dostupné z: <https://www.chicagotribune.com/nation-world/ct-ap-police-database-abuse-20160928-story.html>.
- STATCOUNTER. *Search Engine Market Share Worldwide* [online]. 2021. [cit. 2021-03-15]. Dostupné z: <https://gs.statcounter.com/search-engine-market-share>.
- SÆTRA, H. S. Privacy as an aggregate public good. *Technology in Society*. 2020, 63, s. 101422. ISSN 0160-791X. doi: <https://doi.org/10.1016/j.techsoc.2020.101422>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0160791X20310381>.

- TANKOVSKA, H. *Number of monthly active Facebook users worldwide as of 4th quarter 2020* [online]. 2020. [cit. 2021-03-15]. Dostupné z: <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>.
- THOMPSON, S. A. – WARZEL, C. *Twelve Million Phones, One Dataset, Zero Privacy* [online]. 2019. [cit. 2021-04-05]. Dostupné z: <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>.
- VALÁŠEK, L. – HORÁK, J. *Policisté, politici i jejich děti. Čínský armádní dodavatel sbírá data stovek Čechů* [online]. 2020. [cit. 2021-03-15]. Dostupné z: <https://zpravy.aktualne.cz/domaci/policiste-politici-i-jejich-rodina-cinsky-armadni-dodavatel/r~e3e18dfcf6c211ea842f0cc47ab5f122/>.
- VOKUŠ, P. J. *Poskytovatelé internetového připojení* [online]. 2019. [cit. 2021-04-02]. Dostupné z: <https://www.policie.cz/clanek/poskytovatele-internetoveho-pripojeni.aspx>.
- YANG, Y. C. Web user behavioral profiling for user identification. *Decision Support Systems*. 2010, 49, 3, s. 261–271. ISSN 0167-9236. doi: <https://doi.org/10.1016/j.dss.2010.03.001>. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0167923610000497>.