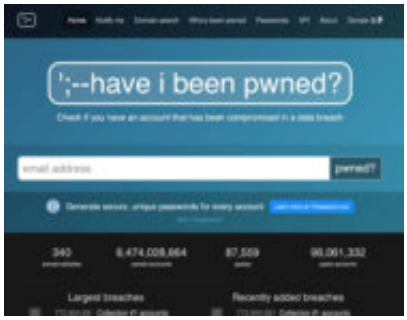
Collection#1 vs GDPR: Data Breach e riflessioni sulla Cyber Security

Author : Michelangelo De Bonis

Date : 4 Febbraio 2019



Come riportato dal noto ricercatore di sicurezza Troy Hunt sul suo blog [1], è recentemente apparso sul servizio di condivisione file Mega un archivio da oltre 87 GB contenente qualcosa come 2,7 miliardi di righe. Un forum specializzato in hacking dopo poche ore ne estrae il contenuto, lo riorganizza in directory per una più facile e veloce ricerca e fruizione. Troy Hunt continua così l'analisi dei dati e scopre che, una volta ripulito, l'archivio contiene oltre 700 milioni di indirizzi mail e oltre 21 milioni di password. Molte di queste informazioni sono frutto di vecchi *data breach* e, forse, sono incluse anche informazioni provenienti da campagne di hacking dirette a singoli individui. Siamo probabilmente davanti ad uno degli archivi più grandi di dati personali mai recuperati, così il ricercatore ne celebra la dimensione e l'importanza battezzando questa grande raccolta di dati Collection#1.

Possiamo, quindi, interrogarci e cercare di fare un'analisi critica di questi dati oramai a disposizione di tutti.

Collection#1: 140 milioni di email e 21 milioni di password

Il numero di indirizzi email contenuti in questo colossale archivio è sicuramente un numero elevato e strabiliante ma, onestamente, l'indirizzo email non è un elemento particolarmente sensibile. Basta pensare a come noi stessi "pubblicizziamo" giornalmente il nostro indirizzo email su biglietti da visita, nei siti professionali come LinkedIn o in altre mille modi. Esistono poi dei servizi online, come il sito www.voilanorbert.com, che permettono la ricerca di indirizzi mail di utenti utilizzando nome, cognome e dominio in cui ricercare, offrono API per l'integrazione del loro servizio in bot leciti o meno. Oppure servizi come www.verificaemail.com che, come mostrato in Figura 1, ricevono uno status OK direttamente dal server del gestore di posta elettronica per la validazione o meno dell'indirizzo stesso.

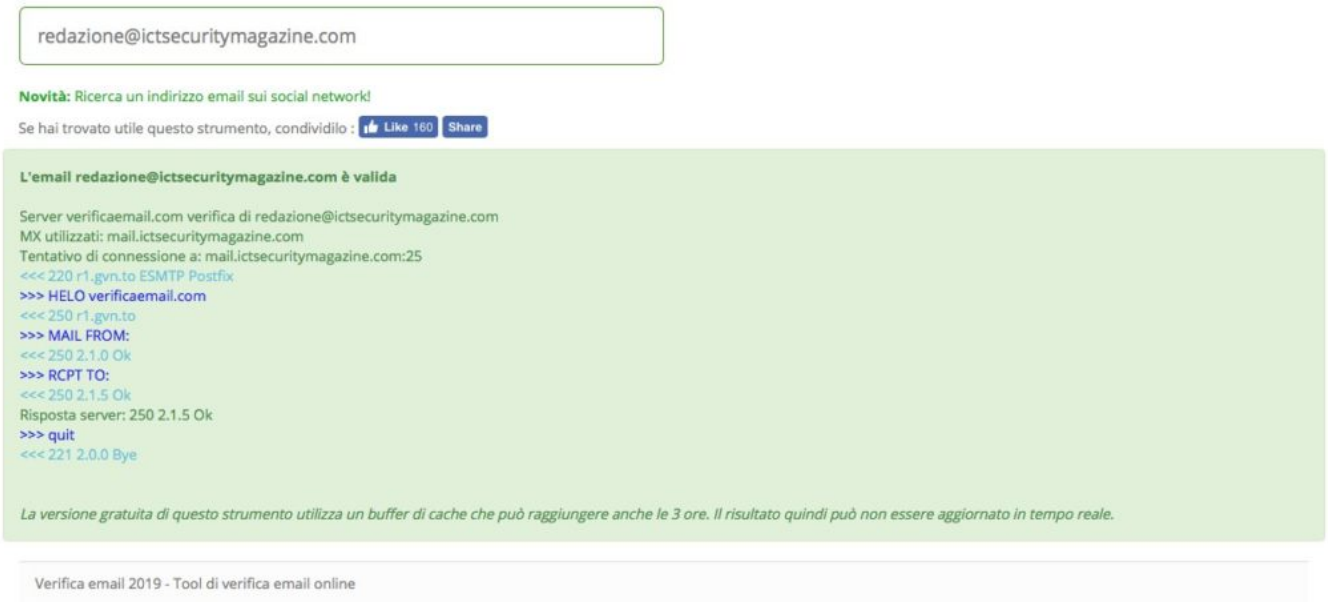


Figura 1 - Esempio www.verificaemail.com

Il ricercatore Troy Hunt offre un servizio sul suo sito **Have I Been Pwned** (HIBP)[2], per controllare se un particolare account sia mai stato interessato o coinvolto in una violazione dei dati. Quindi non c'è da sorprendersi se l'80% degli indirizzi di Collection#1 sono risultati già presenti nel database di HIBP magari perché già coinvolti in precedenti *data breach*. In questo archivio risultano "sconosciuti" *solo* 140 milioni di indirizzi email.

