Testovanie bezpečnosti prihlasovacích údajov

Miroslav Vojtek

3Ib, 2018 - 2019

Abstrakt. Najväčšou slabinou bezpečnosti je stále človek. V dnešnej dobe, keď si veľký objem súkromného obsahu chránime prihlasovacími údajmi, je potrebné vedieť si tvoriť bezpečné prihlasovacie údaje. Hlavným cieľom práce je skúmaním doteraz najväčšieho datasetu uniknutých prihlasovacích údajov zistiť, aké prihlasovacie údaje sú bezpečné a navrhnúť a implementovať systém, ktorý by testoval bezpečnosť prihlasovacích údajov.

Kľúčové slová: prihlasovacie údaje, heslo, bezpečnosť.

1 Úvod

V dnešnej dobe si mnoho ľudí chráni svoje účty na webových službách prihlasovacími údajmi. Často sú prihlasovacími údajmi meno a heslo. Pri menách sa upúšťa od vymyslenej prezývky a ako meno sa čoraz častejšie používa email, čo je určite aj ľahšie zapamätateľné. Je potrebné, aby heslo bolo čo najbezpečnejšie. Čo je však bezpečné heslo? Odpovedať na túto otázku nie je jednoduché. Chcieť poznať odpoveď na túto otázku bolo hlavnou motiváciou, prečo som si vybral ako tému bakalárskej práce tému „Testovanie bezpečnosti prihlasovacích údajov“.

Prvým z cieľov našej práce je porovnanie a spracovanie aktuálnych prístupov a nástrojov na analýzu bezpečnosti prihlasovacích údajov. Existuje veľa online nástrojov na zistenie sily hesla. Vyberieme z nich niekoľko, zistíme, ako určujú silu hesla a porovnáme ich.

Ďalším z cieľov je návrh metód na testovanie bezpečnosti prihlasovacích údajov. Tieto návrhy metód budú vyplývať zo zistení pri skúmaní datasetu, ktorý CSRIT-u UPJŠ poskytol na výskumné účely vládny CSIRT. Je to zbierka dát z menších datasetov reálnych prihlasovacích údajov ľudí z celého sveta uniknutých z rôznych služieb. Dataset je vo forme textových súborov, v ktorých sú umiestnené vo vhodnej adresárovej štruktúre kvôli lepšej prehľadnosti. V súboroch sa v každom riadku nachádza jedna dvojica (email a heslo) oddelená dvojbodkou. Budeme skúmať, aké majú ľudia zvyky pri vytváraní si prihlasovacích údajov.

Tretím cieľom našej práce je návrh a implementácia systému na testovanie bezpečnosti prihlasovacích údajov a jeho verifikácia. Tento systém bude vychádzať z aktuálnych nástrojov, no bude obohatený o pravidlá, ktoré sformulujeme po preskúmaní datasetu.

1.1 Prehľad podobných prác

Práca [1] sa zaoberá analýzou datasetu automatizovaných SSH bruteforce útokov. Tento tím zozbieral dataset, ktorý obsahuje používateľské mená, heslá, čas a zdroj útoku. Skúmali správanie útočníkov a na základe toho sformulovali viaceré odporúčania pre SSH používateľov a administrátorov.

Práca [2] sa zaoberá štatistickou a priestorovou analýzou prihlasovacích údajov(prihlasovacie meno heslo) zozbieraných pomocou honeypotov. Skúma vlastnosti prihlasovacích mien a hesiel, najpoužívanejšie prihlasovacie mená a heslá, mená a heslá podľa krajín, vzťah medzi prihlasovacím menom a heslom.

Práca [3] sa zaoberá analýzou verejne uniknutých prihlasovacích údajov z rôznych služieb, podobne ako naša práca na rozdiel od prác [1] a [2], ktoré skúmajú prihlasovacie údaje, ktoré skúšali útočníci. Pracuje s veľmi veľkým datasetom, ktorý obsahuje približne 1 miliardu dvojíc meno a heslo. Táto práca sa tiež zameriava na opätovné používanie a „recykláciu“ hesiel, tiež regionálne rozdiely pri tvorbe hesiel. Podobne ako sú v tejto práci porovnávané heslá jedného používateľa pomocou ich Levensteinovej vzdialenosti budeme my v našej práci porovnávať podobnosť prihlasovacieho mena a hesla.

