ITAT 2013 Proceedings, CEUR Workshop Proceedings Vol. 1003, pp. 18–23
http://ceur-ws.org/Vol-1003, Series ISSN 1613-0073, © 2013 M. Pataky

Anonymita používateľa v internete

Mikuláš Pataky

Katedra aplikovanej informatiky, FMFI UK, Mlynská Dolina 842 48 Bratislava
pataky@fmph.uniba.sk,

Abstrakt: Sledovanie používateľa internetu a analyzovanie jeho správania sa je čoraz častejším javom. Jednou zo základných techník sledovania používateľa je stopovanie jeho webového prehliadača.

V tomto článku predstavíme teoretické východiská, rozoberieme niekoľko základných techník a predstavíme vlastný systém na detekciu odtlačkov prehliadačov a histórie nimi navštívených stránok s cieľom čo najpresnejšie identifikovať používateľa internetu.

Pri tvorení systému na deanonymizáciu sme kládli dôraz na jednoduchosť implementačného riešenia založeného na bežných webových technológiách ako PHP, CSS, JavaScript a Flash. Cieľom predstavovaného výskumu je ako využiť jeho výsledky v oblasti bezpečnosti (ochrana súkromia či ochrana systémov pred kyber útokmi) , tak aj využiť nazbierané dáta pre dôkladnejšiu analýzu návštevnosti web stránok.

Na overenie navrhnutého systému sme tento nasadili na stránky Univerzity Komenského v Bratislave a jej fakultách, čo nám umožnilo nazbierať dostatočné množstvo dát a v konečnom dôsledku zároveň pomohlo získať dôležité informácie o návštevnosti a správaní sa používateľov na jednotlivých stránkach univerzity.

Ako ukazujú výsledky nášho výskumu, identifikovanie používateľa cez jeho prehliadač, či zistenie jeho histórie, môže uskutočniť ľubovoľná web stránka. Za sledované jednomesačné obdobie sme boli schopní jednoznačne identifikovať 75.74% z 225 154 prehliadačov a 10.01% prehliadačov umožnilo detekciu histórie.

1 Úvod

Strata anonymity pre niektorých ľudí znamená vážny problém. Väčšina z nich si neuvedomuje, že k nej môže dôjsť rôznymi spôsobmi. Oblasti záujmu a informácie o jednotlivých používateľoch sa dajú dobre zistiť aj cez známe internetové vyhľadávače ako Google, Bing atď. Internetové vyhľadávače vedia kategorizovať, prípadne dostatočného množstva dopytov aj identifikovať, používateľov podľa sémantiky vyhľadávacích dopytov a zobrazovať výsledky, ktoré by mali byť pre nich zaujímavejšie.

Pomocou dobre mierených dopytov sa dá dostať i k citlivým informáciám ako sú emailové adresy, zle nakonfigurovaným alebo k zraniteľným serverom. Pre odhaľovanie takýchto dopytov bol vyvinutý framework *SearchAudit*(viď [1]). Identifikuje škodlivé dotazy z logov rozšírených vyhľadávacích strojov s úmyslom odhaliť ich spojitosť s potencionálnym útokom. Špecializované dotazy môžu útočníkom odhaliť veľmi konkrétne informácie z vyhľadávačov, ktoré by vlastnými nástrojmi len veľmi zložito získavali, navyše im poskytuje určitú mieru krytia.

Ďalším zo spôsobov využitia deanonymizačnej techniky je deSEO (viď [2]). Autori tohto článku vytvorili systém *deSEO*, ktorý odhaľuje SEO útokom. *deSEO* je systém na automatické detekovanie útokom nakazených výsledkov vyhľadávania bez skúmania obsahov webových stránok. Za dobu trvania experimentu a preskúmania stoviek miliárd URL adries z vyhľadávačov Bing a Google dosiahli zaujímavé výsledky:

1. deSEO identifikoval viacero skupín škodlivých URL. Každá z týchto škodlivých skupín korešpondovala s kampaňou ovplyvňujúcou tisícky URL.
2. deSEO je schopné identifikovať SEO kampane, ktoré používajú zložité techniky ako je cloaking a majú meniacu sa štruktúru liniek.
3. Odvodenie signatúr regulárnych výrazov na detekciu škodlivých URL skupín umožnilo zistiť, že až 36% všetkých výsledkov hľadania Googlu a Bigu obsahujúci vo svojich top výsledkoch aspoň jeden škodlivý link.

