EE1 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 aggregious opportunity and how one can contain or mitigate such risk going forward. Before really exploring 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 -

321 424374
No: _____ Date: __/__/____

Role in the consumer genetics industry

The company was innovative as the then-cost of getting a DNA tes was prohibitive for most consumers to reap the countless interesting discoveries and benefits of knowing one's DNA sequence can provide, and the collective advantages of maintaining a global genome database which would aid in the development of medicines for countless diseases with genetic linkers.
The basic concept was simple. The idea was to give people access to their genetic information while simultaneously collecting this information for research purposes. Consumers would consent and participate in 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 relevant information.

Since the completion of the Human Genome Project in 2003, wherin the entire human genome was sequenced and base pairs were mapped from both a physical and functional standpoint, the cost of personal genomics has gone down to 1000 US dollars, and even 100 US dollars at the peak of the company's existenc

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, ~~including~~ (Sariyar, Suhr and Schlunder, 2017). This is because DNA is like a unique identifier, much like an identification number, but unlike other unique identifiers which can be ~~recog~~ regenerated and reassigned, this identifier cannot be changed and will identify the relevant person until death. i.e. irrevocabl

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 others 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 suffricient 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 suffing" 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. It is important therefore, before continuing, to gain a strong grasp on the technique used in this attack and the environment that enabled it to take place.

Understanding Credential Stuffing

Credential stuffing is the term given to a cyber attack, where a threat actor takes a list of stolen usernames and passwords from a previous breach and uses it to access online user accounts. More often is the case that the obtained usernames and passwords are not paired and requires automation to try several pairs of credentials until a successful match is obtained; thus credential stuffing is a type of brute force attack.

Data from sources like SpyCloud indicate that more than 64% of users recycle credentials across online services, which makes these users more prone to successful credential stuffing attacks. Over 80% of the cyber attacks of this decade has been from credential stuffing (2024 Data Breach Investigations Report | Verizon) making this technique one of the most common prevention measures like password strength are ineffective to these types of attacks. Additionally, the threat vector is further elevated in risk potential by the fact that if one of the credentials is an organisational credential, then the attack is further disruptive.

321 424374
No:

Date: __/__/____

Some Recent Attacks.

- CBC Canada reported that nearly 10,000 login details for Canada's Revenue Agency were stolen in a credential stuffing attack. It was bad enough that the agency had to take their online services down temporarily.

- The FBI found out that sometime between January and August of 2020, hackers used a whole bunch of stolen logins to withdraw over $3.5 million from a US bank. Yikes!

- Another investmont firm in New York was hit by credential stuffing attacks for months, between 2019 and 2020. They got lucky and no money was actually stolen, but the hackers could have made off with almost $ 2 million.

- The PayPal breach in an early December attack which impacted 35,000 accounts. Although no money was stolen or illegaly transferred, the attack campaign's purpose, as speculated by the company in an official release statement, is to find other credentials that may be recycled for sale on the dark web to other threat actors.

How it works.

As previously stated, the concept is simple. Threat actors use stolen credentials from previous data breaches or purchased from the dark web. They use these credentials with automation on the presumption that people are very likely to reuse the same password across services. The major difference between this and a brute force attack is that in a traditional brute force attack, different combinations of characters are tested until the correct combination is found, while credential suffing involves finding the correct combination of usernames and passwords.

As the number of stolen credentials to check against is usually large and the time taken for brute force attacks scales with input, automation plays a crucial role in this form of cyber attack. Botnets are used extensively, which are networks of compromised computers to automate the task of testing combinations of usernames and passwords. Depending on the size of the botnet used, this allows the threat actor to quickly and reliably assess a large number of credentials over a brief period of time.

There are two phases in a credential stuffing attack:

- Validation - the phase where botnets are used to validate the correct combination of username and password.
- Exploitation - the phase where the correct combination of usernames and passwords obtained from the previous validation phase are used to authenticate and exploit the user account without the consent of the original user.

Understanding the two phases is extremely useful, as it provides two attack surfaces which can be seperately addressed to mitigate future risk of attacks.

321 424 374
No:

Date: __/__/____

Gaining Access

Compromised credentials can be accessed from a variety of different sources, including:

1. Data breaches: one of the more common ways that credential may be stolen.
2. Phishing attacks: a type of social engineering attack, where an attacker mimics the actions of a trusted entity using fake emails, fake login screens, etc., to take the credentials of the victim.
3. Credential leaks: from an organizational perspective, the security infrastructure could be flawed in that sensitive information is stored in plain text or using insecure processes. This could result in the unintentional or intentional leaking of the credentials.

Prevalence of the Technique

Data breaches are such a common occurrence that a recent study by Forbes reveals that there are 15 billion stolen credentials from 100,000 data breaches (Winder, Forbes) out in the wild that threat actors have a nearly endless source of logins to abuse for this type of attack. This leads to a sort of cyber security feedback loop where one data breach leads to downstream impacts on other apps and services on the web.

