Biuletyn WAT
Vol. LXVIII, Nr 2, 2019

Możliwości pozyskiwania adresów e-mail z serwisów internetowych używających Gravatara

PRZEMYSŁAW RODWALD

Akademia Marynarki Wojennej, Wydział Nawigacji i Uzbrojenia Okrętowego,
Instytut Uzbrojenia Okrętowego i Informatyki, ul. Śmidowicza 69, 81-127 Gdynia,
p.rodwald@amw.gdynia.pl

Streszczenie. Mimo że Internet nie został stworzony z myślą o anonimowości, wielu użytkowników publikujących komentarze na forach dyskusyjnych pod różnymi pseudonimami ma nadzieję, że pozostaną anonimowi. W artykule przedstawiony został atak, którego celem jest ujawnienie adresów e-mail użytkowników wybranego serwisu internetowego wykorzystującego Gravatara. Pokazana została metodologia przeprowadzenia ataku oraz zaprezentowane zostały jego rezultaty. Atak ten ma na celu uświadomienie czytelników, że publikując komentarze na stronach korzystających z usług Gravatara, są narażeni na ujawnienie adresów e-mail, a w niektórych przypadkach nawet na pełną identyfikację.
Słowa kluczowe: Gravatar, deanonimizacja, adres e-mail, funkcja skrótu MD5
DOI: 10.5604/01.3001.0013.3003

1. Wstęp

Gravatar to usługa udostępniająca avatary na różnych stronach internetowych. Idea odtwarzania adresów e-mail z postaci, w jakiej przechowuje je serwis Gravatar, została uprzednio zaprezentowana co najmniej dwukrotnie: w roku 2009 [1] autor ukrywający się pod pseudonimem abell po przeszukaniu 80000 skrótów MD5 Gravatara odtworzył 10% adresów e-mail ze strony stackoverflow.com oraz w roku 2013 [2] Bongard po pobraniu 2400 hashy[1] MD5 Gravatara z francuskiego serwisu blogowego fdesouche.com odtworzył 70% adresów e-mail jego użytkowników.

[1] *Hash* to angielskie określenie, które jest już dość powszechne w języku polskim. Oznacza skrót generowany za pomocą kryptograficznej funkcji skrótu.

Główne cele tego artykułu to: dostarczenie szczegółowej metodologii ataku pozwalającego ujawniać adresy e-mail w serwisach internetowych używających Gravatara; pokazanie wyników ataku przeprowadzonego przez autora na wybrany polskojęzyczny serwis blogowy oraz przedstawienie wniosków dotyczących możliwości deanonimizacji użytkowników w oparciu o otrzymane wyniki.

2. Gravatar

Słowo Gravatar to akronim od angielskiego tłumaczenia Globalnie Rozpoznawalnego Avatara (ang. *Globally Recognized AVATAR*). Natomiast avatar to obrazek, który reprezentuje użytkownika online. Będąc bardziej precyzyjnym, to zdjęcie lub wygenerowany obrazek, które pojawiają się przy adresie e-mail bądź nazwie użytkownika, kiedy ten umieszcza komentarze na różnych stronach. Gravatar jest zintegrowany z WordPressem[2], co implikuje jego popularność w Internecie. Szacuje się, że ponad 30% stron internetowych oparta jest właśnie na WordPressie, przy czym dla stron bazujących na CMS-ie (ang. *Content Management System*) udział ten wynosi blisko 60% [3]. Gravatar jest również używany przez wiele innych popularnych serwisów takich jak wspomniany już StackOverflow.com czy TechDirt.com [4].