Systém *HostTracker* (viď [3]) využíva ID odvodené z logov aplikačnej vrstvy, s cieľom vytvoriť jedinečné identifikátory host-a a sledovať väzby host-ov s IP adresami. HostTracker bol nasadený na jednom z najväčších poskytovateľov emailovej komunikácie, kde dokázal 76% udalostí v logoch systému priradiť používateľovi a 92% z nich dokázal vysledovať. Počas jednomesačného nasadenia odhalil 12,6 milióna účtov botov s chybou false positive 0,4%. Za toto obdobie pomohol zablokovať 20,8 milióna škodlivých účtov.

Projekt Panopticlick (viď [4]) podáva zaujímavé výsledky o tom, že aj po 2 428 097 identifikovaných prehliadačoch, stále dokážu unikátne identifikovať nové. Toto je priamy dôkaz toho, koľko informácií sa dá z nich dozvedieť. Autor článku uvádza, že ich identifikačný systém dokáže úspešne identifikovať až 94,2% prehliadačov. Problém updatu prehliadača jednotlivých používateľov dokázali vyriešiť jednoduchou heuristikou, kde využívajú 85% zhodu starého a nového odtlačku. Táto heuristika im dokázala správne zaradiť až 99,1% odtlačkov. Najväčšiu entropiu v tomto projekte dosahovali doplnky a fonty.

Ďalší podobný výskum, ale vo väčšej miere, prebehol na pôde Microsoftu (viď [5]), kde analyzovali a porovnávali IP adresy, informácie prehliadača, cookies a používateľské prihlasovacie ID nazbierané počas jedného mesiaca

z Hotmailu a Bingu. Podľa zistení 60% – 70% návštevníkov sa dalo identifikovať s použitím iba *user agent* reťazca. S pridaním IP adresy sa úspešnosť zvýšila na 80%. User agent reťazec spolu s IP adresou mali v tomto výskume väčšiu entropiu, 20,29 bitov, ako kombinácia doplnkov prehliadača, rozlíšenia, časovej zóny a systémových fontov.

V súčasnej dobe sa stále zvyšuje záujem používateľov internetu o sociálne siete, s ktorými tiež súvisí deanonymizácia. Spojenie skupín v sociálnych sieťach s detekciou histórie v prehliadači viedlo k jednoznačnej identifikácii 42% používateľov (viď [6]). O prepojení informácií z viacerých sociálnych sietí píšu autori v (viď [7]), kde dokázali prepojiť tisícky účtov rôznych sociálnych sieťach a tým zhromaždiť viac súkromných informácií o používateľoch. Zaujímavému a čoraz aktuálnejšiemu prepojeniu sociálnych sietí a mobilných zariadení sa venujú výskumi (viď [8] a [9]).

1.1 *k*-anonymita

S deanonymizačnými technikami nevyhnutne súvisia aj ich inverzné techniky – anonymizačné. Tieto sa snažia upraviť citlivé dáta tak, aby ich časť mohla byť bezpečne zverejnená. S týmto problémom sa často v praxi stretávame pri lekárskych záznamoch. Kompletné lekárske záznamy sú vysoko súkromné údaje, ku ktorým by mali mať prístup iba kompetentné osoby. Na druhej strane je pre všeobecný prospech vedecká analýza týchto dát, ktorá môže pomôcť odhaliť negatívne trendy. Jednoduché odstránenie identifikátorov (napr. rodné čísla, číslo sociálneho poistenia) a mien z takýchto záznamov nezaručuje, aby sa znovu nedali identifikovať konkrétny ľudia.

Výskum ohľadne anonymizácií citlivých dát viedol k formalizovaniu prístupu nazvaného ako *k-anonymita* (viď. [10] a [11]). Tento prístup sa zakladá na výbere podmnožiny atribútov, **kvázi-identifikátorov** QI, danej tabuľky v databáze, ktoré spĺňajú podmienku k-anonymity.

Podmienka *k*-anonymity Všetky zverejnené dáta musia byť také, že každá kombinácia hodnôt z kvázi-identifikátorov môže byť nepriamo zhodná aspoň s k záznamami.

Samotná k-anonymita sa definuje ako:

***k*-anonymita** Nech $T(A_1, A_2, ..., A_m)$ je tabuľka a nech QI je kvázi-identifikátor zviazaný s T. Hovoríme, že T spĺňa k-anonymitu s ohľadom na QI vtedy a práve vtedy, keď každá sekvencia hodnôt $T[QI]$ sa vyskytuje v $T[QI]$ aspoň k-krát.

