Testovanie bezpečnosti prihlasovacích údajov

Študent:	Miroslav Vojtek
Vedúci práce:	Mgr. Tomáš Bajtoš
Konzultant:	RNDr. Veronika Kopčová

Analysis of publicly leaked credentials and the long story of password (Re-) use

- David Jaeger
- Chris Pelchen
- Hendrik Graupner
- Feng Cheng
- Christoph Meinel

International Conference on Passwords 2016, Ruhr-University Bochum, Germany

Príbuzné práce – analýza regionálnych hesiel

- Málo preskúmaná oblasť
- Jaeger a kol. – čisto čínske úniky, rozdiely s medzinárodnými heslami
- Nie latinka, ale azbuka, či arabské znaky
- Dell Amico a kol. – prelamovaine hesiel s jazykovo špecifickými slovníkmi má pre heslo daného jazyka vyššiu úspešnosť

Príbuzné práce – opätovného používanie hesiel

- Zatiaľ robená len na malých vzorkách
- Prieskumy – ťažko získať pravé informácie, ľudia sa boja
- Hunt – 2 úniky, 88 používateľov, 33%
- Das a kolektív – 10 únikov, 6077 používateľov, 43%

Identity Leak Checker Service

- Webová služba - upozornenie na krádež identity, rady, ako správne tvoriť heslo
- Spustené v máji 2014
- 2.5 mil. ľudí si overilo, či ich údaje boli verejných únikoch
- 200 000 z nich bolo upozornených

Postup práce

- Vyčistenie datasetu, snaha previesť všetko do CSV formátu
- Interpretácia dát
- Vloženie do databázy
- Anonymizácia – hešovanie mailov

Analýza hesiel

- Analýza najväčšej databázy toho času
- 31 únikov,
- takmer 1 miliarda prihlasovacích záznamov
- 994 301 846 rôznych záznamov
- Záznam = emailová adresa a heslo alebo jeho heš
- Odstránenie nesprávnych mailov - 884 460 979 rôznych mailov

ID	Name	Passw. routine	Accounts with passw.	Leak date
1	000webhost.com	`$p`	15 035 687	≈ Mar. 2015
2	17.media	`md5($p)`	3 824 575	≈ Sep. 2015
3	51cto.com	`md5(md5($p).$s)`, `md5($p)`	3 923 449	≈ Dec. 2013
4	7k7k.com	`$p`	9 231 185	≈ Oct. 2011
5	aipai.com	`md5($p)`	4 529 928	≈ Apr. 2011
6	ashleymadison.com	`bcrypt($p)`	36 140 796	≈ July 2015
7	badoo.com	`md5($p)`	122 730 419	≈ June 2016
8	csdn.net	`$p`	6 425 905	≈ Oct. 2011
9	duduniu.cn	`$p`	14 192 866	≈ Aug. 2011
10	gawker.com	`des($p)`	487 292	≈ Dec. 2010
11	gmail.com	`$p`	4 925 994	≈ Sep. 2014
12	imesh.com	`md5(md5($p).$s)`	51 308 651	≈ Sep. 2013
13	ispeak.cn	`$p`	8 294 278	≈ Apr. 2011
14	linkedin.com	`sha1($p)`	112 275 414	≈ Feb. 2012
15	mail.ru	`$p`	5 269 103	≈ Sep. 2014
16	mate1.com	`$p`	27 402 581	≈ Feb. 2016

17	mpgh.net	`md5(md5($p).$s)`	3 119 180	≈ Oct. 2015
18	myspace.com	`sha1($p)`	358 986 419	≈ 2008
19	naughtyamerica.com	`md5($p)`	989 401	≈ Apr. 2016
20	nexusmods.com	`md5(md5($s).md5($p))`	5 918 540	≈ Dec. 2015
21	r2games.com	`md5(md5($p).$s), md5($p)`	11 758 232	≈ Oct. 2015
22	renren.com	$p	4 392 208	≈ Nov. 2011
23	sprashivai.ru	$p	3 472 645	≈ May 2015
24	taobao.com	$p	14 769 995	≈ Jul. 2015
25	tianya.cn	$p	29 642 564	≈ Nov. 2011
26	twitter.com	$p	26 121 984	≈ June 2016
27	vk.com	$p	92 144 526	≈ 2012
28	weibo.com	$p	4 529 994	≈ Dec. 2011
29	xiaomi.com	`md5(md5($p).$s)`	8 281 358	≈ May 2014
30	xsplit.com	`sha1($p)`	2 990 112	≈ Nov. 2013
31	yandex.ru	$p	1 186 565	≈ Sep. 2014
Total accounts with email addr.: 994 301 846 , **Total distinct email addr.:** 884 460 979				

