ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2014
Seria: ORGANIZACJA I ZARZĄDZANIA z. 68 Nr kol. 1905

Kamil KRANC, Ireneusz J. JÓŹWIAK, Jacek GRUBER
Politechnika Wrocławska
Wydział Informatyki i Zarządzania

BEZPIECZEŃSTWO HASEŁ WŚRÓD UŻYTKOWNIKÓW INTERNETU

Streszczenie. Niniejszy artykuł jest omówieniem wyników ankiety dotyczącej sposobów układania i obchodzenia się z hasłami przez internautów. Przedstawiono w niej negatywne trendy zauważalne wśród respondentów, takie jak układanie haseł zbyt krótkich, schematycznych, ich częste powtarzanie oraz nie weryfikowanie ich siły. Przeprowadzono również interpretację uzyskanych odpowiedzi pod kątem metod wykorzystywanych przez hakerów do przełamywania haseł.

Słowa kluczowe: bezpieczeństwo hasła, użytkownik Internetu, haker, unikalność hasła, schematyczność hasła.

PASSWORDS SECURITY AMONGST INTERNET USERS

Summary. In this paper the results of a survey on how stacking and handling of passwords from Internet are presented. Negative trends were discussed noticeable among the respondents, such as putting passwords are too short, the creation of passwords by popular templates, failure to verify password strength. Interpretation of the response in terms of the methods used by hackers to break passwords, was also carried out.

Keywords: password security, Internet user, hacker, password uniquenes, schematic of password.

1. Wprowadzenie

Od lat podstawowym sposobem na zabezpieczenie systemu komputerowego przed niepowołanym dostępem jest proces autoryzacji za pomocą identyfikacji i uwierzytelnienia [4]. Identyfikacji użytkownika dokonuje się przez wymaganie podania unikalnego loginu

w obrębie systemu, czyli nazwy użytkownika. Następnym krokiem jest weryfikacja zadeklarowanej przez użytkownika tożsamości przez uwierzytelnienie [2]. Najpopularniejszą metodą uwierzytelniania jest korzystanie z haseł, chociaż badania nad wykorzystaniem biometrii w procesach identyfikacji i uwierzytelniania są prowadzone od lat [9]. Od tego, czy hasło użytkownika jest silne i w jaki sposób jest przechowywane w dużej mierze zależy bezpieczeństwo danych, do których ma autoryzowany dostęp. Jeśli użytkownik układa hasła silne, nie powtarza ich w różnych systemach i nie zapisuje w żadnym łatwo dostępnym miejscu, wtedy zwiększa bezpieczeństwo danych. Jeżeli jednak układa hasła proste, często je powtarza, a w dodatku zapisuje je na karteczce przyklejonej do monitora, to otwiera drogę do systemu dla hakerów. Haker, posiadający właściwą parę login + hasło, dla systemu komputerowego, wykorzystującego tradycyjny proces identyfikacji i uwierzytelniania jest nieodróżnialny od użytkownika, pod którego się podszywa. Niezbędne są zatem starania, aby układane przez użytkowników hasła były silne i poufne. W niniejszym artykule prezentowane są wyniki badań na temat sposobów wymyślania haseł i obchodzenia się z nimi przez użytkowników Internetu.

2. Siła haseł

Hasło silne to takie, które trudno złamać. Definicja taka jest bardzo prosta, ale nie daje tak naprawdę żadnej konkretnej informacji. Aby ocenić siłę hasła, należy posłużyć się miarą. Najczęściej do tego wykorzystuje się entropię Shannona, dotyczącą ograniczenia na minimalną średnią długość słów kodowych w kodowaniu do zapisu symboli generowanych przez dyskretne źródło danych o określonej entropii, to znaczy o określonej średniej liczbie bitów, natów lub ditów na symbol. Entropia jest wyrażona następującym wzorem [7]:

$$H(X) := -\sum_{i} P(x_i) \log_b P(x_i) \tag{1}$$

gdzie: $P(x_i)$ to prawdopodobieństwo zajścia zdarzenia i,

b to podstawa logarytmu, gdy b=2, to jednostką entropii jest bit.

W praktyce wzór ogólny na entropię hasła wygląda następująco [3]:

$$H = \log_2 N^L = L \log_2 N \ , \tag{2}$$

gdzie:

L to długość hasła,

N to długość alfabetu, z którego ułożono hasło.