Użytkownik chcący świadomie korzystać z serwisu Gravatar działa w następujący sposób: tworzy konto na stronie internetowej gravatar.com, następnie przesyła obrazek (avatar) i przypisuje do niego własny adres e-mail. Natomiast każdy użytkownik, który chce umieścić komentarz na stronie wykorzystującej Gravatara (niezależnie od tego, czy ma on konto w serwisie Gravatar, czy też nie), musi oprócz samego komentarza podać swoje imię lub pseudonim oraz adres e-mail. Często istnieje dodatkowa informacja, że adres e-mail nie będzie udostępniany i służy tylko do wyświetlania Gravatara, co jak zostanie pokazane w dalszej części artykułu, nie jest do końca prawdą. Z perspektywy programisty natomiast, aby móc pobierać Gravatary użytkownika i wyświetlać je na tworzonej stronie internetowej, serwis internetowy musi wysłać żądanie do strony gravatar.com. Żądanie to jest oparte na prostym poleceniu HTTP GET i nie jest przy tym wymagane uwierzytelnianie. Strona musi najpierw wygenerować skrót MD5 adresu e-mail użytkownika (HASH), a następnie zażądać avatara przy użyciu określonego adresu URL w formacie: https://www.gravatar.com/avatar/HASH. Na przykład, jeśli adres e-mail użytkownika to p.rodwald@amw.gdynia.pl, to żądanie wygląda następująco: https://www.gravatar.com/avatar/0aec9b599eeb18ae640234f62683104b. Jeśli dla danego adresu e-mail w serwisie Gravatar istnieje utworzone konto i przypisane do niego zdjęcie,

[2] WordPress – popularny system zarządzania treścią (CMS) zaprojektowany głównie do obsługi blogów.

wówczas jest ono wyświetlane na stronie. Jeśli jednak dla danego adresu e-mail konto nie istnieje, wówczas wyświetlany jest wybrany motyw graficzny serwisu Gravatar. Warty podkreślenia jest następujący fakt: mimo że adres e-mail nie jest wyświetlany na stronie ani nie znajduje się w kodzie strony w postaci jawnej, to jednak dla każdego publikowanego komentarza znajduje się on w kodzie strony w postaci skrótu MD5. Właśnie to wykorzystanie kryptograficznej funkcji skrótu MD5 jako unikalnego identyfikatora dla adresów e-mail użytkowników jest jednym z poważniejszych mankamentów Gravatara, jeśli chodzi o zapewnienie anonimowości jego użytkownikom. Algorytm MD5 został zaprojektowany przez Ronalda Rivesta w 1991 roku [5]. Z kryptograficznego punktu widzenia algorytm ten jest uważany za skompromitowany. W 2004 r. przedstawiony został pierwszy atak na znalezienie kolizji [6]. Od tego czasu zaprezentowano prawdziwe kolizje MD5 dla: certyfikatów X.509 z różnymi kluczami publicznymi [7], plików wykonywalnych [8], plików pdf [9] czy plików jpeg [10]. Jak dotąd nie jest znana żadna kolizja MD5 dla adresów e-mail. Mimo że algorytmu MD5 w wielu zastosowaniach nie powinno się już używać [11], to wciąż jest jednym z popularniejszych, szczególnie jako mechanizm przechowywania haseł. Niefortunność użycia algorytmu MD5 w Gravatarze nie polega jednak na problemach związanych z jego bezpieczeństwem (łatwość znajdowania kolizji). Przyczyna leży w postępie technologicznym i związanym z nim wzrostem możliwości obliczeniowej atakującego. Współczesne karty graficzne (na przykład Nvidia RTX 2080 FE) osiągają wydajność 50 GH/s[3] dla algorytmu MD5 [12].

Przechodząc do teoretycznych rozważań, spróbujmy odpowiedzieć na pytanie, czy Gravatar jest bijekcją (funkcją „jeden na jeden"). Z matematycznego punktu widzenia można próbować obliczyć liczbę możliwych adresów e-mail. Adres e-mail składa się z trzech elementów: nazwy użytkownika, @ (znaku tak zwanej „małpki") i części domenowej. Zgodnie z dokumentami RFC [13], [14] nazwa użytkownika może mieć maksymalną długość do 64 znaków, a część domenowa do 253 znaków. Zakładając (co jest pewnym uproszczeniem, a szczegóły są wyjaśnione w dalszej części artykułu), że nazwa użytkownika i część domenowa składają się tylko z małych liter, cyfr i znaków specjalnych: „.", „-" oraz „_", można oszacować całkowitą możliwą liczbę adresów e-mail jako 39^{64} x $39^{253} \approx 2^{1676}$ różnych wartości (oznaczmy to jako e). Z kryptograficznego punktu widzenia wiadomo, że długość skrótu generowanego przez kryptograficzną funkcję skrótu MD5 wynosi 128 bitów, co daje 2^{128} wszystkich możliwych wartości skrótów (oznaczmy to jako g). Można zauważyć, że $e >> g$, co może prowadzić do teoretycznego wniosku, że Gravatar jest surjekcją (funkcją „wiele na jeden"), zakładając jednak przy tym, że wszystkie adresy e-mail są używane i do tego są rozłożone równomiernie. Założenia te jednak w rzeczywistości nie są prawdziwe. Po pierwsze, liczność części domenowej w świecie rzeczywistym