Table 1: Analyzed identity leaks ($p - clear password, $s - salt)

Hash routine	Common name	# of leaks	# of dumps
`$p`	cleartext	16 ($\approx$ 51.6%)	6 ($\approx$ 28.5%)
`md5($p)`	MD5	4 (12.9%)	4 ($\approx$ 19.0%)
`sha1($p)`	SHA-1	3 (9.7%)	3 ($\approx$ 14.3%)
`des($p)`	descrypt	1 ($\approx$ 3.2%)	1 ($\approx$ 4.8%)
`md5(md5($p).$s)`	vBulletin-Hash	5 ($\approx$ 16.1%)	5 ($\approx$ 23.8%)
`md5(md5($s).md5($p))`	MyBB-Hash	1 ($\approx$3.2%)	1 ($\approx$ 4.8%)
`bcrypt($p)`	bcrypt	1 ($\approx$ 3.2%)	1 ($\approx$ 4.8%)

Table 2. Password routines of all identity leaks

Analýza hesiel

- Hľadanie hešu v dúhovej tabuľke
- Hľadanie hešu – gúglenie – niektoré už mohli byť prelomené
- Hľadanie hešu na niektorých heš-krekovacích stránkach
- Stiahnutie zoznamov už kreknutých hesiel z heš-krekovacích fór
- 848 miliónov cleartext záznamov, 320 miliónov rôznych hesiel

Table 3. Credentials with cleartext passwords and percentage of recovered encrypted password, - was used for cleartext only leaks

Name	Clear cred.	Rec.	Name	Clear cred.	Rec.
000webhost.com	15 035 687	-	mpgh.net	247 499	8%
17.media	2 709 893	71%	myspace.com	328 152 578	91%
51cto.com	2 228 479	67%	naughtyamerica.com	911 781	92%
7k7k.com	9 231 185	-	nexusmods.com	2 691 088	45%
aipai.com	2 221 875	49%	r2games.com	364 927	3%
ashleymadison.com	2 559 028	8%	renren.com	4 392 208	-
badoo.com	114 090 491	97%	sprashivai.ru	3 472 645	-
csdn.net	6 425 905	-	taobao.com	14 769 995	-
duduniu.cn	14 192 866	-	tianya.cn	29 642 564	-
gawker.com	439 449	90%	twitter.com	26 121 984	-
gmail.com	4 925 994	-	vk.com	92 144 526	-
imesh.com	15 908 834	32%	weibo.com	4 529 994	-
ispeak.cn	8 294 278	-	xiaomi.com	1 167 052	14%
linkedin.com	104 955 280	93%	xsplit.com	2 904 588	97%
mail.ru	5 269 103	-	yandex.ru	1 186 565	-
mate1.com	27 402 581	-			
Total cleartext cred.: 848 590 922 , **Cleartext passwords:** 320 201 615					