Przykładowo, polski alfabet składa się z 32 znaków. Hasło o długości 12 znaków wziętych z alfabetu polskiego ma entropię $12*\log_2 32 = 60$ bitów. Jeśli w tym samym alfabecie będziemy rozróżniać małe i wielkie litery, w efekcie wydłużając dwukrotnie alfabet, otrzymamy już entropię $12*\log_2 64 = 72$ bity. Tabela 1 prezentuje entropię pojedynczego znaku hasła w zależności od wykorzystanego alfabetu. Z takim obliczaniem siły haseł wiąże się jednak jedno założenie, że są one w pełni losowe, a to oznacza, że wyglądają przykładowo w taki sposób: „fA4#nzJ7C`v@3*", „}l6mEmJ2{fdlD&", „|K+DwFiNt5oo@". W praktyce hasła losowe wykorzystywane są rzadko. Jest to spowodowane tym, że ciąg znaków taki jak „|K,+DwFiNt5oo@" jest dla człowieka bardzo trudny do zapamiętania. Dlatego też użytkownicy wymyślają takie hasła, które są w stanie zapamiętać, co przekłada się do obniżenia entropii.

Tabela 1

Długości różnych alfabetów i entropia ich znaków

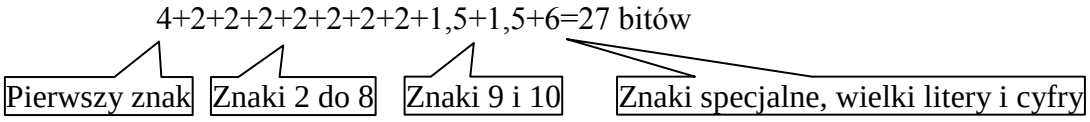
Alfabet	Długość alfabetu N	Entropia na znak hasła H [bit]
Cyfry arabskie	10	3,333
Kod szesnastkowy	16	4,000
Łaciński (bez rozróżniania wielkości liter)	26	4,700
Polski (bez rozróżniania wielkości liter)	32	5,000
Alfanumeryczny łaciński (bez rozróżniania wielkości liter)	36	5,170
Alfanumeryczny polski (bez rozróżniania wielkości liter)	42	5,390
Łaciński (wielkie i małe litery)	52	5,700
Alfanumeryczny łaciński (wielkie i małe litery)	62	5,950
Polski (wielkie i małe litery)	64	6,000
Alfanumeryczny polski (wielkie i małe litery)	74	6,210
ASCII	95	6,570

W przypadku haseł układanych przez ludzi, ich siłę można nie tyle obliczyć, co oszacować. Jedną z metod takich szacunków zaproponował NIST, czyli National Institute of Standards and Technology [3]. Jest to Amerykańska Agencja Federalna spełniająca funkcję analogiczną do polskiego Głównego Urzędu Miar. W swoim dokumencie SP 800-63-1, agencja proponuje następujący sposób oceny entropii haseł układanych przez człowieka [3]:

- Entropia pierwszego znaku hasła wynosi 4 bity.
- Entropia kolejnych 7 znaków wynosi 2 bity na znak.
- Dla znaków od 9 do 20 entropia pojedynczego znaku wynosi 1,5 bita.

- Dla każdego znaku powyżej dwudziestego entropia wynosi 1 bit.
- Dodatkowe 6 bitów entropii hasło uzyskuje, jeśli zawiera małe i wielkie litery oraz znaki specjalne.

