Organizational and Societal Impact of the 2013 and 2014 Yahoo Data Breaches

Ivan Adinata[1], Boxiang Fu[2], Dominic Henley[3] and Lingshan Zhang[4]

[1] *The University of Melbourne, Student ID 1163497*
[2] *The University of Melbourne, Student ID 998092*
[3] *The University of Melbourne, Student ID 1186484*
[4] *The University of Melbourne, Student ID 1136442*

Information Security and Privacy (INFO30006)

Team DIWA (Group Number 34)

Date: 23/09/2022

Abstract—This paper analyzes the 2013 and 2014 Yahoo data breaches and its implications on the organization and society in general. It was found that the breach resulted in the significant destruction of shareholder wealth and reputation. Additionally, the attack posed significant risks to the privacy of its 3 billion users, as well as to over 150,000 highly sensitive government and military employees. Yahoo's negligent breach response is then compared to Github's efficient response to analyze their differences. It was found that comprehensive internal controls with active anomaly detection systems in addition to a transparent and forthcoming communication channel is what made Github's response efficient and what Yahoo was lacking. Several recommendations were then put forward encouraging organizations to be more transparent and work more closely with their security team. In this way, future cyber-attack threats could hopefully be diminished or mitigated and its blast radius reduced.

Keywords—Yahoo data breach, Github DDoS, loss of personal information, negligence, transparency

CONTENTS

1. Introduction

The 2013 and 2014 Yahoo data breaches were the largest leak of private information in recorded history [Perlroth, 2017]. In total, over 3 billion users' private information were compromised [Richter, 2017]. This has led to severe negative organizational and societal effects, ranging from decreases in shareholder value to the leak of highly classified government and military employee profiles [May, 2016].

The response from Yahoo's data breach is then compared with the response from the 2018 Github DDoS attack. The distinct trade-offs the decision makers had to make in each incident reflected the effectiveness of their data breach policies. Github's response was swift and effective, while Yahoo's response was slow and inadequate. The contrasting responses are then compared to evaluate what results in an effective response and what should be avoided.

In the final sections of this report, the timely transparency and disclosure of the data breach is discussed. In particular, it examines how Yahoo's lackluster and opaque response ultimately led to its downfall and destruction of shareholder value. Lastly, a list of recommendations and "best practice guides" is provided to help mitigate similar events in the future.

2. Background

Two separate attacks on Yahoo's data servers occurred in 2013 and 2014. In total, it was estimated that the attacks led to the compromise of over 3 billion and 500 million user accounts respectively [Perlroth, 2017]. Affected accounts in Yahoo's application ecosystem include but are not limited to: Yahoo Mail, Yahoo Finance, Yahoo Fantasy Sports, and Flickr [Perlroth, 2017].

It was found that the data breach resulted in the compromise of users' personal information such as their passwords, names, email addresses, dates of birth, security questions, and their answers. Luckily, the systems housing payment information and bank accounts were not affected [National Cyber Security Centre, 2017].

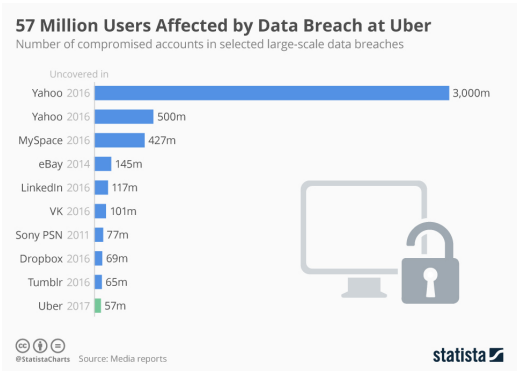

Fig. 1: Number of compromised accounts in major data breaches [Richter, 2017]

Later investigations found that the hackers gained entry into Yahoo's internal systems by spear-phishing employees with root access. They then exploited additional system vulnerabilities to obtain a backup version of servers housing the user-bases' private information [Ng, 2017].

The data breaches remained unknown to the public for about 2 years until leaked Yahoo account details started selling on the Dark Web in 2016 [Tsukayama et al., 2016]. Even to this day, the two breaches still tops the list of the most number of compromised accounts in a data breach incident (see Fig. 1).