Ďalšou možnosťou anonymizácie je generalizovanie údajov daného atribútu. Napríklad vynechanie posledných dvoch čísel v PSČ. Týmito spôsoby sa môže časť citlivých údajov zverejniť ďalším výskumom bez toho, aby bola porušená anonymita pôvodných dát.

Avšak, ani vyššie uvedená k-anonymita nemusí vo všeobecnosti zabezpečiť úplnú anonymitu údajov, a teda vylúčiť budúcu deanonymizáciu. Útoky na k-anonymitu sú založené na ďalších informáciach, ku ktorým sa útočníci môžu dostať. Napríklad použitím anonymizovaných lekárskych a volebných údajov obyvateľov sa dajú získať o niektorých pacientoch úplné informácie (viď. [12]). Ochranu proti takémuto útoku zabezpečuje silnejšia podmienka na anonymitu ℓ-*Diversity*.

Pre náš výskum bolo zaujímavé prepojenie nášho detekčného systému s teóriou k-anonymity na zistenie stupňa anonymity daného prehliadača. Tento prístup rozpracovávame v časti výsledky.

1.2 Zhrnutie analýzy

Z uvedených príkladov je zrejmé, že deanonymizačné techniky majú veľký význam v počítačovej bezpečnosti, a to hneď z dvoch dôvodov. Jednak ako narušenie súkromia a odhaľovanie identít používateľov útočníkmi a jednak na odhaľovanie útočníkov, ktorí sa snažia narušiť bezpečnosť systému. Rôznych spôsobov deanonymizácie je niekoľko. V našom projekte sme sa rozhodli vytvoriť systém na deanonymizáciu používateľov internetu za pomoci zachytávania odtlačkov a histórií prehliadačov. Pri analýze nazbieraných dát zase využijeme k-anonymitu na určenie anonymít prehliadačov.

Pre stále rastúci počet užívateľov, ako aj webových stránok, či aplikácií, sme si dali za cieľ zistiť, ako dobre sa dá bežný užívateľ internetu identifikovať. Pri každom dopyte na webovú stránku, webový prehliadač za sebou zanecháva stopy – odtlačky, podľa ktorých ho možno ďalej sledovať. V niektorých kombináciách typu prehliadača a operačného systému je možné z prehliadača zistiť aj časť histórie navštívených stránok.

1.3 Organizácia článku

V prvej časti sme opísali viacero prístupov a využití deanonymizačných techník. V druhej časti popisujeme architektúru a činnosť vytvoreného detekčného systému. Ďalšia, tretia časť sa venuje analýze nazbieraných dát. V nej prehľadne prezentujeme najzaujímavejšie dosiahnuté výsledky. V predposlednej, štvrtej časti, vyvodzujeme závery a objasňujeme niektoré dosiahnuté výsledky, v ktorej predkladáme aj niekoľko usmernení pre aplikačnú prax, t.j. rád pre používateľov ako ostať na internete čo najviac anonymný. V poslednej časti predstavujeme plány do budúcna a ďalšie možnosti využitia implementovaného systému.

2 Detekčný systém

Prezentovaný výskum o anonymite sme zavŕšili vytvorením systému na detekciu návštevníka webovej stránky, teda na získavanie otlačkov jeho prehliadača, ktorého principiálnym znakom je jednoduchosť riešenia. Funkčné riešenie, ktoré sa nám podarilo implementovať do podoby

jedného <iframe>, je založené na bežných webových technológiách ako PHP, JavaScript a Flash.

Z vyššie uvedeného vyplýva, že podobný detekčný systém si môže v dnešnej dobe naprogramovať každý tvorca webových stránok. Tento fakt je pre nás výskum dôležitý, lebo dokazuje schopnosť ľubovolnej internetovej stránky identifikovať užívateľov prehliadač. Tieto dáta sa dajú následne z viacerých stránok spájať a tak budovať pomerne presný pohyb užívateľa.

Jednoduchosť a štandardnosť riešenia nám zároveň odstraňuje problém s označením nášho systému za škodlivý. Pri použití ďalších technológií ako napríklad Java Appletov by sa dalo pristúpiť k ďalším zaujímavým informáciám, akými sú odhadnutie výpočtového či zobrazovacieho výkonu daného zariadenia. Tieto vlastnosti sú aj pri rovnakých zariadeniach (napríklad mobilných zariadeniach jedného typu) čiastočne rozdielne. Ich výkon totiž závisí aj od rozdielnej verzie systému a najmä od aplikácií bežiacich na pozadí. Nevýhodou tohto riešenia je, že niektoré prehliadače (IE 10+ a Chrome) sú prednastavené, aby podobný obsah blokovali a oznamovali to používateľovi. Ten následne môže buď daný prvok zakázať, alebo povoliť. Takéto správanie systému, na často navštevovanej stránke ako je univerzitná stránka, nebolo žiaduce, na čo sme pri vývoji museli brať dôraz.