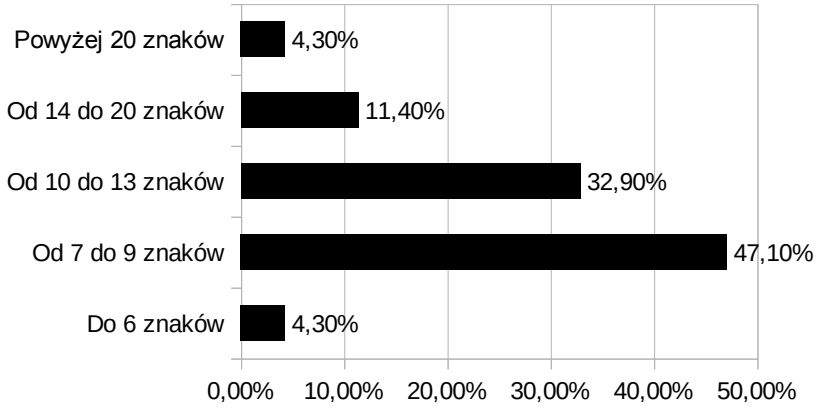
Hasło może być wzmocnione maksymalnie o 6 dodatkowych bitów, jeśli przy jego kreowaniu zostało sprawdzone, czy nie występuje w słowniku. Słownik rozumiany jest tutaj jako jakiś zbiór haseł, tzw. czarna lista. Dodatkowe bity entropii zależą od dokładności przeprowadzonego testu słownikowego. Ponadto, reguła zakłada, że długie hasła (powyżej 20 znaków), w celu zapewnienia możliwości ich zapamiętania, muszą być zwrotem, przez co ich składowe zawierają się w słownikach i tym samym hasło takie nie otrzymuje dodatkowych bitów z tej kategorii. Powyższa metoda zakłada korzystanie z alfabetu i reguł budowy słów w języku angielskim. Zawartość informacyjna języków zawiera się jednak w przedziale od 1 do 1,5, więc reguły dla innych języków niż angielski nie są drastycznie różne. Losowe hasło składające się z 10 znaków standardowego alfabetu ASCII ma entropię 65,7 bita, ale hasło o tej samej długości ułożone na tym samym alfabecie przez człowieka będzie już miało entropię:



3. Metodologia badań

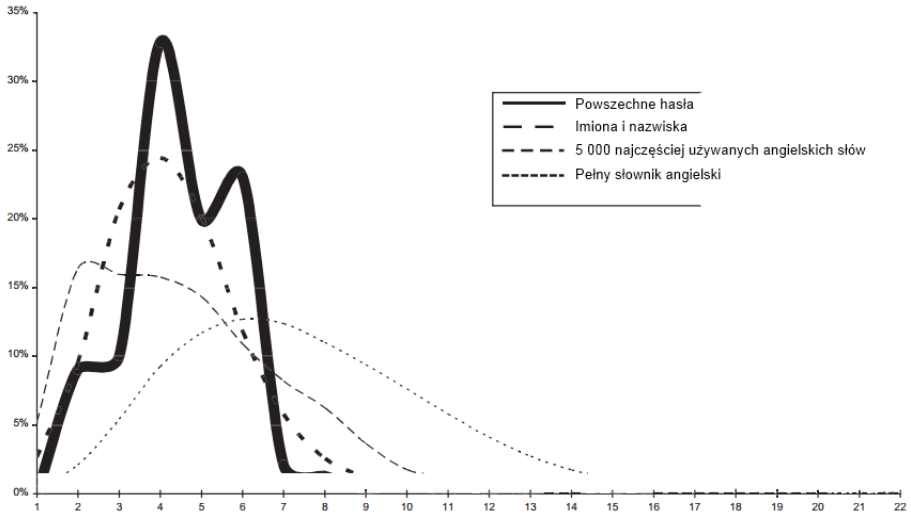
W kolejnych akapitach przedstawione zostaną najważniejsze wyniki ankiety przeprowadzonej na przełomie kwietnia i maja 2013 roku, wraz z ich omówieniem. Ankieta została przeprowadzona za pośrednictwem Internetu i wzięły w niej udział 144 osoby. Tak mała grupa ankietowanych, w dodatku w dużej mierze przypadkowych, nie pozwala na uznanie wyników za reprezentatywne, niemniej w odpowiedziach ankietowanych widoczne są wyraźne trendy.

Odpowiedzi na pytanie o długość układanych haseł ankietowani odpowiedzieli tak, jak zaprezentowano na rysunku 1. Wyniki pokazują, że użytkownicy preferują hasła krótkie.