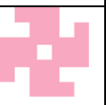
[3] GH/s – 10^9 skrótów (hashy) na sekundę.

nie powinna być aproksymowana do wartości 39^{253}, a jedynie do około 350 mln faktycznie zarejestrowanych domen [15]. Po drugie, szacowana liczba kont e-mail w roku 2018 wynosiła 6690 mln[4]. Ponad połowa ludności świata korzysta z poczty elektronicznej, przy blisko dwóch[5] kontach e-mail przypadających na jednego użytkownika [16]. Liczba rzeczywistych adresów e-mail jest więc mniejsza niż 2^{33}, a co za tym idzie $2^{33} << g$. Prowadzi to do wniosku, że w rzeczywistości Gravatar może być bijekcją, czyli każdemu rzeczywistemu adresowi e-mail odpowiada unikalny Gravatar[6].

Gravatar nie jest jedynym systemem oferującym funkcjonalność podobną do tej opisanej powyżej. Wybrane serwisy zostały zaprezentowane w tabeli 1.

Tabela 1

Wybrane serwisy oferujące usługę avatara

lp	strona www	nazwa	algorytm	przykładowe wywołanie	przykład
1	gravatar.com	globally recognized avatar	MD5	gravatar.com/avatar/ 547d20f2c04a3dc4 838aae94b1ff06e1	
2	evatar.io	email-linked avatar	SHA256	evatar.io/8d5218972a45ba 5309db7d70f3373d4bfdbae04 0a336df4091cab8b66f642149	
3	dicebear.com	pixel-art avatar	Tekst jawny	avatars.dicebear.com/ v2/male/art@wp.pl.svg	
4	github.com/ tobiaslins/ avatar	gradient avatar images	Tekst jawny	avatar.tobi.sh/art@wp.pl	
5	robohash.org	robohash - robot images	Tekst jawny	robohash.org/art@wp.pl	
6	adorable.io	adorable avatar	Tekst jawny	api.adorable.io/avatars/ 100/art@wp.pl.png	

[4] Wartość ta została obliczona jako iloczyn: ogólnoświatowej liczby użytkowników poczty e-mail (3823 mln) oraz średniej liczby kont e-mail przypadających na jednego użytkownika (1,75). Dane pochodzą z [16].

[5] Liczba kont e-mail przypadających na jednego użytkownika w 2018 roku szacowana jest na 1,75, natomiast w 2022 roku estymowana jest na 1,86 [16].

[6] Twierdzenie to pozostanie prawdziwe, dopóki żadna kolizja MD5 dla adresów e-mail nie zostanie znaleziona.

Niektóre z serwisów (oznaczone numerami 3-6 w tabeli 1) oferują tylko usługę generowania pewnej zindywidualizowanej grafiki (avatara, gradientu kolorów) dla zadanego ciągu wejściowego, który nie musi być adresem e-mail. W serwisach tych wywołanie avatara następuje poprzez przekazanie adresu e-mail (lub dowolnego innego ciągu) w postaci jawnej. Serwisy oznaczone numerami 1 oraz 2 w tabeli 1 oprócz generowania elementu graficznego umożliwiają także przypisanie własnej grafiki (zdjęcia) do konkretnego adresu e-mail, a wywołanie avatara następuje poprzez przekazanie adresu e-mail w postaci skrótu (MD5 lub SHA-256).

3. Metodologia ataku

Zaprojektowany i przeprowadzony przez autora atak składa się z trzech głównych fraz: 1 – wyodrębnianie skrótów MD5, 2 – badania statystyczne, 3 – wykonanie ataku.