3. Organizational Impact

The discovery and subsequent widespread media coverage of the data breach led to substantial negative sentiment towards Yahoo. In the short term, Yahoo's share prices experienced a drop in market value. However, perhaps more importantly, Yahoo was in the middle of being acquired by Verizon when the data breach was discovered in 2016. The widespread negative media coverage of the compromise eventually led a decrease in Verizon's buyout price of Yahoo by $350 million to $4.48 billion [Fiegerman, 2017].

Almost immediately, Yahoo was sued by its affected user-base and investigated by the United States Congress. In 2020, the case was settled in court and Yahoo was ordered to compensate $117.5 million to those affected by the breach [United States San Jose District Court, 2020]. In addition to this, the United States Securities and Exchange Commission also fined Yahoo for a

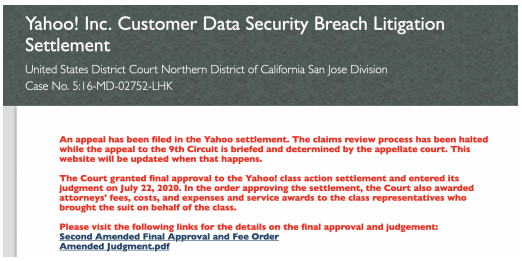

Fig. 2: Yahoo litigation settlement website

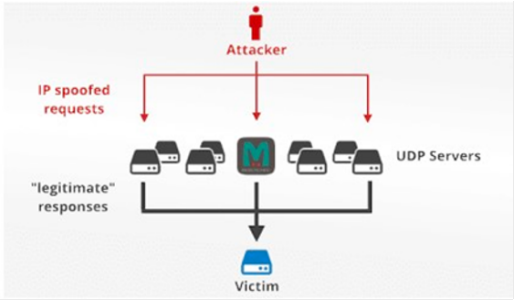

Fig. 3: Memcached amplification mechanism [Kottler, 2018]

further $35 million for its lack of disclosure and negligence in handling the data breach (see Fig. 2) [Securities and Exchange Commission, 2018].

In the long term, Yahoo suffered irreparable damage to its reputation and public image. It also raised doubt on the future cooperation between its business partners. For example, Verizon executives were originally divided as to whether or not to continue with their acquisition of Yahoo [Samuel and Kalluvila, 2017]. In the end, the remediation fees and reputation damage significantly decreased the wealth of the organization and its shareholders.

4. Societal Impact

In the short term, the compromise of over 3 billion individual user accounts poses significant risks to privacy and personal information. The names, email addresses, and dates of birth of over 3 billion users worldwide can now be readily found on the Dark Web. This in turn increases the risk of specific individuals being targeted my malicious entities using the leaked data.

In addition, the data breach also leaked the identities of over 150,000 highly sensitive government and military employees [May, 2016]. Not only would this possess a national security risk, it also endangers the lives of the leaked individuals and their families.

However, despite the negative impacts on society, some potential benefits can be attributed to the data breach. It can be argued that the widespread media coverage of the hack raised public awareness on the impact of cyber-crime on privacy intrusion. This may be a leading factor for more widespread implementation of cyber-security safeguards such as two-factor authentication and stronger database security measures.

5. Github DDoS

I. Background

On February 28th 2018, Github was targeted by a distributed denial-of-service (DDoS) attack intended to disrupt the normal traffic flow of its servers. As a result, it was unavailable from 17:21 to 17:26 UTC [Kottler, 2018]. At that time, it was the largest DDoS attack in recorded history.

The attackers abused Memcached instances that were inadvertently exposed to the public over the internet. Memcached is a distributed memory-caching system originally intended to store external data in RAM to speed up dynamic web applications [Kottler, 2018]. However, the hackers used this as an attack vector by making it an amplification tool (see Fig. 3). Using this, the attackers amplified the attack by 51,000 times, meaning that 51 Kilobytes of data is sent to Github's server for every 1 byte sent by the attacker [Kottler, 2018].

During the peak of the attack, 168.75 Gigabytes of data was sent to Github's servers every second [Kottler, 2018]. Amazingly, Github was able to mitigate the attack in under 10 minutes. The anomalous traffic was first detected at 17:21 UTC. At 17:26 UTC, a directive was given to redirect traffic to a different server. Finally, at 17:30 UTC, the system was reported to have recovered and fully online [Kottler, 2018]. An impressive feat considering that DDoS attacks generally lasts hours if not days.