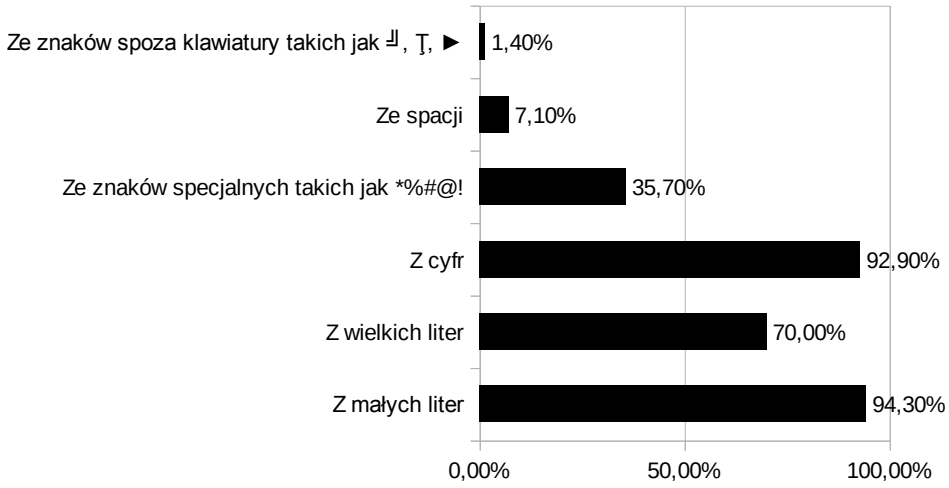


Rys. 1. Długość haseł układanych przez respondentów
Źródło: opracowanie własne.
Fig. 1. Length of passwords created by respondents

Przy obecnej mocy obliczeniowej komputerów zbyt krótkie hasła nie można uznać za bezpieczne. Dodatkowo, hakerzy często przeprowadzają ataki słownikowe, zamiast ataków typu brute-force. Badania pokazują, że hasła znajdujące się w takich słownikach sięgają 14 znaków [1], co pokazuje rys. 2. Tym samym dopiero układanie haseł o długości powyżej 14 znaków zapewnia ochronę przed atakami słownikowymi. Jedynie niecałe 16% ankietowanych deklarowało korzystanie z tak długich haseł.



Rys. 2. Procentowy rozkład długości wyrażeń w znakach [1]
Fig. 2. Percentage distribution of the length of phrases in characters



Rys. 3. Odpowiedzi na pytanie: z jakich znaków korzystasz przy układaniu haseł
Źródło: opracowanie własne.
Fig. 3. Answers to the question: which characters you use when creating passwords

Im większy jest alfabet, na którym zbudowano hasło, tym większa jest jego entropia, co wprost wynika ze wzoru (2). Mimo to, wielu użytkowników nie stosuje się do tej zasady. W wyniku analizy milionów haseł odkryto, że blisko 90% używanych haseł składa się jedynie z małych liter i cyfr [10]. Ponadto, jeśli już w haśle pojawi się wielka litera, to zazwyczaj na pierwszej lub drugiej pozycji [1]. Rysunek nr 3 pokazuje wyniki własnych badań, tzn. że sytuacja ta nie ulega zmianie. Znaki specjalne wykorzystywane są jedynie przez niecałe 36% ankietowanych. Znak spacji, pozwalający w prosty sposób układać długie, wieloczłonowe hasła, wykorzystywany jest zaledwie przez 7% respondentów. Wykorzystanie znaków spoza klawiatury (np. w systemie Windows przez wciśnięcie kombinacji Alt+XXXX, gdzie każdy X odpowiada jednej cyfrze z klawiatury numerycznej, pozwala na wprowadzenie znaku spoza drukowalnej tabeli ASCII) jest znikome. Zapewne jest to związane w dużej części z restrykcjami serwisów webowych, niezezwalających na wykorzystywanie znaków spoza klawiatury, niemniej znaki tego typu z powodzeniem mogą być wykorzystywane w obrębie systemów operacyjnych do układania haseł unikalnych, nieznajdujących się w hakerskich słownikach.

Najpopularniejszymi schematami, wedle których budowane są hasła, a więc takie, których nigdy nie powinno się wykorzystywać, wg Burnetta i Kleimana [1] są:

1. korzystanie z wyrazów słownikowych, czyli wykorzystywanie jako hasła słowa, które znajduje się w takim, czy innym spisie: słowniku, liście imion, krajów itd. Takie hasła bardzo łatwo złamać przez atak słownikowy, czyli sprawdzanie po kolei słów w takim, czy innym słowniku. O ile dla człowieka wpisanie po kolei wszystkich słów z języka angielskiego zajęłoby bardzo dużo czasu, to dla współczesnych komputerów nie jest to żaden problem.