3.1. Faza 1 – wyodrębnianie skrótów MD5 Gravatar

Jako jeden z pierwszych etapów została wykonana analiza znanych polskich blogów i forów internetowych pod kątem używania przez nie Gravatara do prezentacji avatarów użytkowników. W rezultacie wybrano stronę internetową o adresie jakoszczedzacpieniadze.pl (polski znany blog o oszczędzaniu pieniędzy). Dokonano analizy struktury kodu HTML stron tego serwisu. Fragment kodu odpowiedzialny za prezentowanie Gravatarów użytkowników na poszczególnych podstronach serwisu zaprezentowany został jako kod 1.

```
...
<dt class="comment even thread-odd thread-alt depth-1" id="comment-375475">
<span class="avatar">
<img
src='https://secure.gravatar.com/avatar/b3272d913b1f8dd5e416b9ad1ded89ff?s=66&r=g'
alt='' class='avatar avatar-66 photo' height='66' width='66' />
</span>
<span class="comment_author">milosz</span>
…
</dt>
...
```

Kod 1. Fragment kodu HTML odpowiedzialny za wyświetlanie Gravatarów użytkowników w serwisie jakoszczedzacpieniadze.pl

Następnie uruchomiono zaprojektowanego pająka indeksującego (ang. *web-crawler*) w celu przeszukania wszystkich podstron dostępnych w wybranej domenie i wyodrębnienia z nich pseudonimów używanych przez użytkowników wraz ze

skrótami MD5 Gravatara. Uproszczony kod źródłowy pająka w języku PHP został przedstawiony jako kod 2. W rezultacie pająk indeksujący wyodrębnił 13935 unikatowych skrótów MD5 Gravatara wraz z pseudonimami. Wszystkie dane zostały zapisane w utworzonej na potrzeby ataku bazie danych MySQL.

```
<?php
$website = "jakoszczedzacpieniadze.pl";
$subpage = " jak-wypelnic-wniosek-rodzina-500plus";    //sample subpage
$html_content = file_get_contents("http://".$website."/".$subpage);
$comment_beg = "<dt ";
$comment_end = "</dt>";
$author_beg  = "<span class=\"comment_author\"> ";
$author_end  = "</span>";
$comm = returnSubstrings($html_content, $comment_beg, $comment_end);
if ($comm)  {
  for ($i = 0; $i < count($comm); $i++)     {
   $gravatar=substr($comm[$i],strpos($comm[$i],"gravatar.com/avatar/")+20,32);
   $nickname = returnSubstrings($comm[$i], $author_beg, $author_end);
     echo $gravatar." - ".$nickname;
}}
?>
```

Kod 2. Uproszczony pająk indeksujący

3.2. Faza 2 – statystyki polskich adresów e-mail

Jak określono w RFC6531 [13], adresy e-mail mają określoną strukturę: część zwaną nazwą użytkownika i część domenową oddzieloną ogranicznikiem @. Część domenowa jest zgodna z regułami nazewnictwa domen internetowych określonymi w RFC1035 [14] i może składać się z maksymalnie 253 znaków ASCII. Natomiast maksymalna długość nazwy użytkownika to 64 znaki, które są ograniczone do podzbioru znaków ASCII: wielkie i małe litery łacińskie (a-z, A-Z); cyfry (0-9); znaki specjalne (# - _ ~! $ & ′() * +,; =:) oraz znak kropki „.”, o ile nie jest to pierwszy lub ostatni znak. Inne znaki specjalne są dozwolone, ale używane tylko w konkretnych przypadkach, które można zignorować w tych rozważaniach.

Atak brutalny (ang. *brute-force attack*) [18] użyty bezpośrednio w celu łamania adresów e-mail jest nieskuteczny, ponieważ rozmiar typowego adresu e-mail jest znacznie większy niż 9-10 znaków. Można jednak skorzystać z połączenia ataku słownikowego (ang. *dictionary attack*) oraz mocy reguł dostarczanych przez wyspecjalizowane oprogramowanie, takie jak na przykład Hashcat czy John the Ripper.

Atak łączący obie wspomniane techniki nosi nazwę ataku kombinacyjnego (ang. *combination attack*). Aby przygotować atak, przeprowadzone zostały dwa dodatkowe badania: analiza statystyczna najpopularniejszych polskich domen oraz analiza wzorców występujących w nazwach użytkowników adresów e-mail.