Ďalším prvkom, ktorý by mohol byť užitočný pri identifikácii používateľa, je poloha jeho zariadenia. Moderné mobilné zariadenia môžu svoju polohu poskytnúť navštívenej stránke pre zobrazenie relevantnejšieho obsahu. Aj pri tejto technológii je však obdobný problém ako s Appletmi, nakoľko používateľ musí toto poskytnutie informácie o polohe potvrdiť súhlasom.

2.1 Otlačky

Odtlačky sú informácie, ktoré za sebou používateľ, resp. jeho webový prehliadač, zanecháva. Podľa rôznych kombinácií zozbieraných informácií z odtlačkov vieme následne identifikovať daný prehliadač. Za predpokladu, že bežný užívateľ používa iba jeden či dva obľúbené prehliadače [1], je pomerne dobre vystopovateľný.

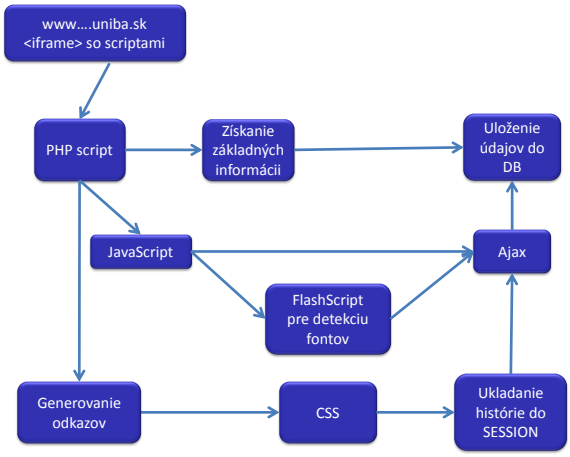
Samozrejme, neexistuje zaručený spôsob ako prepojiť viaceré prehliadače daného používateľa bez prihlásenia sa do systému, ktorý by jednoznačne identifikoval užívateľa. Sú prípady, kedy by odhalenie danej súvislosti bolo možné. Napríklad, ak existujú stránky, ktoré sú navštevované niektorými prehliadačmi častejšie v istých častiach dňa [2] alebo ojedinelá kombinácia navštívených stránok.

Ďalšia možnosť prepojenia viacerých prehliadačov k jednému užívateľovi, môže viesť cez detekciu histórie. Avšak, táto možnosť je omedzená iba na niektoré prehliadače.

V našom systéme zbierame 38 rôznych odtlačkov. Tie najvýznamnejšie uvádzame tu:

[1] jeden vo svojom notebooku a druhý v mobile

[2] cez pracovnú dobu z počítača v práci, po nej z mobilu, alebo z počítača doma



Obr. 1: Architektúra detekčného systému

User Agent obsahuje mikroverziu prehliadača, verziu operačného systému, jazyk, toolbary a občas aj ďalšie informácie, ktoré sa líšia s použitým prehliadačom.

Hlavičky HTTP ACCEPT

Povolenie cookies

Rozlíšenie obrazovky sa dá zisťovať v dvoch smeroch a to: rozlíšenie celej obrazovky a rozlíšenie prehliadača.

Farebná hĺbka obrazovky

Časová zóna

Doplnky prehliadačov

Systémové fonty sa dajú zisťovať za pomoci flashu. Ten má prístup k ich zoznamu. Okrem zistenia jednotlivých fontov je zaujímavý už len ich počet, ktorý sa od používateľa k používateľovi mení. Identifikácia fontov v podobe, ako ju používame, nefunguje na mobilných zariadeniach.

Test na partial supercookie či sa dajú vytvoriť cookie domény 1. úrovne.

Trieda CPU dá sa zistiť iba v niektorých prehliadačoch.

História dá sa sledovať najmä prehliadačoch Opera a v Interne Exploreri vo Windows XP.

2.2 Architektúra

Vyvinutý detekčný systém využíva niekoľko webových technológií. Prepojenie jednotlivých častí je znázornené na obrázku 1.