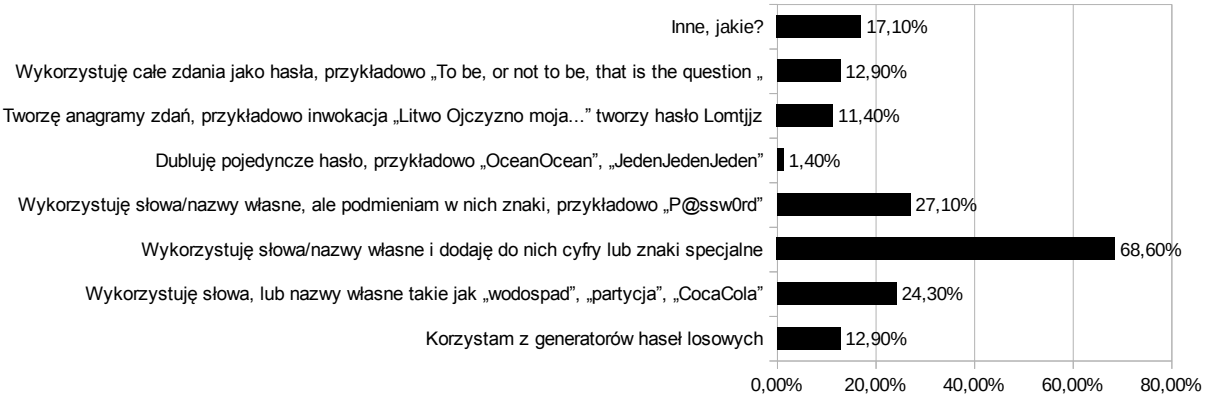
2. Wyrazy słownikowe z dodanymi cyframi. Teoretycznie dodanie cyfr do hasła, czyli wydłużenie alfabetu na którym powstało o 10 znaków zwiększa jego bezpieczeństwo. Istniejące wyniki badań pokazują jednak, że aż 64% haseł zawierających cyfry ma je na samy

końcu hasła, a 6% na początku [10]. W efekcie hasła te mają następującą postać: „Monika92”, „Porsche64”, „0kaloryfer”. Popularne narzędzia hakerskie, takie jak Jack the Ripper sprawdzają tego typu kombinacje, toteż zwiększenie bezpieczeństwa hasła w ten sposób jest minimalne.

3. Proste podstawienia. Jest to metoda wyraźnie bezpieczniejsza niż korzystanie z wyrazów słownikowych czy proste dodanie cyfr, lecz zbyt często wykorzystywana, aby pozostawała skuteczna. Polega ona na zamianie liter w słowie na inne znaki, przykładowo: „h45ł0”, „K4r0l1n4”, czy „w@k@cj3”. Tak jak w schemacie numer 3, programy hakerskie sprawdzają proste podstawienia.

4. Dublowanie hasła. Schemat opiera się na układaniu haseł takich „haslohaslo”, „schodyschody”, „icecreamsicecreams”. Hakerom takie techniki są doskonale znane i odpowiednio dostosowują do nich swoje narzędzia, przez co dublowanie słów nie daje dużej poprawy siły hasła.

Przeprowadzone badania wykazały, że wciąż najpopularniejszym schematem układania haseł jest schemat drugi. Schematy pierwszy i trzeci mają zbliżoną popularność, za to wśród ankietowanych schemat czwarty był niemal nieużywany. Popularność wykorzystywanych schematów pokazuje rysunek 4. Pod pozycją inne, ankietowani udzielali odpowiedzi takich, jak: wykorzystywanie wyrazów obcojęzycznych, literowanie słów angielskojęzycznych w sposób fonetyczny (np. CAT=kejejti), korzystanie z sekwencji klawiszy na klawiaturze (piąty najpopularniejszy schemat wg Burnetta i Kleimana [1]) oraz własnoręczne układania i zapamiętywanie pseudolosowych ciągów znaków.



Rys. 4. Odpowiedzi na pytanie: czy korzystasz z któryś z podanych sposobów układania haseł
Źródło: opracowanie własne.
Fig. 4. Answers to the question: are you using any of the following methods to create passwords

Z powyższym wiąże się brak nawyku sprawdzania siły ułożonego hasła przez ankietowanych. Pomimo tego że narzędzia obliczające siłę haseł są bardzo łatwo dostępne w Internecie, jedynie około 25% respondentów zadeklarowało regularne korzystanie z nich. Pozostali używali ich sporadycznie albo w ogóle. I tak na pytanie czy układając hasła starasz

się korzystać z dostępnych narzędzi oceniających ich siłę? ankietowani odpowiadali: zawsze, przeważnie zawsze korzystam 5,70%, odpowiedzi, przeważnie zawsze korzystam 21,40% odpowiedzi, rzadko korzystam 40,00% odpowiedzi, nigdy nie korzystam 32,90% odpowiedzi.