Aby korzystać z reguł, należy utworzyć zestaw najpopularniejszych krajowych dostawców poczty e-mail. Do najpopularniejszych bezpłatnych dostawców poczty internetowej w Polsce należą: wp.pl, gmail.com, o2.pl, interia.pl, op.pl, poczta.onet.pl, vp.pl, tlen.pl, poczta.fm, gazeta.pl, onet.eu, interia.eu, onet.pl, hotmail.com, autograf.pl, go2.pl, yahoo.com, neostrada.pl, post.pl, outlook.com, home.pl, orange.pl. Lista ta jest oparta na badaniach statystycznych przeprowadzonych przez autora na podstawie trzech list e-mail-owych: bezpłatnej bazy danych e-maili znalezionej w Internecie (zawierającej 7 mln adresów e-mail z domeny .pl pochodzących z różnych wycieków) i dwóch baz pochodzących z istniejących prawdziwych polskich stron internetowych. Otrzymane wyniki są podobne do innych publikowanych zestawień statystycznych [19].

Przygotowując się do analizy wzorców dla nazw użytkowników, wykonano dodatkowe czynności polegające na: wyszukaniu list polskich imion i nazwisk, przygotowaniu słownika nicków internetowych (wykonane wcześniej przez pająka indeksującego) i rozpoznaniu wzorców w nazwach użytkowników. W Internecie można łatwo zlokalizować bazy zawierające: spis polskich imion (około 1200 rekordów) czy listę polskich nazwisk (około 400 tys. rekordów). Wykaz zidentyfikowanych najpopularniejszych wzorców nazw użytkowników wraz z przykładami przedstawiono w tabeli 2.

Tabela 2

Najpopularniejsze wzorce dla nazw użytkowników

Wzorzec	Przykłady
[nazwisko][?s][?d]{0,4}	rodwald, rodwald_1990
[imię][?s][?d]{0,4}	przemek, przemek-007
[imię][?s][nazwisko][?d]{0,4}	przemekrodwald, przemek_rodwald2017
[nazwisko][?s][imię][?d]{0,4}	rodwald.przemek, rodwald_przemek77
[?l][?s][nazwisko][?d]{0,4}	prodwald, p.rodwald, p_rodwald33
[nazwisko][?s][?l][?d]{0,4}	rodwaldp, rodwald_p, rodwald-p01
[pseudonim][?s][?d]{0,4}	zipper, ziuta_82
[słowo][?s][?d]{0,4}	polska, kwiatek-1111
legenda: ?d- cyfra, ?l- litera, ?s- jeden ze znaków{ , ., -, _}	

3.3. Faza 3 — atak

Po przeprowadzeniu dwóch powyższych etapów przygotowawczych, przystąpiono do przeprowadzenia samego ataku. Został on podzielony na trzy etapy: łamanie brutalne, łamanie kombinacyjne i przeszukanie adresów e-mail pozyskanych z różnych wycieków.

Etap pierwszy — atak brutalny

Atak siłowy, zwany też brutalnym, polega na obliczaniu funkcji skrótu MD5 dla każdej możliwej kombinacji liter, cyfr, znaków reprezentującej nazwę użytkownika w adresie e-mail w połączeniu z najpopularniejszymi domenami (w niniejszym ataku wzięto pod uwagę dziesięć najpopularniejszych domen). Obliczony skrót MD5 podlega sprawdzeniu i ustaleniu, czy odpowiada on skrótowi MD5 Gravatara. Liczba obliczeń uzależniona jest od długości nazwy użytkownika oraz przestrzeni znaków w niej występujących. Przykładowo dla co najwyżej 9-znakowej nazwy użytkownika składającej się z liter, cyfr i znaków specjalnych uzyskujemy 2688656×10^9 możliwych skrótów (oznaczanych często jako GH – liczba gigahashy). Natomiast dla 10-znakowej nazwy użytkownika składającej się wyłącznie z liter należy przeszukać 1468138 GH. Efektywność ataku siłowego zależy od następujących czynników: platformy, którą wykorzystuje atakujący, czasu, który może poświęcić na atak, oraz środków finansowych, którymi dysponuje (związanych głównie z kosztem energii elektrycznej). Do ataku została wykorzystana przeznaczona do tego platforma sprzętowa, złożona z 6 kart graficznych MSI GeForce GTX 1080 GAMING X [20], o sumarycznej wydajności 102,9 GH/s dla algorytmu MD5 [21].