E l'analisi delle password invece? Degli oltre 21 milioni di password elencate, la metà è risultata nuova al sito HIBN. Analizzando questo numero, l'ottica del "bicchiere mezzo vuoto" ci porterebbe ad affermare che sono state rese pubbliche e violate una mole enorme di password (che dovrebbero essere assolutamente cambiate molto in fretta...). D'altro canto il bicchiere può essere anche visto come mezzo pieno: perché il rapporto numerico tra password e indirizzi email presenti in Collection#1 indica che una buona metà degli utenti a cui è stata rubata la password, non la usava già su altri siti hackerati precedentemente.

Il rischio più immediato è che queste credenziali siano usate per effettuare "crediantial stuffing" cioè l'uso della combinazione email/password su svariati siti nella speranza di poter accedere contando sul fatto che, di solito, l'utente riutilizza la stessa combinazione per comodità.

Non esiste una valida alternativa alle password, questo lo sappiamo, specialmente quando si tratta di proteggere la propria casella di posta elettronica, il conto bancario online e tutti i dispositivi che archiviano i nostri dati personali. Anche nei casi di autenticazioni a due fattori o tramite token generati al volo, c'è sempre una prima combinazione utente/password a "dare il La" alla sequenza. Inoltre, le evidenze portano alla luce che gli utenti non amano sistemi complessi di autenticazione (es. solo il 10% degli utenti Google usano il sistema a due fattori [3]) e che questi sistemi di ridondanze soffrono anche di criticità gravi [4]. Ecco perché è cruciale utilizzare delle combinazioni uniche e affidabili, e non solo per i siti e i servizi più

importanti. Il sito specializzato in sicurezza informatica Kaspersky offre uno strumento che permette di verificare e testare la robustezza di una password utilizzabile gratuitamente [5].

Altro strumento utilissimo che segnaliamo dal sito Kaspersky è, invece, di tenore metodologico: una metafora che ci aiuti a cambiare l'atteggiamento verso le nostre password. Gli esperti di sicurezza dell'azienda russa consigliano di pensare alle proprie password come se fossero biancheria intima [6]. Questo inusuale cambio di prospettiva da vita a tre regole fondamentali di gestione delle password:

1. non sono mostrate a chiunque;
2. sono cambiate regolarmente;
3. non sono esposte ben in vista sulla propria scrivania.

Se a questo saggio trittico di regole si aggiunge il buon senso di usare combinazioni lunghe, non direttamente derivabili dalla propria vita, con lettere maiuscole, minuscole, numeri e simboli, le password diventano sempre più affidabili e difficili da hackerare.

Il *data breach* non è questione di password

Il grosso problema dei *data breach* però è sostanzialmente un altro: i nostri dati sono esposti in modo totalmente indipendente dalla nostra accortezza e accuratezza. Infatti possiamo essere stati attenti ed efficaci nella scelta delle password, ma se queste sono rubate dal sito in cui sono depositate non possiamo farci nulla!

Tuttavia uno strumento utile a gestire il tema del *data breach* è stato recentemente introdotto in Europa. Il 25 maggio 2018 è entrato, infatti, in vigore in tutti gli Stati membri il Regolamento Ue 2016/679, noto come GDPR (General Data Protection Regulation)[7]. Tale regolamento è incentrato sulla protezione delle persone fisiche con riguardo al trattamento e alla libera circolazione dei dati personali.

Il regolamento europeo GDPR è uno strumento importante per veicolare una maggiore consapevolezza e cultura sulla cyber security nelle aziende di ogni dimensione. I meriti principali di una corretta applicazione del Regolamento sono:

- l'obbligo di definire nelle aziende un processo di gestione della sicurezza dei dati
- chiarire in modo inequivocabile a soggetti piccoli e grandi (dalla multinazionale alla PMI) che la sicurezza informatica è un tema non sacrificabile
- per "contagio" aumentare la consapevolezza degli utenti sull'importanza dei dati personali e sugli strumenti che si hanno per assicurarsi della loro corretta conservazione

Lo sforzo – e la sfida – per tutti dovrebbe essere volto a combattere la consunzione della norma e la tentazione di incasellare il messaggio culturale e di innovazione di processo della GDPR come l'ennesima vessazione burocratica. La GDPR è l'opportunità per introdurre una nuova forma di mentalità nei comportamenti e nei processi, trasformando il tema della sicurezza in un elemento strutturale, in un cambiamento culturale.

Riferimenti

- [1] https://www.troyhunt.com/the-773-million-record-collection-1-data-reach/
- [2] https://haveibeenpwned.com
- [3] USENIX Enigma 2018 - Anatomy of Account Takeover - https://youtu.be/W2a4fRalshI
- [4] https://hackernoon.com/why-do-most-people-ignore-two-factor-authentication-1bbc49671b8e
- [5] https://password.kaspersky.com/it/
- [6] https://www.kaspersky.it/blog/passwords-are-like-underwear/6959/
- [7] https://www.garanteprivacy.it/regolamentoue

Articolo a cura di **Michelangelo De Bonis** e **Matteo De Simone**