Ułożenie silnego hasła to jeszcze za mało, aby zapewnić bezpieczeństwo autoryzacji. Pozostaje kwestia poprawnego przechowywania i wykorzystywania swoich haseł. Ludzie mają tendencję do używania jednego klucza dostępu do wielu systemów: serwisów społecznościowych, komunikatorów, sklepów internetowych itd. [8]. W efekcie włamanie hakera do jednego z tych systemów powoduje kompromitację wielu innych. Tym samym hakerzy są zainteresowani dostępem do haseł dowolnego serwisu. Użytkownicy często wychodzą z założenia, że ich konto na jakimś forum internetowym dla cyber-przestępców nie przedstawia żadnej wartości, ale są w błędzie. Przejęcie jakiegokolwiek konta pozwala hakerom na podszycie się pod inną osobę w sieci i maskowanie przez to własnej tożsamości. Zdobyte hasła pozwalają na aktualizowanie wykorzystywanych do łamania zabezpieczeń słowników. Przede wszystkim jednak zdobycie pary login + hasło pozwala na wykorzystanie takiej pary do hakowania innych serwisów. W dobie rozrywki elektronicznej istnieje wiele portali i gier internetowych wymagających założenia swojego konta. Firmy, do których należą takie serwisy, a dokładniej ich specjaliści od bezpieczeństwa zauważają w ostatnich latach nowy trend w działaniach hakerów. Zamiast przeprowadzać zmasowany atak i testować tysiące haseł, hakerzy podają bardzo konkretne zestawy loginów i kluczy dostępów, przykładowo „userKowalski@poczta.pl" - „aligator100". Jeżeli hasło się nie zgadza, wypróbowywane zostają jego proste wariacje, po czym procedura przechodzi do następnej nazwy użytkownika itd. Podawane są głównie loginy, które nawet nie istnieją w danym serwisie. Dzieje się tak, ponieważ hakerzy już znają parę login – hasło z innych serwisów, które zostały wcześniej zhakowane [5]. Korporacja Sony, twórca konsoli PlayStation i powiązanej z nią usługi PlayStation Network oraz właściciel Sony Online Entertainment uległa atakowi, przez który w 2011 roku wyciekły dane 24,6 miliona kont SOE i 70 milionów kont PSN. Wykradzione dane zawierały między innymi adresy e-mail i numery kart kredytowych [6]. Podobnym atakom uległy m.in. spółki Blizzard oraz Valve, właściciele bardzo popularnych platform do gier sieciowych BattleNet i Steam. Każdy taki wyciek danych to miliony skompromitowanych kont, których dane bardzo szybko wykorzystywane są do kolejnych włamań w samonapędzającej się spirali.

W obliczu powyższej prawidłowości bardzo istotne jest, aby użytkownicy wymyślali indywidualne hasła dla różnych serwisów. Wyniki ankiety pokazały jednak, że rzeczywistość jest daleka od takiego ideału. Na pytanie „w ilu miejscach (serwisy internetowe, konta e-mail, konta systemowe) zdarza ci się powtarzać lub wykorzystywać jedno hasło?" tylko 15,70% respondentów układa unikalne hasła dla wszystkich portali, a 25,70% wykorzystuje jedno hasło w więcej niż pięciu miejscach, a 18,60% wykorzystuje jedno hasło w 4 do 5 miejscach, natomiast 40,00% respondentów wykorzystuje hasło w 2 do 3 miejscach.

Ankietowani większą wagę przywiązywali do bezpieczeństwa haseł w serwisach finansowych (sklepy internetowe, bankowość elektroniczna itp.) oraz zawierających ich prywatne dane (np. portale społecznościowe), gdzie większość osób zadeklarowała wymyślanie unikalnych haseł. Jednakże 8,60% respondentów nie uważa za stosowne układania indywidualnych haseł do tego typu serwisów, a 34,3% zdaża się powtarzać wymyślone na potrzeby takich portali hasła w innych witrynach. Pozostałe osoby w ilości 57,1% dla każdego tego typu serwisu finansowego tworzą inne hasło.

