The Yahoo Data Breach of 2013-2016

Cody Murrer

**Abstract**

This paper will serve as a case study and analyzation of data breaches. However, it will be focused specifically on the Yahoo Data Breach that had spanned the years of 2013 all the way to 2016. We will analyze how attackers managed to breach Yahoo’s servers and security counter measures to access sensitive data. We will also look at the scale of the breach, and the response in which Yahoo had toward the public as well as internally to ensure the breach was patched. The result of this paper is to possibly show how preventable this breach could have been or if there was nothing else Yahoo could have done to prevent the situation from occurring.

**Introduction**

Data breaches have plagued the internet since it had begun growing in popularity after its release for public use. This is caused by companies and corporations moving operations online and having the need to store different kinds of user data for their members to retain some sort of user records for their accounts or business. Because companies begun storing user information malicious attackers have decided to attempt to steal that data and sell it online for large profits. Due to attackers trying to steal data the efforts to counter attacks and bolster security to prevent the theft of data has been ever evolving from the internet’s conception up until present day.

This paper will talk about one of the biggest data breaches that the internet has ever had. This data breach had taken place over multiple years but is known to be one whole data breach since both breaches were announced on the same day years after each one had occurred. There have been many data breaches, but one of which known as the Yahoo data breach was called the biggest breach of user data in the history of the internet. We will analyze every aspect of the data breach and how Yahoo had handled the breach when notified of such a large-scale attack taking place years prior to their knowledge.

**Background**

Yahoo has been a notorious company and a staple in the history of the internet. The company was founded in 1994 and was one of the most visited websites worldwide in the year 2000. Yahoo hosts a large variety of services like email, search engines, finance information, entertainment, news, and many other services. With these services Yahoo allows the creation of user accounts to use a service like their email. Accounts made with Yahoo around the time of the breaches all included personal information containing your full name (first and last middle optional) as well as a password, security question with its answer, and lastly an optional backup email address for if you get locked out of your account. With all these items of personal information Yahoo is storing onto their servers they must maintain a high level of security as at the time there were over 4 billion accounts worth of user information being stored.

Although a high level of security is a must, Yahoo’s countermeasures were quite lacking and their urgency to acknowledge the issue before all their user’s data was leaked was also very slow and handled not in any way one would expect such a large company like Yahoo to do. While this was going on Yahoo was during settling a deal with the company known as Verizon to sell to, but only then did they begin to look at their data breaches as to show that they were able to sell the company and all its data with it without any form of issues. During the sale Yahoo had announced that it had been the victim of multiple data breaches between the years 2013 to 2016 with the most notable ones being from 2013 and 2014. These breaches had affected every single user account and stole all personal data pertaining toward those accounts.

In the September of 2016 Yahoo had announced that it was the victim of a data breach, but the severity was yet to be known at the time. At first, they had estimated the breach had occurred during 2014 but after more research into the topic the breach had in fact happened during 2013, 2014, and leading into 2015 before it had gotten noticed by the large company. The main breaches were the ones from 2013 and 2014. In 2013 the breach had gone unnoticed for a long duration until the sale of data had been found on the dark web and law enforcement begun to notify the company. Similarly in 2014 the data was also being sold online but this time was suspected to be sold by foreign state sponsored actors since the data being sold had pertained information for accounts on government officials like congressmen and senators. The 2014 leak was instead of being found on the dark web by police it was found on the dark web by a cyber security researched Andrew Komarov who had then notified Yahoo of their findings online.

**How The Breach Occurred**

The Data breach had taken place over many years from 2013 to 2016. There were many ways each breach had taken place however the most notable ones were from 2013 and 2014. The 2013 breach had occurred when cross site scripting known as XSS was found to be able to penetrate Yahoo’s webpage and run malicious scripts from one client’s browser that could affect others that were connected to the webpage at that time. On top of this sort of attack constantly taking place on Yahoo’s webpage their email list was leaked and had been targeted with phishing links in which one had successfully managed to trick an employee. Because of this they were able to gain access to things like the source code of the website which was where the beginning of the 2013 attack had taken place. With the use of the source code forged user cookies were created and sent out to almost all users using the Yahoo account servers which then gave access to every single account that logged in to those attackers without the need of password access.

The 2014 breach was done in a very similar way except they had already had the forged cookies on users’ browsers so all they had to do was abuse the cross-site scripting that was still being used to attack users on Yahoo. Combining the forged cookies and cross site scripting attacks had allowed the foreign hackers to eventually gain access to the server in which that held all user data for the entirety of Yahoo’s database. At this point they were able to steal every single person’s data that had ever used an account on Yahoo as well as steal personal information on high-ranking government officials from the United States of America. Which had forced Yahoo to do a full-scale investigation on the leaked data to ensure they it was entirely their user data and not faked data that could have been created to be passed off as their own.

**The Scale**

There is a reason as to why this breach is known as one of the internet’s biggest data breaches and that’s because it had affected every single user that was using the platform during the attacks. This means that it had managed to steal the data of every single account that was active from the start of the breach in 2013 all the way up to when Yahoo alerted the public in 2016. Because of this every single user account that Yahoo managed and had stored data on was breached and was up for sale on the dark web. The scale of which was upward to 4 billion affected users.