Systém je nasadený na jednotlivých stránkach pomocou jedného <iframe>. V ňom sa sa spúšťajú všetky skripty, ktoré odchytávajú jednotlivé odtlačky. Dôležitou

podmienkou je, aby veľkosť <iframe> bola aspoň jeden pixel. V prípade, ak táto podmienka nebude splnená, v <iframe> sa nič nevyrenderuje a skripty sa nezavolajú. Tým, že sme umožnili, aby <iframe> mal minimálnu veľkosť iba jeden pixel, sme sa nestretli s tým, aby narušil dizajn stránky, a teda môže byť široko používaný.

Základné otlačky ako meno, verzia prehliadača, IP adresa, URL navštívenej stránky, sa zisťujú pomocou **PHP**. Snažili sme sa, aby sme pomocou PHP odchytávali čo najviac odtlačkov pre prípad vypnutého JavaSriptu. **JavaSript** odchytáva väčšinu ďalších otlačkov, z ktorých najvýznamnejší je zoznam rozšírení (pluginov) prehliadača. **Flash** sa využíva na detekciu nainštalovaných fontov v OS.

2.3 História

Na zisťovanie *histórie* sa požíva špeciálne **CSS**, ktoré formátuje odkazy na vybrané stránky. Ak je stránka odkazu navštívená, prehliadač ju naformátuje pomocou pseudotriedy `visited`, ktorá sa odvoláva na náš detekčný systém.

Zisťovanie histórie navštívených webových stránok má viacero zaujímavých využití. Okrem lepšej identifikácie užívateľa používajúceho viaceré prehliadače, prípadne aj operačné systémy, sa dajú celkom dobre odhadnúť aj jeho záujmy. Takáto informácia by bola veľmi žiadaná napríklad v reklamnom priemysle, kde by užívatelia dostali iba reklamy, ktoré by ich mali zaujímať. Na druhú stranu história sa dá považovať za súkromnú informáciu a asi väčšina ľudí by ju nechcela zverejniť. Aj preto sa snažia prehliadače históriu dobre chrániť.

V roku 2010 bol publikovaný článok Feasibility and real-world implications of web browser history detection (viď [13]), ktorý popisuje zisťovanie histórie navštívených stránok. Uvádza dve odlišné metódy, ktoré sa zakladajú na CSS štýlovom príznaku `visible`.

Príznak `visible` umožňuje zvoliť rozličný štýl pre webové odkazy, ktorých stránky už používateľ navštívil a ktoré nie. Táto vlastnosť je pri tvorbe webových stránok pomerne obľúbená u používateľov, ktorí takto vedia zistiť, ktoré stránky z daného menu už navštívili.

Využiť na detekciu histórie sa dá príznak `visible` dvomi spôsobmi. Prvý je použitím JavaSrciptu. Konkrétne objektu `style` a jeho funkcie `getPropertyValue`. Táto funkcia vracia hodnotu štýlu, ktorý je použitý na nejaký HTML element. Za predpokladu vytvorenia dostatočne veľkej databázy URL najrôznejších stránok, vieme zistiť na ktorých stránkach z danej databázy užívateľ bol a urobiť si tak predstavu o jeho záujmoch. Tento spôsob však bol odhalený a všetky bežne používané prehliadače nedovoľujú dotazovanie na štýl odkazu, ktorý má nastavený príznak `visible`.

Druhý spôsob je využitie CSS vlastnosti `background`, ktorá nastavuje pozadie elementu. Do tejto vlastnosti sa dá vložiť URL adresa, ktorá môže odkazovať na ľubovoľné umiestnenie. Tým môže byť napríklad php skript, ktorý ako výsledok vráti obrázok do pozadia, a zároveň si vie zistiť pomocou get parametra odkaz, z akej stránky ho zavolal. Aj tento spôsob sa podarilo prehliadačom MSIE, Firefox, Chrome a Safari úspešne zablokovať tým, že nepodporujú URL v nastaveniach pozadia pre príznak `visible`. V prehliadačoch Opera je táto možnosť stále otvorená. Jej funkčnosť sme overili vlastným detekčným systémom odtlačkov.

