EEI 5270 - Information Security

Case Study

In October 2023, 23andMe, a DNA testing company, acknowledged that its customers' profile data was compromised by malicious actors who employed a credential stuffing tactic.

Part One

1.1 Introduction

Concept and Inception

6.9 million users at risk and general public trust eroded in what was once a beacon of innovation. In light of recent events, it is important to understand what really happened, what circumstances allowed such an agregious opportunity and how we can contain or mitigate such risks going forward. Before we really explore into the crux of the matter, let's take a quick refresher on the context of this disaster.

Hailed as the "Invention of the year" by the Time magazine in 2008, 23andme was a pivotal entrepreneur in the consumer genetics industry. Born of Sillicon valley blood, to a father who served as a professor at Stanford university and a mother who taught some of the greatest technical minds to emerge in the 21st century, Anne Wojcicki took on the bet of providing direct-to-consumer genetics testing with the help of Linda Avey and Paul Cusenza. This bet paid off as the company hit the ground running and peaked at a market cap of almost 6 billion US dollars around 2021.

Role in the consumer genetics industry

The company was innovative as the then-cost of getting a DNA test was prohibitive for most consumers to reap the countless interesting discoveries and benefits of knowing one's DNA sequence can provide and and building a comprehensive genetical database was missed out on. The basic concept was simple. The idea was to give people access to their genetic information while collecting this information at the same time for research purposes. Consumers would consent and participate with the program by providing samples of saliva, which would then go through autosomal analysis (which itself was a pioneering technique soon to be adopted by all other similar companies soon after) to provide important genetic information like ancestral, health and similar information.

Since the completion of the Human Genome Project in 2003, wherein the entire human genome was sequenced, base pairs mapped from both a physical and functional standpoint, the cost of

personal genomics has gone down exponentially, thanks to the contributions of companies like 23andme to the industry. Thanks to these organisations, the cost of sequencing an individuals DNA has gone down to 1000 US dollars, and even 100 US dollars at the peak of the company's existence.

User growth

Starting in a small office, comprising of just the three founders, the company grew from humble beginnings to a peak of 816 employees and 14 million users. This meant that the company now possesses genomic data of those 14 million users including data on those who hadn't taken the DNA test as that could be extrapolated from the existing user data.

Sensitivity of the data

The essential challenge with genetic data is the information it may contain, including ethnicity and health-related information, together with the potential for reidentification (Sariyar, Suhr and Schlünder, 2017). This is because DNA is like a unique identifier, much like an identification number, but unlike other unique identifiers which can be regenerated and reassigned, this identifier cannot be changed and will identify the relevant person until death, i.e. irrevocable.

The misuse of such information in the form of targeted exploitation, discrimination, or other unforeseen consequences can not only affect the individual but also the biological relatives of the individuals regardless of whether the individuals are aware of each other's existence or the fact that they are related. This can compound the privacy concern as the related individuals may not even be aware of their data being collected as it can be inferred through sufficient samples. To put this in context, a carefully chosen set of 45 single nucleotide polymorphisms is more than enough to identify an individual to an error of 10^-15 (Pakstis et al., 2007).

While the protection of sensitive data, including genomic, is covered in legislation like the EU GDPR, the vast majority of consumers are unaware of the implications of data breaches and the services or products at offer by these companies.

Nature of the breach

In early October of 2023, 23andme officially filed a report with the Securities and Exchange Commission (SEC), USA, that they had identified that a large amount of data from individual user accounts had been compromised and scraped by a threat actor. This information was then sold online through the use of various of black-hat (hacking for nefarious purposes) forums.

Investigations into the event began and subsequently revealed that the attack had been the result of a "credential stuffing" technique employed by the threat actor. This was disclosed to the public, which resulted in an immediate loss of trust, and followed up with updates on the company's official blog as the investigation progressed.