Etap drugi — atak kombinacyjny (słowniki z regułami)

Jako drugi atak wykorzystano różne kombinacje imion i nazwisk, jak te przedstawione w tabeli 2. W niektórych przypadkach użyto bezpośrednio posiadane słowniki (na przykład dla wzorców takich jak [nazwisko][?s][?d]{0,4}@domena), a w innych należało przygotować dodatkowo połączone słowniki (dla wzorców takich jak [nazwisko][?s][imię][?d]{0,4}@domena), gdyż bezpośrednie użycie programu Hashcat z trzema oddzielnymi słownikami (imiona, nazwiska, nazwy domen) było niemożliwe.

Etap trzeci — e-maile z wycieków

W ostatnim podejściu wykorzystano bazę wycieków polskich adresów e-mail (jako źródło wykorzystano serwis exploit.in) i sprawdzono 7 milionów unikalnych rekordów. Dane pozyskane z takiego wycieku mają jedno ograniczenie: zawierają

tylko adresy e-mail w domenie .pl, bez innych popularnych domen międzynarodowych, takich jak gmail.com, yahoo.com czy outlook.com. Dla przeprowadzonego ataku 20,05% adresów e-mail (2794 z 13935) znalazło się w bazie wycieków, przy czym tylko 1,9% (265 e-maili) nie zostało wcześniej złamanych podczas pierwszego (siłowego) lub drugiego (kombinacyjnego) ataku. Warty podkreślenia jest tutaj fakt, że atak ten pozwala odzyskać około 20% skrótów MD5 Gravatar i powinien zostać wykonany na początku, przed atakami słownikowym i brutalnym.

Etap dodatkowy — wizualna deanonimizacja

Jako dodatkowy krok sprawdzono, ilu użytkowników korzysta z własnego obrazu na koncie Gravatar, tzn. nie używa domyślnych motywów graficznych udostępnianych przez usługę Gravatar. W celu sprawdzenia dla każdego wyodrębnionego adresu e-mail wywołano funkcję zaprezentowaną w kodzie 3.

```
function get_gravatar( $email ) {
        $url = "http://www.gravatar.com/avatar/".md5($email)."?d=404";
                $headers = @get_headers($url);
                if (!preg_match("|200|", $headers[0])) $has_valid_avatar = false;
                else $has_valid_avatar = true;
        return $has_valid_avatar;
}
```

Kod 3. Funkcja sprawdzająca istnienie indywidualnego pliku graficznego w serwisie Gravatar dla zadanego adresu e-mail

Aby pokazać, że niektórzy użytkownicy używają prawdziwych zdjęć, przygotowano obraz zbudowany z 27 losowo wybranych Gravatarów i przedstawiono go na rysunku 1.

Rys. 1. Losowo wybrane avatary użytkowników Gravatara

3.4. Wyniki ataku

Wybrane najciekawsze wyniki dotyczące przeprowadzonego ataku i odzyskanych adresów e-mail przedstawiają się następująco:

- 63,54% (8854 z 13935) wszystkich skrótów MD5 Gravatar zostało odwróconych do pierwotnych adresów e-mail;
- 8,7% e-maili ma ustawiony indywidualny obraz w serwisie Gravatar;
- blisko 10% adresów e-mail zawiera w sobie nazwę użytkownika;
- ponad 40% e-maili pochodzi z domeny gmail.com;
- prawie 80% e-maili pochodzi z pięciu najpopularniejszych domen.

Losowo wybierane skróty MD5 wraz z odpowiadającymi im adresami e-mail (przedstawionymi w postaci zanonimizowanej, na przykład p******d@wp.pl) są prezentowane dynamicznie na przeznaczonej do tego witrynie [22].