1.2 Prehľad súčasných testerov sily hesla

V súčasnosti existuje už mnoho nástrojov (aj online) na testovanie sily prihlasovacích údajov, konkrétne nástroje na testovanie sily hesla. Silu hesla posudzujú podľa rôznych kritérií. Najčastejšie podľa dĺžky a rôznorodosti použitých znakov.

[1] pri posudzovaní sily hesla zohľadňuje jeho dĺžku a to, či sa v ňom nachádzajú opakujúce sa znaky. Tiež uvádza približný čas, ktorý je potrebný na prelomenie tohto hesla hrubou silou pomocou priemerného domáceho počítača.

[2] silu hesla posudzuje podľa viacerých kritérií. Zohľadňuje dĺžku hesla, rôznorodosť znakov hesla(či sú použité písmená, číslice alebo symboly) a tiež aj to, či testované heslo môže byť meno alebo slovo, ktoré sa nachádza v slovníku. Tento nástroj tiež uvádza odhadovaný čas, ktorý je potrebný na prelomenie hesla.

Minimálnymi požiadavkami na heslo nástroja [3] sú: dĺžka aspoň 8 znakov a výskyt znakov z aspoň 3 z týchto kategórií: malé písmeno, veľké písmeno, číslica, symbol. Silu hesla zvyšuje počet znakov hesla, počet malých, veľkých písmen, číslic, symbolov, tiež počet symbolov vo vnútri hesla a celkové splnenie minimálnych požiadaviek. Naopak, sila hesla je menšia v prípadoch, keď heslo obsahuje len písmená, len číslice, len malé písmená, len veľké písmená, opakujúce sa znaky, súvislú postupnosť malých písmen, veľkých písmen alebo číslic.

1.3 Prehľad súčasných minimálnych požiadaviek na prihlasovacie údaje

Keďže chceme hodnotiť bezpečnosť prihlasovacích údajov, tak je dobré vedieť, ako minimálne požiadavky na prihlasovacie údaje majú v dnešnej dobe rôzne populárne webové služby. Vybrali sme rôzne slovenské a celosvetové služby a zistili, aké majú stanovené minimálne požiadavky na heslo. Tento prehľad je robený ku dňu 1.3.2019.

Tabuľka 1. Minimálne požiadavky vybraných služieb na prihlasovacie údaje.

Služba	**Požiadavky na heslo**
Amazon.com Centrum.sk Instagram.com MySpace.com Profesia.sk Spotify.com Wikipedia.org	Dĺžka aspoň 6 znakov.
Ebay.com	Dĺžka aspoň 6 znakov. Musí obsahovať číslicu alebo symbol. Musí obsahovať aspoň 1 písmeno.
Aliexpress.com	Dĺžka aspoň 6 znakov. Musí obsahovať aspoň 2 z: malé písmeno, veľké písmeno, číslica, interpunkčné znamienko. Môže obsahovať len písmená, číslice, interpunkčné znamienka.
Facebook.com Twitter.com Pinterest.com	Dĺžka aspoň 6 znakov. Musí obsahovať aspoň 2 z týchto: malé písmeno, veľké písmeno, číslica.
Azet.sk	Dĺžka aspoň 7 znakov.
LinkedIn.com	Dĺžka aspoň 8 znakov. Musí obsahovať aspoň 1 číslicu alebo špeciálny znak.
Google.com	Dĺžka aspoň 8 znakov. Musí obsahovať aspoň 3 z týchto: malé písmeno, veľké písmeno, číslica, symbol.
GitHub.com	Dĺžka aspoň 8 znakov, ak obsahuje číslicu a malé písmeno alebo dĺžka aspoň 15 znakov bez ďalších obmedzení.
Sme.sk AIS2.sk	Dĺžka aspoň 8 znakov. Musí obsahovať aspoň 1 malé písmeno. Musí obsahovať aspoň 1 veľké písmeno. Musí obsahovať aspoň 1 číslicu.
Zoznam.sk	Dĺžka aspoň 8 a najviac 16 znakov. Musí obsahovať kombináciu aspoň 3 z: malé písmená, veľké písmená, číslice, špeciálne znaky(~!@#$%&*).
WordPress.org	Dĺžka aspoň 9 znakov. Musí obsahovať aspoň 3 z týchto: malé písmeno, veľké písmeno, číslica, symbol.