II. Comparison

Comparing the Github DDoS attack with the Yahoo data breach, some facts are immediately obvious as to why Github was able to so effectively mitigate the threat while Yahoo was not. Perhaps the most important aspect is that the internal controls at Github was comprehensive and effective. Once an anomaly was detected, decision making was instant, and the malicious traffic was routed elsewhere without delay [Kottler, 2018].

Github should also be commended in its use of active anomaly detection systems. Github was able to detect the anomalous traffic in real time, thus enabling them to implement countermeasures within minutes of the attack. In contrast, Yahoo did not or was only using a rudimentary active detection system. As such, they were not able to detect the intrusion as it was occurring, and only realized they were hacked after the damages had been done [Securities and Exchange Commission, 2018].

There was also a stark contrast between how Github and Yahoo dealt with their user-base post-breach. Github was transparent and honest, issuing a public statement the day after the attack detailing what had happened and reassured users that data was not compromised [Kottler, 2018]. On the other hand, Yahoo intentionally concealed information in regards to the breach, and only publicly acknowledged the breach two years after the hack [Securities and Exchange Commission, 2018].

Finally, after the attack, Github publicly announced its commitment to further strengthening their network and improve their automatic intrusion detection system [Kottler, 2018]. They also announced that they will be spending money to expand their edge networks so that DDoS attacks can be mitigated earlier and faster [Kottler, 2018]. Contrast this with Yahoo. Yahoo adamantly denied any wrongdoing despite the overwhelming evidence, and remained opaque and uncommunicative with their user-base, preferring to solve issues in Court instead [Securities and Exchange Commission, 2018, United States San Jose District Court, 2020].

III. Analysis

When analyzing the comparisons in hindsight, it is obvious that Github's approach did a much better job at mitigating most of the damages of the attack when compared to Yahoo. The swift response to redirect all incoming traffic made by Github during the DDoS attack meant that there were almost no lasting impact on the company. On the other hand, Yahoo's slow response time meant that the hackers were able to fully extract the user-bases' private information before their security team became aware. As such, much more detrimental organizational and societal side effects were evident, ultimately leading to huge fines and the leak of highly confidential government information.

The difference between Yahoo's lackluster post-breach actions compared with Github's transparent communications also played a crucial factor on the long term impact on both companies. Github's approach meant that the company did not suffer any reputational damage, whereas Yahoo's long term reputation was irreversibly damaged and viewed as untrustworthy and opaque. This would have played a crucial factor in the number of lawsuits and investigations the companies faced in the aftermath of both attacks – Github received none, while Yahoo was flooded with lawsuits and fines.

6. TRANSPARENCY (LACK THEREOF)

Yahoo's disclosure and transparency of the data breach can be called lackluster at best. Their disclosure of the two data breaches was over two years late. They were only compelled to acknowledge the breach after private user account details started selling on the Dark Web [Tsukayama et al., 2016].

What was even more shocking was that Yahoo's annual reports contained materially misleading information. They claimed that the company "only faced the risk of potential future data breaches", and stipulated that they were not aware of any previous "security breaches" [Securities and Exchange Commission, 2018].

However, later investigations made by the Security and Exchange Commission found that

Yahoo was well aware of the data breaches almost immediately after the attacks, and concluded that Yahoo was intentionally concealing the data breach and deliberately misleading both investigators and stockholders [Securities and Exchange Commission, 2018].

The investigation also found that various internal reports were produced and circulated by the Chief Information Security Officer to a number of senior management personnel within days after the breach asking for resources to further investigate the extent of the compromise. However, the management and legal team had decided that it not "warrant substantial further inquiry" – despite the fact that the original internal report indicated that the personal data for at least 108 million users were stolen [Securities and Exchange Commission, 2018].

As a consequence of this negligence, Yahoo faced several lawsuits from Congress and affected parties. It decreased Verizon's buyout price of Yahoo by $350 million to $4.48 billion [Fiegerman, 2017]. Furthermore, it was fined $35 million by the SEC for their intentional concealment of the data breach, and settled in court for another $117.5 million to those affected by the breach [United States San Jose District Court, 2020].

7. Mitigation and Prevention

Much of Yahoo's shortcomings in dealing with the data breach can be mitigated and prevented. They are broadly classified into two categories – pre-breach and post-breach. These are the measures Yahoo should have reasonably taken so as to protect itself from potential threats and respond effectively once a breach has occurred.