Można zadać pytanie, jakie adresy e-mail są ukryte za skrótami MD5 Gravatara, których nie udało się odzyskać? Do możliwych odpowiedzi zaliczyć można między innymi adresy e-mail pochodzące z: domen osobistych (np. rodwald.pl), domen regionalnych (np. krakow.pl), domen instytucji (np. amw.gdynia.pl) lub domen rządowych (np. mf.gov.pl), mało popularnych domen (np. t-net.com.pl), błędnie wpisanych domen (np. gmial.com), domen z innych krajów (np. mail.ru), czy też nieistniejących domen (np. hflgssa@cjttalf.oj).

Warto zauważyć, że w wielu krajach, w tym w Polsce, adres e-mail zaliczany jest do danych wrażliwych. Generalny Inspektor Danych Osobowych wydał opinię traktującą adresy e-mail jako dane osobowe w rozumieniu ustawy o ochronie danych osobowych. Wśród odzyskanych podczas ataku adresów e-mail istniały także takie, które zawierały w sobie zarówno imię, jak i nazwisko, czyli dane, które w pewnych przypadkach mogą jednoznacznie identyfikować właściciela.

4. Podsumowanie

W artykule przedstawiono praktyczny atak ujawniający adresy e-mail użytkowników wybranego serwisu internetowego umożliwiającego dodawanie własnych komentarzy. Skuteczność ataku (rzędu 63 procent) jest ściśle skorelowana z możliwościami (czasowymi, sprzętowymi i finansowymi) atakujących. Najważniejszym czynnikiem jest tutaj wydajność zastosowanego rozwiązania sprzętowego, jednak przy obecnej popularności platform wydobywających kryptowaluty (często z 6, 12, a nawet 18 kartami graficznymi) nie stanowi to już większego problemu.

Przedstawiony atak ma z jednej strony na celu zwiększenie świadomości użytkowników publikujących komentarze na różnych forach internetowych, w zakresie potencjalnego ujawnienia ich adresów e-mail, mimo że są przechowywane w postaci zakodowanej. Z drugiej strony natomiast powinien uczulić blogerów korzystających

bardzo chętnie z WordPressa, żeby ostrożnie podchodzili do umieszczania na swoich stronach systemów komentarzy korzystających z zewnętrznych serwisów, takich jak na przykład Gravatar, gdyż może to narażać ich czytelników na ujawnianie adresów e-mail, a czasami nawet na pełną identyfikację. Projektanci rozwiązań informatycznych powinni natomiast zaprzestać używania klasycznych kryptograficznych funkcji skrótu (takich jak MD5, SHA-1, a nawet SHA-2) w tych elementach systemów, gdzie skuteczny atak na przeciwobraz może mieć istotne znaczenie dla bezpieczeństwa (przechowywanie haseł, opisywany Gravatar) i zastępować je funkcjami adaptacyjnymi: pamięciowo lub obliczeniowo trudnymi (przykładowo: bcrypt, PBKDF2, ARGON2, Balloon).

Źródło finansowania badań – środki własne autora.

Artykuł wpłynął do redakcji 17.01.2019 r. Zweryfikowaną wersję po recenzjach otrzymano 23.05.2019 r.

Przemysław Rodwald https://orcid.org/0000-0003-4261-8688

BIBLIOGRAFIA

[1] abell, *Gravatars: why publishing your email's hash is not a good idea*, developer.it 2009, http://www.developer.it/post/gravatars-why-publishing-your-email-s-hash-is-not-a-good-idea [dostęp: 15.01.2019].

[2] Bongard D., *De-anonymizing Users of French Political Forums*, 0xcite LLC, Luxembourg, 2013, http://archive.hack.lu/2013/dbongard_hacklu_2013.pdf [dostęp: 15.01.2019].

[3] Web Technology Surveys, *Usage of content management systems for websites*, https://w3techs.com/technologies/overview/content_management/all [dostęp: 15.01.2019].

[4] Gravatar – strona startowa, http://pl.gravatar.com [dostęp: 15.01.2019].

[5] Rivest R., *The MD5 Message-Digest Algorithm RFC 1321*, 1992, https://tools.ietf.org/html/rfc1321 [dostęp: 15.01.2019].

[6] Wang X., Yu H., *How to Break MD5 and Other Hash Functions*, In: R. Cramer (eds), *Advances in Cryptology – EUROCRYPT 2005*, Lecture Notes in Computer Science, vol. 3494, Springer, Berlin, Heidelberg, DOI: https://doi.org/10.1007/11426639_2.