Ankietowane osoby nie miały także w zwyczaju okresowego zmieniania haseł w serwisach finansowych i społecznościowych. Prawie 50%, tzn. 48,6% respondentów w ogóle nie zmienia swoich haseł, a jedynie 4,3% zmienia je co kilka tygodni, 5,7% zmienia je co miesiąc, 12,9% co kwartał, 10% co pół roku, a pozostała ilość – 18,6% zmienia je co rok lub rzadziej. Częstotliwość zmian haseł często wymusza polityka bezpieczeństwa firmy, w których pracują respondenci.

4. Podsumowanie

Podsumowując, błędy popełniane przez użytkowników w procesie układania haseł są systematyczne, z czego korzystają hakerzy. Znają oni najpowszechniej wykorzystywane hasła oraz schematy ich powstawania, co obniża bezpieczeństwo systemów komputerowych. Błędy ludzkie dotyczące generowania kluczy dostępu powstają na etapie planowania (ułożenie słabego hasła) i przygotowania (niepoprawne przechowywanie hasła). Kluczowe wytyczne dotyczące wymyślania haseł to zapewnienie im stosownej długości i oparcie ich na jak najdłuższym alfabecie. Niestety klucze dostępu układane przez ludzi nie są tak silne, jak te generowane losowo, dlatego wymuszenie ich odpowiedniej złożoności obliczeniowej na wypadek prób ich łamania jest bardzo istotne. Ponieważ współcześnie sieci komputerowe są wszechobecne, a co za tym idzie istnieje wiele usług wymagających autoryzacji użytkownika, hakerzy mogą przeprowadzać ataki na wiele serwisów, z których korzysta użytkownik. Jeśli dana osoba wykorzystuje jeden klucz dostępu do licznych usług, włamanie do pojedynczego serwera kompromituje bezpieczeństwo wszystkich usług współdzielących klucz. Korzystanie z unikalnych haseł jest zatem priorytetowe. W opinii autorów pracy, należy przekonywać użytkowników do wymyślania dłuższych haseł opartych na zdaniach, niż zmuszać ich do korzystania z 12-znakowych ciągów o narzuconej skomplikowanej strukturze, którą to i tak schematycznie upraszczają. Ponadto, należy przekonywać ludzi, by nie wykorzystywali jednego hasła w wielu serwisach. Biorąc pod uwagę fakt, że hakerzy powszechnie korzystają z ataków słownikowych i gotowych baz danych wykradzionych haseł, posiadanie różnych haseł w postaci prostych wyrażeń o długości ponad 15 znaków, w rodzaju 'Mam na biurku drukarkę', jest lepszym rozwiązaniem niż zapamiętanie jednego ciągu "+rUb@dUr2y!" i jego powtarzanie.

Bibliografia

1. Burnett M., Kleiman D.: Perfect Passwords. Syngress Publishing Inc., Rockland 2006.
2. Haley K.: Password Survey Results, http://www.symantec.com/connect/blogs/password-survey-results. [dostęp z dnia 04.03.2013].
3. NIST Special Publication 800-63-1, 2011.
4. Norma ISO/IEC 27000:2009.
5. O'Brien M., Mike O'Brien on Account Security, https://www.guildwars2.com/en/news/mike-obrien-on-account-security/. [dostęp 02.03.2013].
6. Schreier J.: Sony Hacked Again; 25 Million Entertainment Users' Info at Risk, http://www.wired.com/gamelife/2011/05/sony-online-entertainment-hack/ [dostęp 02.03.2013].
7. Shannon C.E.: A Mathematical Theory of Communication. Bell System Technical Journal, 1948, Vol. 3, No. 27, p. 379-423.
8. Sophos Online Password Survey, 2009.
9. Tuyls P., Goseling J.: Capacity and Examples of Template-Protecting Biometric Authentication System. Lecture Notes in Computer Science, Springer, Heidelberg 2004, Vol. 3087, p. 158-170.
10. Weir M., Aggarwal S., Collins M., Stern H.: Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords. Proceeding CCS '10. Proceedings of the 17th ACM Conference on Computer and Communications security, ACM New York, USA 2010, p. 162-175.

Abstract

One of the main problems of security of information systems is discussed in this article. The analysis of responses to a questionnaire on the creation and use of passwords are presented. Very careful analysis was performed on responses related to the creation and use of passwords to applications and web-based systems. These responses were interpreted in the context of the current methods used by hackers to crack passwords. The results show a negative trend noticeable among the respondents, such as creating passwords too short passwords, passwords that are easy for hackers to crack using brute force, the creation of passwords by popular templates, and failure to verify password strength.