Príklad štýlu, ktorý v našom systéme používame na zisťovanie histórie:

```
.n1:visited {
  color: green;
  background-image:
    url(http://...sk/visited?web=1&user=46);
}
```

Príklad odkazu, ktorý používame na zisťovanie histórie:

```
<a href="http://google.com"class="n1»</a>
```

3 Výsledky

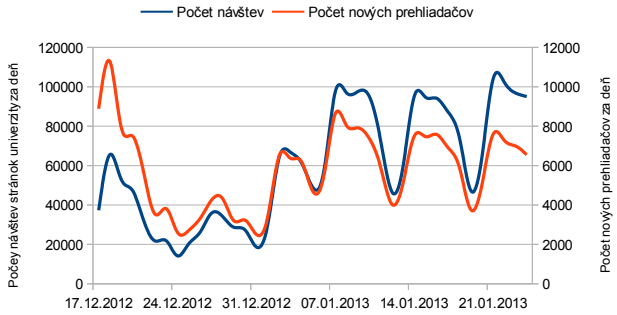
Presnosť nášho systému sme overovali za pomoci cookies s dlhou exspiračnou dobou. Vo výslednej analýze nazbieraných dát sme zreťazili všetky odtlačky prehliadača do jedného reťazca. Vzniknutý reťazec tvoril pre daný prehliadač identifikátor, pomocou ktorého sme skúmali anonymitu daného prehliadača. Využili sme teóriu *k*-anonymity, kde sme naše odtlačky prehliadača dosadili za kvázi-identifikátor. Vo všeobecnosti sme zistili, že anonymitu prehliadača najviac ovplyvňujú odtlačky ako rozšírenia prehliadača a systémové fonty. Počty prehliadačov spĺňajúcich daný stupeň *k*-anonymity je uvedený v tabuľke 1.

Systém bol nasadený na všetky stránky fakúlt Univerzity Komenského v Bratislave počas dvoch období. Testovacia 8 dňová prevádzka zaznamenala 263 070 zobrazení fakultných stránok s 41 007 prehliadačmi. Jednoznačne sa nám podarilo pomocou zozbieraných otlačkov *identifikovať* **84,93%** prehliadačov. *Históriu* sa nám podarilo zistiť u **13,21%**.

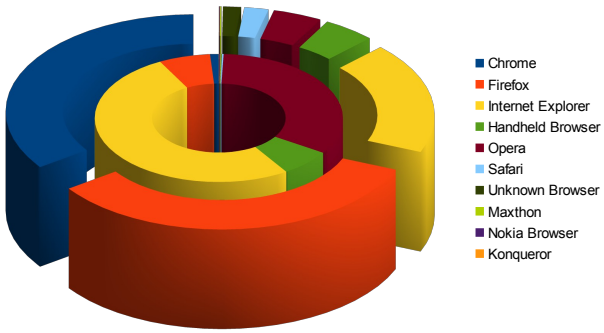
Počas ostrého vyše jednomesačného nasadenia od **17.12.2012** do **24.1.2013** sme zaznamenali úspešnosť identifikácie **75.74%** z 225 154 prehliadačov. Všetky údaje v tomto článku sú za toto obdobie.

V tabuľke 1 prehľadne uvádzame, koľko nami zachytených prehliadačov spĺňa daný stupeň *k*-anonymity. V prvom stĺpci uvádzame stupeň *k*-anonymity, v druhom počet unikátnych identifikačných reťazcov prehliadačov a v treťom počet prehliadačov spadajúcich do daného stupňa *k*-anonymity. Prehliadače patriace do prvého riadku tabuľky vieme jednoznačne identifikovať. Prehliadače v druhom riadku už len s presnosťou 50% atď.

10.01% prehliadačov sa za skúmané obdobie aspoň raz prihlásilo z IP adresy univerzity. Počet zobrazených stránok bol 2 334 351 z čoho 19.22% tvorili návštevy z IP ad-



Obr. 2: Graf zobrazuje vývin zaznamenávania dát od 17.12.12 do 24.1.13



Obr. 4: Zastúpenie prehliadačov pristupujúcich na stránky univerzity (vonkajší kruh) a zastúpenie prehliadačov pri úspešnej detekcii histórie (vnútorný kruh)

Tabuľka 1: k-anonymita prehliadačov

k-anonymita	# id. reťazcov	# prehliadačov
1	170533	170533
2	14272	28544
3	1275	3825
4	693	2772
5	309	1545
6	250	1500
7	150	1050
8	103	824
9	78	702
10	66	660

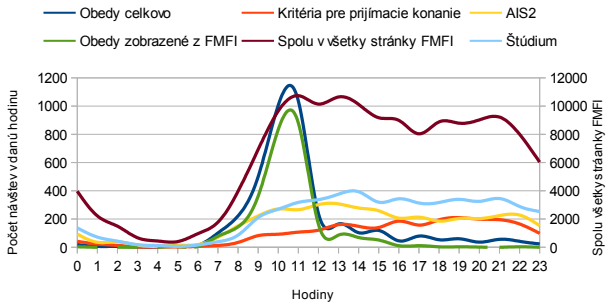
Tabuľka 2: Najnavštevovanejšie stránky FMFI