[7] Lenstra A., Wang X., de Weger B., *Colliding X.509 Certificates*, Cryptology ePrint Archive Report 2005/067, 2005, https://eprint.iacr.org/2005/067 [dostęp: 15.01.2019].

[8] Selinger P., *MD5 Collision Demo*, 2006, https://mathstat.dal.ca/~selinger/md5collision/ [dostęp: 15.01.2019].

[9] Stevens M., Lenstra A., de Weger B., *Predicting the winner of the 2008 US Presidential Elections using a Sony PlayStation 3*, 2007, https://www.win.tue.nl/hashclash/Nostradamus/ [dostęp: 15.01.2019].

[10] McHugh N., *Create your own MD5 collisions*, 2015, https://natmchugh.blogspot.com/2015/02/create-your-own-md5-collisions.html [dostęp: 15.01.2019].

[11] Rodwald P., Biernacik B., *Zabezpieczanie haseł w systemach informatycznych*, Biuletyn Wojskowej Akademii Technicznej, 67, 1, 2018, s. 73-92, DOI: 10.5604/01.3001.0011.8036.

[12] *Nvidia Gigabyte RTX 2080 TI Hashcat Benchmarks*, https://gist.github.com/codeandsec/1c1f2c7bd81abba6fa9736b061944675 [dostęp: 15.01.2019].

[13] Yao J., Mao W., *RFC 6531 – SMTP Extension for Internationalized Email*, IETF 2012, http://www.ietf.org/rfc/rfc6531.txt [dostęp: 15.01.2019].

[14] Mockapetris P., *RFC 1035 – Domain Names – Implementation and Specifications*, IEFT, 1987, http://www.ietf.org/rfc/rfc1035.txt [dostęp: 15.01.2019].

[15] Verisign, The Domain Name Industry Brief, vol. 16, iss. 1, March 2019, https://www.verisign.com/assets/domain-name-report-Q42018.pdf [dostęp: 22.05.2019].

[16] The Radicati Group, *Email Statistics Report 2018-2022*, 2018. www.radicati.com/wp/wp-content/uploads/2017/12/Email-Statistics-Report-2018-2022-Executive-Summary.pdf [dostęp: 22.05.2019].

[17] Dougherty C.R., *Vulnerability Note VU#836068 MD5 vulnerable to collision attacks*, Vulnerability notes database, CERT Carnegie Mellon University Software Engineering Institute, 2008, https://www.kb.cert.org/vuls/id/836068 [dostęp: 15.01.2019].

[18] Rodwald P., *Wybór strategii łamania hasła przy nałożonych ograniczeniach czasowych*, Biuletyn Wojskowej Akademii Technicznej, 2019, 68, 1, 2019, s. 79-100. DOI: 10.5604/01.3001.0013.1467.

[19] Heartle A., *Hasła ponad 10 milionów polskich kont email dostępne do pobrania w sieci*, zaufanatrzeciastrona.pl, 2017, https://zaufanatrzeciastrona.pl/post/hasla-ponad-10-milionow-polskich-kont-email-dostepne-do-pobrania-w-sieci/ [dostęp: 15.01.2019].

[20] *Hashkiller 1080 – specyfikacja sprzętu*, https://www.rodwald.pl/blog/432/ [dostęp: 15.01.2019].

[21] *Hashkiller 1080 benchmark*, https://www.rodwald.pl/blog/1161/ [dostęp: 15.01.2019].

[22] *Losowo wybierane adresy e-mail odtworzone z Gravatarów w serwisie jakoszczedzacpieniadze.pl*, https://www.rodwald.pl/blog/1195/ [dostęp: 15.01.2019].

P. RODWALD

E-mail recovery from websites using Gravatar

Abstract. Although the Internet has not been created for anonymity, many users posting comments on blogs hope that they will remain anonymous. The article presents an attack on revealing the e-mail addresses of users posting comments on the sample website using Gravatar. This attack aims to make readers aware of the fact that by posting comments on sites using Gravatara services, we are exposed to the disclosure of our e-mail addresses and sometimes even of our real identity.

Keywords: Gravatar, deanonimization, e-mail, MD5 hash function
DOI: 10.5604/01.3001.0013.3003