2 Práca s datasetom

1.2 Úprava datasetu

Keďže dataset, s ktorým pracujeme, vznikol spojením mnohých menších datasetov z únikov prihlasovacích údajov z rôznych služieb, bolo potrebné ho upraviť do jednotného tvaru, prípadne vymazať údaje, ktoré nie je možné previesť na požadovaný tvar. Každý riadok v súboroch datasetu predstavuje jeden záznam, najčastejšie je v tvare „email:heslo“. Niektoré záznamy nemali ako oddeľovač emailu a hesla použitú dvojbodku, ale bodkočiarku, takže tú sme nahradili dvojbodkou. Záznamy, ktoré boli v tvare „heslo:email“ alebo „heslo;email“ boli prevedené na jednotný tvar „email:heslo“. V záznamoch, ktoré mali v emaile dva zavináče za sebou, boli tieto dva zavináče nahradené jedným. Odstránené boli záznamy, ktorých heslo bolo dlhšie ako 50 znakov alebo email nebol v korektnom tvare.

2.2 Transportovanie datasetu do databázy

Aby bolo skúmanie datasetu jednoduchšie, rozhodli sme sa vložiť prihlasovacie údaje z datasetu do relačnej databázy. Ako databázu sme zvolili MySQL. Tabuľka s prihlasovacími údajmi okrem emailu a hesla obsahuje viaceré vlastnosti daného záznamu, ktoré boli dopočítané pred samotným vkladaním. Mali by uľahčiť skúmanie datasetu. Atribútmi tabuľky s prihlasovacími údajmi sú: id, meno, email, top-doména, sub-doména, heslo, dĺžka mena, dĺžka hesla, entropia hesla, dĺžka najdlhšej postupnosti číslic v hesle, počet malých písmen v hesle, počet veľkých písmen v hesle, počet číslic v hesle, dĺžka najdlhšieho spoločného podreťazca mena a hesla, normalizovaná Levensteinova vzdialenosť mena a hesla, Jaccardov index mena a hesla a priemerná vzdialenosť znakov hesla na vybranom modeli klávesnice.

Prihlasovacie meno, v našom prípade email, popisujú atribúty meno, email, top-doména, sub-doména, dĺžka mena. Meno je časť emailu od začiatku po zavináč. Email je celý email. Top-doména je časť emailu za poslednou bodkou.
Sub-doména je časť emailu od zavináča po poslednú bodku.

Heslo je popisované atribútmi heslo, dĺžka hesla, entropia hesla, dĺžka najdlhšej postupnosti číslic v hesle, počet malých písmen v hesle, počet veľkých písmen v hesle, počet číslic v hesle a priemerná vzdialenosť znakov hesla na vybranom modeli klávesnice. Atribút heslo obsahuje samotné heslo.

Normalizovaná Levensteinova vzdialenosť prihlasovacieho mena a hesla je počítaná ako rozdiel 1 a podielu Levensteinovej vzdialenosti a dĺžky dlhšieho reťazca.

$$1 - (\text{Levensteinova vzdialenosť} / \text{dĺžka dlhšieho reťazca})$$

Levensteinovu vzdialenosť predstavuje počet operácií, ktoré musíme vykonať na to, aby sme zmenili jeden reťazec na druhý. Rozlišujeme 3 operácie: vloženie znaku, nahradenie znaku a vymazanie znaku [3].