Value of Compromised Accounts.

Credential stuffing is also well established in the hacker economy. Rather than profit off the compromised account directly, money is made and risk of discovery is passed on to other threat actors at a more beneficial rate. A recent trend observed in the cybersecurity space, which this phenomenon is a direct example of, is the ongoing specialisation of threat actors. That is, some threat actors will specialise in gaining access while others specialize in using the compromised accounts in committing fraud.

These attacks have cemented their place in the hacker economy, such that accounts have a well-established commodity price. For example, financial accounts like PayPal may go for $30 to $120 depending on the amount of money in the account already. (Dark Web Price Index 2021 - Dark Web Prices of Personal Data, 2021).

Financial accounts are not the only compromised accounts for sale, social media accounts are also available for sale. These accounts may be used to spread malware or for phishing campaigns as well as astroturfing campaigns, where a marketing or public relations campaign is conducted under the guise of unsolicited comments by the general public.

Bad Cyber Hygiene or How Threat Actors Gained Access.

The vulnerability wasn't just the existence of the compromised accounts, but also a wide range of security complacencies from 23andme which allowed the two step process of credential stuffing to take place. Despite what the company claims in the defense built by their lawyers, there was significant bad cyber hygiene which lead to the attackers accessing a large data set instead of being failed at the very beginning or limited to a smaller set of accounts once the pattern was identified.

Until the breach occurred, the login process had no rate limiting, ip block lists, and account login timeouts. These measures are used from the server side to prevent the threat actor from abusing automation to make many requests or attempts at trying the credential combinations. Ideally, the service should detect abnormal login rates at high frequencies and should try to limit or block the traffic coming from the specific duration to deincentivize repeated login attempts.

Lack of multi-factor athentication mechanisms also proved vital in enabling this sort of attack. The lack of initiative on the company's part to alert the users on potential risks of recycling logins also contributed. One of the major contributing factors was the unawareness of the regular user ~~account~~ about such risks and dangers of using the products and services offered by the company.

Timeline of the Attack.

- In August 2023, the threat actor known as Golem on Breach Forums, claimed to have stolen 300 TB of 23andme data.
- On the 10th of October 2023, the company confirmed that their data had indeed been compromised. Stolen data emerged that verified that the idea indeed matched the certain users' public geneology.

Data Exfiltration

Despite the company claiming that the attacker was only able to affect 0.1% of the userbase, because the attacker was able to exploit certain features like Ancestral tree and DNA relatives, the actual number of users affected by the data scrape totalled around 6-7 million (Different sources provide differing figures).

Discovery and Actions Taken

This opt-in feature, as the site once described it, allows the users to share their information with other users on the platform to find distant genetic relatives. This information includes broad descriptions of the users' genetic make-up but with no raw data.

Exposed Data

According to the listing found on the dark web, the data included:

- Origin estimation
- Phenotype (the observable traits of an individual like height, eye color, etc.)
- Health information
- Name and Photograph
- Identification data.

The hacker further claimed that they had organized the data points by ethnicity and origin estimation. The data profile price points range from $1000 to $100,000 (Alder, 2023). The attackers seemed to have primarily targeted individuals of Ashkenazi Jews and Chinese descent. The kind of data that was leaked, like the name, photograph, etc left the attack with a highly politicised undertone. That is to say that experts believed the attack to be racially and politically motivated.

1.2. Impact

1.2.1. Immediate Consequences.

Customer Concerns

Especially in the information industry, these types of attacks highlight a key risk that organizations must address. Sensitive consumer data, when exposed without consent, will lead to serious violations of privacy and the erosion of public goodwill and trust. Consumers may feel vulnerable that their locations, identification, et may be in nefarious hands, especially minorities facing the danger of persecution. Furthermore, this kind of data breach results in feeling like the users have lost control of their own data

Immediate impacts on the consumer's wellbeing may be due to employers, governments, insurance companies, or others that may use the information in an unfair manner, often to refuse services or jobs based on certain information that must be private. Additionally, this informatio could lead to blackmail, identity theft, persecution and further social engineering to commit fraud.

The biological relatives of the compromised users, who may not have even used the 23andme service, may now face the same threats as the users themselves. Their data and information which was derived involuntarily may also be compromised. There have been cases, where families of political members or individuals wanted by governments have been hunted down using genomic information; thus endangering whole demographics.

Reputational Damage

Regardless of whether 23andme feel that they have done no wrong and that the attack wasn't their fault, the public scrutiny has severly tarnished the reputation of the company. Once hailed as the invention of the century by the Times magazine, the erosion of public trust has weighed heavily on the company.