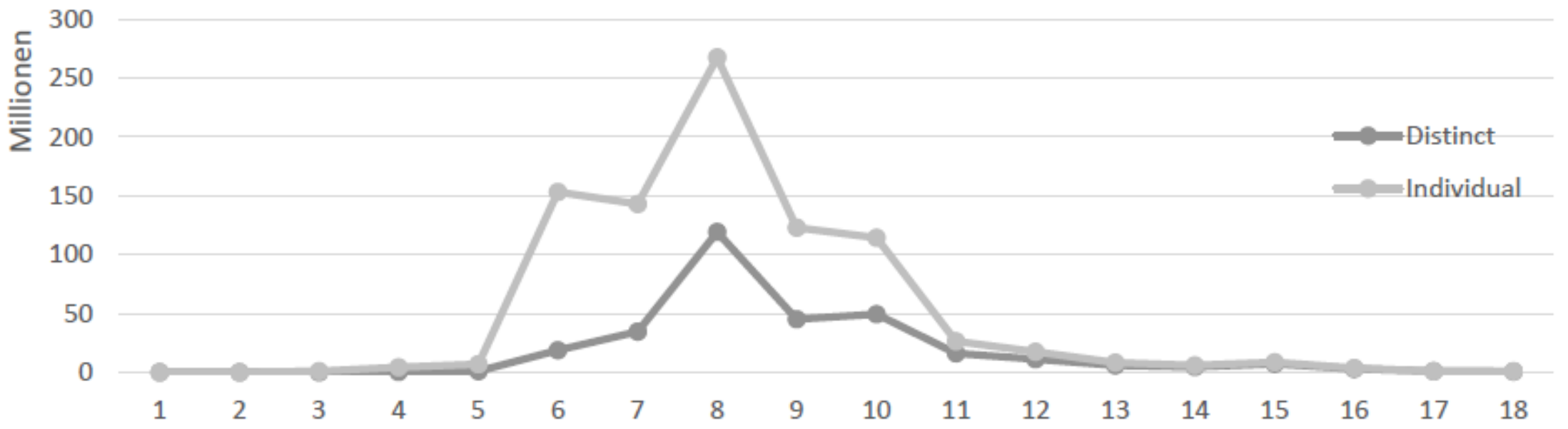

Fig. 2. Distribution of password lengths (distinct - each password only once, individual - password used by a user in a leaked source)

Analýza hesiel

- Najviac hesiel má 8 znakov, veľká časť 6-10
- 81% hesiel je tvorených len malými písmenami a číslicami
- 64% hesiel obsahuje postupnosť malých písmen, po ktorých nasledujú číslice ([a-z]*[0-9]*)
- 20 top hesiel – všetky môžeme považovať za slabé heslá

Table 4. Normalized top passwords

Top 1-5		Top 6-10		Top 11-15		Top 16-20	
1	123456	6	password	11	000000	16	abc123
2	111111	7	1q2w3e4r	12	1234567890	17	123qwe
3	12345678	8	1qaz2wsx	13	666666	18	654321
4	123456789	9	1234567	14	123321	19	112233
5	123123	10	iloveyou	15	qwerty	20	11111111

Opätovné použitie hesla

- Skúmané záznamy s rovnakým mailom
- Miera podobnosti – normalizovaná Levensteinova vzdialenosť

  1 - (vzdialenosť / dĺžka dlhšieho reťazca)
- Každý unikátny mail – ohodnotený graf, vrcholy – heslá k danému mailu, hrany medzi nimi – podobnosť hesiel
- Hľadanie kompletného podgrafu s mierou podobnosti viac ako 0.7
- Výpočet priemernej podobnosti v takom podgrafe

Opätovné použitie hesla

- 65 miliónov mailov vo viacerých únikoch
- Cca 19 miliónov, kde miera podobnosti bola viac ako 0.7
- 13.7 mil. použitie úplne rovnakého hesla
- Z nich:
  - 12.9 mil. rovnaké heslo pre 2 služby
  - 825 000 rovnaké heslo pre 3 služby
  - 60 000 rovnaké heslo pre 4 služby

Jazyková analýza

- Maily s doménou .com sa použiť nedajú (73%)
- Vybrané domény: .nl, .cn, .ru, .fr, .it, .uk, .de
- Vytvorený zoznam top 1000 hesiel z celého datasetu
- Vytvorené zoznamy top hesiel pre každú doménu(ak heslo bolo v top 1000, tak nie je špecifické pre daný jazyk)

Jazyková analýza

- .uk: futbalové kluby, mená členov kráľovskej rodiny, mestá
- .fr: „azerty“ – chod po klávesnici, mestá
- .de: typické mená

- Celkovo: mestá, typické mená

Table 5. Country-specific passwords

ID	Domain	Language	number of addresses	Top 5 passwords
1	.uk	British English	18 604 736	liverpool, arsenal, chelsea
2	.fr	French	32 207 859	azerty, marseille, doudou
3	.de	German	15 401 823	passwort, ficken, qwertz
4	.it	Italian	21 856 935	juventus, andrea, francesco
5	.nl	Dutch	3 513 385	welkom, welkom01, wachtwoord
6	.cn	Chinese	12 213 153	5201314, woaini, 1314520
7	.ru	Russian	119 002 753	qwertyuiop, UsdopaA, 1q2w3e4r5t

Zlepšenie efektívnosti prelamovanie hesiel

- Hľadanie rovnakého mailu v iných únikoch – jeho heslá
- Únik Ashley Madison – bcrypt, obnovených 2.9 mil. hesiel
- Ashley Madison a Mate1 – zoznamky, 1.94 mil. rovnakých mailov,

  clear text heslá z Mate 1 -> bcrypt, 5 h – cca 913 000 hesiel

  heslá v top 20, 5h – cca 1800 hesiel

Máte nejaké otázky?