Jaccardov index prihlasovacieho mena a hesla je ďalšou mierou podobnosti dvoch reťazcov. Ak A je množinou znakov prihlasovacieho mena a B je množinou znakov

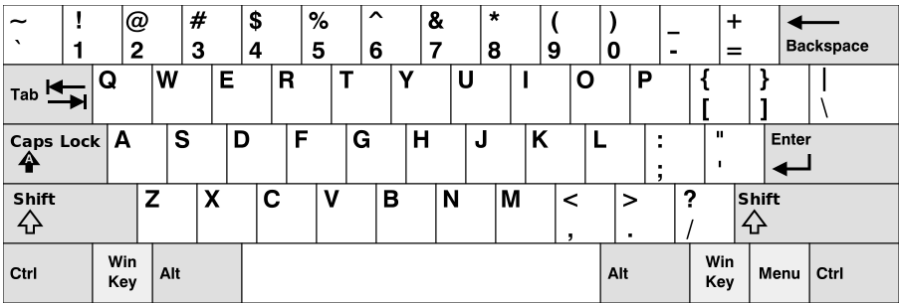
hesla, tak Jaccardov index je počítaný ako podiel veľkosti prieniku týchto množín a veľkosti zjednotenia týchto množín [5].

Entropia je mierou neusporiadanosti, neurčitosti, v našom prípade neurčitosti hesla. Vypočítame ju podľa nasledovného Shannonovho vzorca, kde H_0 je entropia, M je počet rôznych znakov reťazca a P_i je podiel počtu výskytu i-teho znaku v hesle a dĺžky hesla. [4]

$$H_0 = -\sum_{i=1}^{M} P_i \log(P_i)$$

Obrázok 1. Shannonov vzorec na výpočet entropie [4].

Priemer vzdialeností znakov hesla na klávesnici je priemer Euklidovských vzdialeností stredov susedných znakov v slove na vybranom modeli QWERTY klávesnice.



Obrázok 2. Vybraný model QWERTY klávesnice.

3 Záver

Naštudované sú podobné práce, aktuálne nástroje na testovanie sily prihlasovacích údajov a aj minimálne požiadavky na prihlasovacie údaje rôznych populárnych slovenských ale aj celosvetových webových služieb.

Dáta zo skúmaného datasetu sú očistené, upravené a s dorátanými pozorovanými vlastnosťami vložené do databázy. Aktuálne pracujeme na analýze týchto dát, chceme skúmať každú z týchto vlastností, tiež vzťahy medzi niektorými vlastnosťami, ako sú napríklad entropia hesla a dĺžka hesla alebo počet číslic v hesle a dĺžka najdlhšej súvislej postupnosti číslic v hesle. Chceme overiť hypotézu, že medzi týmito dvojicami vlastností existuje závislosť. Na základe týchto skúmaní chceme sformulovať pravidlá na tvorbu hesla, ktoré budeme považovať za bezpečné. Malo by to heslo také, ktoré má čo najmenej spoločných vlastností s heslami, aké máme v skúmanom datasete, teda s heslami, aké si ľudia bežne vytvárajú.

Po sformulovaní pravidiel na tvorbu bezpečného hesla navrhneme a implementujeme systém na testovanie bezpečnosti prihlasovacích údajov a overíme jeho funkčnosť.

Literatúra

1. ABDOU, AbdelRahman; BARRERA, David; VAN OORSCHOT, Paul C. What lies beneath? Analyzing automated SSH bruteforce attacks. In: *International Conference on Passwords*. Springer, Cham, 2015. p. 72-91.
2. SOKOL, Pavol; KOPČOVÁ, Veronika. Lessons Learned from Honeypots-Statistical Analysis of Logins and Passwords. In: *International Conference on Research and Practical Issues of Enterprise Information Systems*. Springer, Cham, 2016. p. 112-126.
3. JAEGER, David, et al. Analysis of publicly leaked credentials and the long story of password (Re-) use. In: *Proc. Int. Conf. Passwords*. 2016.
4. W. Ma, J. Campbell, D. Tran and D. Kleeman, "Password Entropy and Password Quality," *2010 Fourth International Conference on Network and System Security*, Melbourne, VIC, 2010, pp. 583-587.
5. NIWATTANAKUL, Suphakit, et al. Using of Jaccard coefficient for keywords similarity. In: *Proceedings of the international multiconference of engineers and computer scientists*. 2013. p. 380-384.
6. https://password.kaspersky
7. https://howsecureismypassword.net
9. http://www.passwordmeter.com