The company stock once valued at $6 billion at market cap, is now at the danger of being delisted from NASDAQ as the company is being hit with multiple lawsuits from the 7 million users affected from the data breach (23andMe Blames Users for Recent Data Breach as It's Hit with Dozens of Lawsuits | WIRED)

1.2.2. Long-term Consequences

Heightened User Caution

Due to the ensuing events of the data breach, even months later, users are cautious about approaching the company. Critics have argued, beyond the specific legal complaints, that there has been a lack of accountability and transparency from the company.

It has also brought to light, some of the ~~unfi~~ vulnerabilities that telehealth (health services over electronic means) companies face and the information that users should be aware of. Up until the breach, there was a massive interest in Direct-to-Consumer DNA testing from companies like 23andme and Ancestry. After events such as this, customer churn and weariness will slow the growth of the industry.

Regulatory Scrutiny

This breach and the subsequent lawsuits have brought the spotlight to the broader consumer data protection laws. Consequently, multiple states in the United States now have consumer protection laws, some even with DNA specific clauses included. This incident highlights the ever growing need for more stringent data protection laws and fines to hold companies accountable.

Competitive Pressures

2023 was the year, where the most interest was shown in the use of DNA testing. Both 23andme and Ancestry.com grew in userbase amassing a vast amount of DNA data. After the data breach, use of 23andme has decreased and users have started to diversify in their reliance on companies to test their DNA. Other companies capitalizing on the same opportunities that 23 andme focus on, have had an easier time approaching the market while 23 andme has been tied up in legal and regulatory troubles. (23andme Reports FY2023 Fourth Quarter and Full Year Financial Results | 23andMe, Inc.)

1.3. Analysis

Economic Implications

After the initial discovery in October 2023, 23andme spent considerable resources in IT, legal and public affairs, to contain the situation. The costs of halting regular operations to send notifications and messages to all customers, especially those affected.

Additionally, third party forensics and government ~~regulation~~ regulators were brought in for investigation. These force regular operations to take the backseat and more money to be spent on these third parties to conduct the investigations. Additionally, as news of the event spreads, class action lawsuits by their customers were filed against the company. Legal matters such as this is also extremely expensive and the company funds took a large hit as evidenced by the annual financial report.

Possibly other costs may be incurrated for providing identity theft insurance to those affected consumers. Revenue will also be lost from these products and services due to customers leaving the company en masse and reduced patient trust. Investor confidence in the company will also reduce, leaving it to be stranded without financial support.

Companies like GSK (formerly GlaxoSmithklein) depend on the data that 23andme and other similar companies collect for research and development of drug cures for genetically linked diseases. Breaches lead to disruption of healthcare and delays in patient care as the information that other pharmaceutical companies depend on may be tied up in legal trouble.

Productivity of employees will be lost, due to incident response, recovery and containment efforts. The stigma that follows from public scrutiny may also lead to an uptick in employee burnout and turnover. Especially those engineers in charge of security and authentication will be under strain from investigative efforts. Thus, the damaged reputation will further hinder the attraction of future investors and cash inflow, as well as the hiring of talent due to the lack of cash to maintain high labor costs and the social stigma surrounding the company.

Humanitarian Implications

The information should, but doesn't to a satisfying degree, embrace the principles of transparency. Many telehealth, direct-to-consumer genetic testing companies don't disclose the unique tradeoffs that consumers make when signing up for these services. Consumers often spend a premium on these services and are then surprised to find that the company not only profits from the provided service but also the mined data.

Additionally, when data is not stored securely and internal access control mechanisms are not provided, unauthorized access may be given to threat actors to sensitive medical decords, including diagnoses, treatment plans, and genetic information.

Especially minorities like Ashkenazi Jews and persecuted citizens are at risk from unauthorized access to health information. Ultimately, the intangible efforts of psychological duress to these patients whose data was leaked can be immeasurable. The potential consequences of such can be detrimental to even those unaffected by the breach. This may lead to patients voluntarily withholding potentially life-saving information from

medical professionals delaying much needed patient care.

Social Implications.

Once patients lose trust in the organization, they are ~~not~~ more inclined to refer their medical information to third parties which may or may not be varified. This can very easily lead to misdiagnosis and potentially harmful treatments or avoidance of treatments altogether.

Less technically literate users who are unaware of cyber risks and how to setup measures to counter such risks are more likely to be exploited in cyber attacks. These individuals come from demographics of society without access to good education and good IT awareness training. Breaches and attacks such as this, lead to furthering the disparity and inequality in healthcare.

Part Two.

2.1. Recommendations.

Criticality of Strong Authentication Mechanisms.

Strong authentication processes are necessary for a secure exchange of information between the user and the organisation. Mandatory uniqueness and complexity requirements as well as the use of multi-factor authentication mechanisms are also important to create a secure environment.