Počet návštev	Názov stránky
3928	AIS2
3020	Štúdium
2539	Bakalárske štúdium
2334	Študijné plány
2269	Bakalárske štúdium
2117	Kritéria pre prijímacie konanie
1895	Štúdium
1720	Aktuálne jedálne lístky
1637	Prvý stupeň
1615	Jedáleň FMFI

ries patriacim univerzite. Počet „migrujúcich“ prehliadačov sme zaznamenali **6 495** čo tvori **26.30%** zo všetkých prehliadačov prihlásených z univerzity (24 693).

U **23 503** (**10.39%**) prehliadačov bola zistená *história* jedného z 18tich odkazov. Ak by sme rozšírili databázu odkazov, vieme pomerne dobre odhadnúť, aké stránky používatelia daných prehliadačov navštevujú. Detekciu histórie najviac umožňoval Internet Explorer, Opera a Handheld Browser (viď obr. 4).

Pomocou nášho systému sa dajú zistiť aj iné zaujímavé štatistické zistenia ako najpoužívanejšie prehliadače (Chrome 34.79%, Firefox 33.99%, IE 18.62%), operačné systémy (Windows 7 49.69%, Windows XP 28.74%, Windows Vista 7.29%, Android 2.96%, Mac OS X 2.35%), najmenej vyťažené dni (24.-25.12.), najaktívnejší prehliadač/používateľ (11 899, 7 768, 6 049 zobrazení), najnavštevovanejšia stránka univerzity http://www.fphil.uniba.sk/index.php?id=5885 *Prijímacie konanie 2013* 13072 zobrazení (0.56%), používatelia si zobrazili 2 743 stránok FMFI spolu 150 250 krát, atď.



Obr. 3: Počet zobrazení vybraných stránok FMFI v jednotlivých hodinách

4 Záver

Výsledky nášho výskumu poukazujú na to, že s pomerne dobrou presnosťou môže identifikovanie používateľa, či v niektorých prípadoch zistenie jeho histórie, uskutočniť ľubovoľná web stránka. S porovnaním iných podobných projektov, ktoré sme uviedli v úvode (viď [4] a [5]), sme dosiahli menšiu úspešnosť. Tento fakt môže byť spôsobený nasadením systému v akademickej sfére, kde sa nachádza pomerne veľa zhodných počítačov. Ďalším dôvodom môže byť, že uvedené projekty zobrazujú výsledky spred niekoľkých rokov a dnes si vývojári prehliadačov dávajú ešte viac záležať na rýchlej a častej aktualizácii, ktorá môže ovplyvňovať naše výsledky.

Účinnosť takéhoto systému, môže závisieť aj od konkrétnej web stránky, pre akých používateľov s akými pre-

hliadačmi je zaujímavá. Ak stránka zaujme používateľov s ojedinelými prehliadačmi, či so špeciálnymi rozšíreniami, je vysoká pravdepodobnosť veľmi dobrej účinnosti systému.

Takéto identifikovanie prehliadača sa dá následne jednoducho prepojiť aj s ďalšími informáciami, ktoré na internete poskytujeme. Napríklad prepojenie prehliadača a sociálnej siete či e-mailového účtu by umožňovalo zistenie komplexných informácií o danom používateľovi internetu, od jeho osobných údajov až po jeho aktivity v internete.

4.1 Ako o sebe prezradiť menej

Ideálne pre zabezpečenie čo najväčšej miery anonymity pri prezeraní webového obsahu by bolo, používať najpoužívanejší prehliadač webových stránok, s najrozšírenejším nastavením a množinou doplnkov, ktoré má najviac jeho používateľov. Ďalší dobrý spôsob je následne použiť jeden z anonymizérov, ktorý všetku komunikáciu presmeruje cez sériu proxy serverov. Otázkou samozrejme zostáva, kto je v pozadí takéhoto programu. Na internete sa ich dá nájsť pomerne dosť aj v podobe open source. Šesť takýchto aplikácií je recenzovaných v článku [14].

Ďalšou možnosťou je používať prehliadač s najprísnejším bezpečnostným nastavením, bez povolenia cookies, ukladania histórie, zakázaného JavaSriptu, nenainštalovaných žiadnych rozšírení, dokonca ani Flash, ci Java. Niektoré z týchto vlastností dnes všetky bežne používané prehliadače podporujú pod súkromným prezeraním. Samozrejmosťou je spúšťanie všetkej komunikácie, cez niekoľko proxy serverov. Týmto spôsobom sa síce sťaží stopovanie daného prehliadača, ale veľa webových aplikácií nebude vôbec funkčných, a aj obsah stránok môže byť veľmi obmedzený.

