Mystery blurs dump of over 1 billion people's personal data

November 25 2019, by Nancy Cohen

Credit: CC0 Public Domain

Two security sleuths last month discovered an enormous amount of data that was left exposed on a server. Data found on the server belonged to around 1.2 billion people.

Kartikay Mehrotra wrote about it on Friday for Bloomberg, in a story, along with one from *Wired*, that was frequently quoted over the weekend. The data was left unprotected on a Google Cloud server.

The FBI were contacted and the server was shut down. Not trivial. *Wired* referred to the situation as a "jumbo" data leak. *Wired* said the information was sitting exposed and easily accessible on an unsecured server.

The data left unprotected was actually a database, aggregating 1.2 billion users' personal information, e.g., social media accounts, email addresses and phone numbers.

The incident was relayed on the Data Viper blog.

"On October 16, 2019 Bob Diachenko and Vinny Troia discovered a wide-open Elasticsearch server containing an unprecedented 4 billion user accounts spanning more than 4 terabytes of data. A total count of unique people across all data sets reached more than 1.2 billion people."

They were out to do just a routine scan for unprotected data and that is when the trove was spotted. The FBI were contacted.

Appearing in a Bloomberg interview, Troia, Data Viper founder, elaborated on the discovery. "To be honest this was just a part of our normal research process where we were just looking through open web servers to look for any databases that potentially have valuable information in them, and we just kind of came across it."

The 4 terabytes of personal information, about 1.2 billion records, did *not* include passwords, credit card numbers, or Social Security numbers, said Lily Hay Newman in *Wired*.

She spelled out what it did reveal. "It does, though, contain profiles of hundreds of millions of people that include home and cell phone numbers, associated social media profiles like Facebook, Twitter, LinkedIn, and Github, work histories seemingly scraped from LinkedIn, almost 50 million unique phone numbers, and 622 million unique email addresses."

Bloomberg quoted Troia. "There are no passwords related to this data, but having a new, fresh set of passwords isn't that exciting anymore. Having all of this social media stuff in one place is a useful weapon and investigative tool."

After all, just nabbing names, phone numbers and account URLs delivers ample information to get attackers started.

Harrison Van Riper, analyst at security firm Digital Shadows, made a similar point in *Wired*. "Van Riper notes that while passwords, credit card numbers, and government IDs are the most obviously threatening pieces of information for scammers to have, it's important not to underestimate the significance of all the supporting data that helps build out profiles of consumers."

Who owned the server? It is unclear how the records got there in the first place, said *Wired*. The data that Troia discovered seemed to be four data sets cobbled together. Welcome to the world of those who abuse "data enrichment."

Jacob J in International Business Times noted a "vastly unprotected and unregulated data enrichment business scene." The data sets appeared to originate from different data enrichment companies.

Cory Doctorow in Boing Boing also drew a blank: "No one knows who owns the Google Cloud drive that exposed 1.2 billion user records," he

wrote. Doctorow explained that data-brokers like People Data Labs and Oxydata "may have simply sold the data to a customer that performed the merge operation and then stuck the resulting files on an unprotected server."

"The owner of this server likely used one of our enrichment products, along with a number of other data-enrichment or licensing services," said Sean Thorne, cofounder of People Data Labs, in the *Wired* report.

So, whose trail does one pursue to figure out how the data was exposed in the first place? Experts said good luck with that.

"Identification of exposed/nameless servers is one of the most difficult parts of an investigation. In this case, all we can tell from the IP address...is that it is (or was) hosted with Google Cloud," said the Data Viper blog. "Because of obvious privacy concerns cloud providers will not share any information on their customers, making this a dead end."

Robert Prigge, President of Jumio, discussed the news with *Digital Journal*:

"We live in an era where information from disconnected data breaches, as well as legitimate data-selling companies, are often combined to create comprehensive identity profiles on the dark web, incorporating everything from personal identifiable information, to job history, to shopping preferences, to dating profiles, and more. The deep level of intel available is frightening, and it's making it extremely easy for criminals to commit digital identity fraud via a number of different ways."

Citation: Mystery blurs dump of over 1 billion people's personal data (2019, November 25)