Each user’s personal information they had inputted during their account creation was leaked. This means every username, email, password, country in which they lived in, phone number, security question with answer, as well as their back up email in case they forgot their password.Text

Description automatically generated This is a big security issue because many people often use the same passwords for many different accounts which causes this security breach to also lead to other possible accounts being taken on different sites. In figure one it shows a small sample of what the data had looked like as a plaintext document where it listed account information alphabetically using the accounts username or email when it also shows the persons date of birth, country of origin, and their hashed password. Normally only being able to see the hashed password is fine, but Yahoo had later told the public that it was transitioning from the MD5 hashing algorithm for their passwords to the bycrypt hashing algorithm for their passwords. This meant that Yahoo had alerted those who were in possession of these hashed password the different kinds of hashing algorithms they must use in order to attempt to crack or brute force in order to obtain the users passwords. Which was a flaw in their security by telling the attackers the different types of hashing algorithms as they were actively still trying to fix the vulnerability.

Figure Part of the data dump that had been leaked

**Yahoo’s Response**

The response in which Yahoo had for the public about the breach was instead of alerting every user about the breach they had not notified the public for years despite knowing about the different breaches. Before alerting the public Yahoo had decided to go through all the data dumps that was for sale to ensure its authenticity, but when they had reviewed all the stolen data and learned that it was authentic, they had alerted the public as soon as possible. Their response was one of the most lacking responses in terms of giving the public reassurance in their security measures. One of the ways they had alerted the public was they had directed the affected users to a blog within their website about the different security notices instead of creating a public statement directed toward the affected users to tell them everything they are doing to help resolve the issues taking place. Figure two is a screenshot in one of the few webpages that Yahoo had directed users toward when they needed answers to their questions.Timeline

Description automatically generated

Figure help page that Yahoo had redirected users to when they had questions

Other than redirecting questions their users were having to a webpage with prewritten answers they had also taken steps to invalidate all passwords. By doing this they had made all the data that the attackers had collected during the breach null and void as they were no longer able to access these accounts since their passwords were invalidated. On top of invalidating the user passwords causing every user to resubmit a new password to use their accounts and have access to their emails they had also invalidated the user’s security questions and answers. This had made it harder for accounts that were still breached after the password invalidation to recover their account unless they had gone through the backup email in order to reset the account security questions and password.

**How It Could’ve Been Prevented**

This could have been easily prevented from the start with Yahoo giving better training to employees for attacks like malware and phishing sent through email. With this they could have prevented their source code from being leaked and prevented forged cookies from being sent out to users causing part of the data breach from happening. Another prevention that could have been used was not store sensitive user data all in one place. This had allowed the attackers to come in and take all user data with little issue because they didn’t have to search through multiple databases for all the user data. They also had used the MD5 hashing algorithm to hash all their passwords in the beginning which is a very unsecure hashing method where with enough computing power anyone can brute force it if they get their hands on a good enough dictionary and algorithm.

**Conclusion**

The Yahoo data breach was one of the worst data breaches the internet had ever seen. This was the outcome of poor security and a terrible sense of urgency in trying to resolve the breach that had fell into their lap. The company had shown they it was ill-equipped for a data breach let alone one to this scale and it had showed through their response to the public and how they addressed the issues when it continued to get increasingly more severe.

**References**

[**https://yahoodatabreachsettlement.com/**](https://yahoodatabreachsettlement.com/)

[**https://www.ncsc.gov.uk/news/yahoo-data-breach-ncsc-response**](https://www.ncsc.gov.uk/news/yahoo-data-breach-ncsc-response)

[**https://www.cshub.com/attacks/articles/incident-of-the-week-multiple-yahoo-data-breaches-across-4-years-result-in-a-1175-million-settlement**](https://www.cshub.com/attacks/articles/incident-of-the-week-multiple-yahoo-data-breaches-across-4-years-result-in-a-1175-million-settlement)

[**https://money.cnn.com/2017/10/03/technology/business/yahoo-breach-3-billion-accounts/index.html**](https://money.cnn.com/2017/10/03/technology/business/yahoo-breach-3-billion-accounts/index.html)

[**https://www.npr.org/sections/thetwo-way/2017/10/03/555016024/every-yahoo-account-that-existed-in-mid-2013-was-likely-hacked**](https://www.npr.org/sections/thetwo-way/2017/10/03/555016024/every-yahoo-account-that-existed-in-mid-2013-was-likely-hacked)

[**https://www.nytimes.com/2017/10/03/technology/yahoo-hack-3-billion-users.html**](https://www.nytimes.com/2017/10/03/technology/yahoo-hack-3-billion-users.html)

[**https://www.sec.gov/news/press-release/2018-71**](https://www.sec.gov/news/press-release/2018-71)

[**https://help.yahoo.com/kb/SLN27925.html**](https://help.yahoo.com/kb/SLN27925.html)

[**https://bpbonline.com/blogs/news/yahoo-data-breach-what-actually-happened**](https://bpbonline.com/blogs/news/yahoo-data-breach-what-actually-happened)

[**https://asamborski.github.io/cs558\_s17\_blog/2017/04/02/yahoo.html**](https://asamborski.github.io/cs558_s17_blog/2017/04/02/yahoo.html)