Aj napriek tomuto sa dá užívateľ vystopovať a identifikovať pomocou jeho dotazov a prezeraných stránok, napríklad v rozšírenom vyhľadávači.

5 Plány do budúcna

V našom výskumnom projekte sa aj naďalej zaoberáme novými deanonymizačnými technikami, ktoré postupne zapracovávame do vytvoreného a implementovaného systému s cieľom zvýšiť jeho presnosť. Skúmame ďalšie možnosti identifikácie a dlhodobo by sme chceli zistiť, napríklad koľko terajších uchádzačov o štúdium následne v septembri navštívi univerzitu, so svojimi prehliadačmi.

Náš systém je zároveň aj ideálnou možnosťou pre analyzovanie efektívnosti propagačných akcií na univerzite, či na jednotlivých fakultách. V tejto oblasti rozvíjame spoluprácu s Centrom Informačných Technológii (CIT) Univerzity Komenského v Bratislave.

Vyvinutý systém by sme chceli použiť aj na ďalšie praktické aplikácie, napríklad ako sledovanie správania sa používateľa na stránkach a odhadovanie jeho záujmov, vyhodnocovanie návštevnosti jednotlivých noviniek na stránkach, dynamické tvorenie webového menu atď.

Literatúra

[1] J. P. John, F. Yu, Y. Xie, M. Abadi, and A. Krishnamurthy, "Searching the searchers with searchaudit," in *Proceedings of the 19th USENIX conference on Security*, USENIX Security'10, (Berkeley, CA, USA), pp. 9–9, USENIX Association, 2010.

[2] J. P. John, F. Yu, Y. Xie, A. Krishnamurthy, and M. Abadi, "deseo: combating search-result poisoning," in *Proceedings of the 20th USENIX conference on Security*, SEC'11, (Berkeley, CA, USA), pp. 20–20, USENIX Association, 2011.

[3] Y. Xie, F. Yu, and M. Abadi, "De-anonymizing the internet using unreliable ids," *SIGCOMM Comput. Commun. Rev.*, vol. 39, pp. 75–86, Aug. 2009.

[4] P. Eckersley, "How unique is your web browser?," in *Proceedings of the 10th international conference on Privacy enhancing technologies*, PETS'10, (Berlin, Heidelberg), pp. 1–18, Springer-Verlag, 2010.

[5] T.-F. Yen, Y. Xie, F. Yu, R. P. Yu, and M. Abadi, "Host Fingerprinting and Tracking on the Web: Privacy and Security Implications," in *Proceedings of the 19th Annual Network & Distributed System Security Symposium*, Feb. 2012.

[6] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A practical attack to de-anonymize social network users," in *Proceedings of the 2010 IEEE Symposium on Security and Privacy*, SP '10, (Washington, DC, USA), pp. 223–238, IEEE Computer Society, 2010.

[7] M. Korayem and D. J. Crandall, "De-anonymizing users across heterogeneous social computing platforms," *The 7th international aaai conference on weblogs and social media (ICWSM 2013)*, 2013.

[8] N. D. Lane, J. Xie, T. Moscibroda, and F. Zhao, "On the feasibility of user de-anonymization from shared mobile sensor data," in *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*, PhoneSense '12, (New York, NY, USA), pp. 3:1–3:5, ACM, 2012.

[9] M. Srivatsa and M. Hicks, "Deanonymizing mobility traces: using social network as a side-channel," in *Proceedings of the 2012 ACM conference on Computer and communications security*, CCS '12, (New York, NY, USA), pp. 628–637, ACM, 2012.

[10] V. Ciriani, S. D. C. di Vimercati, S. Foresti, and P. Samarati, "*k*-anonymity," in *Secure Data Management in Decentralized Systems*, pp. 323–353, Springer US, 2007.

[11] L. Sweeney, "k-anonymity: a model for protecting privacy," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 10, pp. 557–570, Oct. 2002.

[12] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discov. Data*, vol. 1, Mar. 2007.

[13] A. Janc and L. Olejnik, "Feasibility and real-world implications of web browser history detection," 2010.

[14] J. Sedlák, "Hlavně nenápadne," *Computer*, vol. 18, pp. 32–34, Sept. 2011.