Multi-Factor Authentification

Multi factor authentification is a form of authentication which requires multiple "factors" to authenticate. This basically means that a registered device or email is required to verify the identity of the person using the account credentials.

The basic authentication flow goes like this:

1. User enters credentials
2. User receives one-time-passwords via email, sms or authenticato - apps
3. Entering the top verifies the identity of the user.

Where solutions like rate limiting (limiting the amount of traffic flowing through the network so that requests cannot be made several times in a row) and ip block lists may not be as effective as the attacker may spread the attack over regional botnets (Credential Stuffing Prevention - OWASP Cheat Sheet Series), multi factor authentification stumps password based attacks. This authentication mechanism requires the user to have the secondary authentication device in person, invalidating brute force automation.

The downside ~~of~~ is that multifactor authentication is counter intuitive to setup, impacts UX and non-techsavvy users may forego this process entirely. Convenience is usually the reciprocal of Security.

Strong Password Policies

Updating the password requirement for both new and existing users will be extremely helpful so that unique passwords with strong lengths (to prevent character or dictionary brute force attacks) and mixed character combinations will hold up better against these types of attacks.

Uniqueness can also be mandated using password management tools to generate the password for you. This alleviates the ~~long~~ cognitive load of having to remember the password. The main reason for password recycling is that it is difficult to remember many unique passwords for many different services. These tools help manage your passwords such that you don't have to remember more than one.

Data Access Controls

Regular reviews of users and user activities to discover malicious patterns early. Embrace the principle of least privilege, that is limit the access to data for features like DNA relatives to allow only the information that the user needs to do their job or get the service they requested.

Robust Network Segmentation

This means to partition systems and networks such that an exploit in one system cannot affect another. Redundancy measures are also recommended to reduce the impact such attacks may have on the systems.

Sector-wide Collaboration

Organisations are fighting lone battles against threat actors in the health sector. United, they may fare better. It is important for health sector organisations to share threat intelligence and lessons learned with each other so that as a whole, the health sector is secure.

Technical Safeguards.

It is difficult to correctly verify that user login activity / traffic is friendly and not nefarious. This makes credential stuffing attacks difficult to monitor.
However, systems like intrusion detection and prevention systems can still help look for early signs of a potential attack.
Account ~~lokout~~ lockout mechanisms may also help but these are still susceptible to botnets spread across regions so that account lockout code is never triggered.
Using existing information and shared threat intelligence it may be possible to identify regions or ip address ranges from which botnets commonly attack from. This is errorprone, but firewalls can be used to thus lock out entire ip address ranges.

2.2. Strategic Approach: Awareness Training Investment.

Targeted Education and Reinforcement.

The biggest complaint from critics in the 23andme disaster, is the lack of transparency. The company could have done more to inform users about the potential risks, what they were signing up for and the measures have been implemented from the user side.

From an organisational perspective, staff must be trained in cyber security and preventive measures must be stressed. To prevent organisational credentials from being compromised, it is vital for staff to undergo training on phishing attempts, password hygiene and to avoid password recycling across accounts. Organisations may even consider opting for physical key devices for multi-factor authentificd

Safety training at airports happen regularly, every few months. This highlights the importance of reinforcement of safety procedures in life critical environments. Cybersecurity is no different, it is techniques employed by threat actors.

Businesses often are reluctant to include security training as they see them as cost centers and hence non-profitable. Often times, it is the opposite case, where 80% of employees after a strong cybersecurity training programs ~~setup~~ conducted regularly show a significant decrease in likelihood of falling for a phising scam or social engineering attempt (2020 State of the Phish: An in-depth look at user awareness, vulnerability and resilience; 2021). Additionally, such training programs set up convenient procedures for users and employees to report suspicious behavior when spotting it.

No: Date: / /

Key Lessons (Q1.3) and Conclusion.

23andme suffered a devastating data breach in October, 2023 where attackers ~~would~~ used the "Credential Stuffing" technique to scrape the data of upto 7 million users, critically violating the privacy of all these users. The lessons learned upon the completion of the investigation are the vital importance of strong authentication practices, especially multi-factor authentication (MFA), the dangers of credential recycling and the importance of awareness training. The unique sensitivity of genetic data which exposes both the direct owner as well as the related individuals and the politicized nature of the exploit complicates the situation for the healthcare industry. The need has been identified, for a series of strong, united, defensive measures for the healthcare sector to fight off the inevitable horde of threat actors focused on exploiting the vulnerable for money. It is no longer enough, that companies are reactive to the attacks from hackers; that is rather ~~wait~~ to implement MFA after an attack, better to take a proactive approach & important to review the level of trust placed on passwords and to take on more accountability and responsibility for user health information as it is clearly shown to be a highly profitable industry. The hope going forward is ~~that information as this~~ such needless disasters are prevented in the future by following security best practices, adopting zero-trust policies and staying up to date, in an ever-changing threat landscape.