I. Pre-Breach

One of the most important measures for protection against cyber-security threats is to listen to your Chief Information Security Officer (CISO). As discussed earlier, the executive team essentially ignored their CISO's internal reports and recommendations [Securities and Exchange Commission, 2018].
There were also reports indicating the executive team sidelining the CISO and disregarding the protection of their user-bases' data [Securities and Exchange Commission, 2018].

Fig. 4: Previous CISO of Yahoo Alex Stamos [Stamos, 2022]

This uncooperative work environment has led the CISO position to change hands 3 times in 6 months [Gagliordi, 2015]. It is pretty self-evident that something has to have gone wrong when the CISO position changes so often. Even so, such frequent shuffles in the CISO position can prove to be detrimental to the company's cyber-security measures. Each incoming CISO has to learn the most important attack vectors faced by the company. This comes with experience working with the company, it is not something one could learn in a mere 6 months.

As previously mentioned, it was found that the hackers gained entry into Yahoo's internal systems by spear-phishing employees with root access (see Fig. 5) [Ng, 2017]. Therefore, a segmented network could have been implemented beforehand so as to prevent lateral movement within the internal network. Even if the hackers gained access to one part of the network, they would have had trouble traversing to the other parts and would have minimized the blast radius of the breach.

Furthermore, Yahoo could have also implemented a zero-trust architecture for their network. The architecture would have required all users to verify and authenticate on a continuous basis and limited users to a least privileged access mechanism (see Fig. 6) [Dreyfus, 2022]. Therefore, the hackers would have a much harder

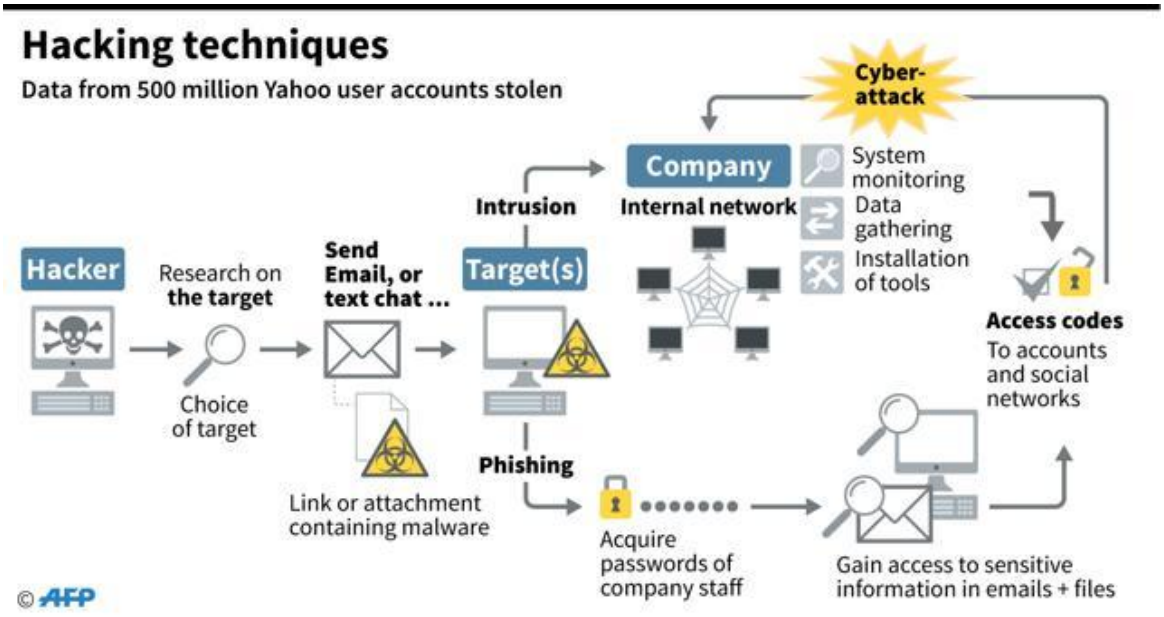

Fig. 5: How hackers gained access to Yahoo's network [Jackson, 2016]

time getting into the system as their entry into the network would most likely be flagged as an anomaly. Even if they somehow obtained access to the network through a phished employee account, they would not be able to extract useful information as the employee's account access would be limited to their own work area, and not to the database holding confidential user information.

Finally, as eluded to earlier, the hackers obtained a backup version of the database containing user information [Sat, 2020]. Therefore, don't make the backups available to be accessed via the internet. It is relatively low-cost for Yahoo to remove any physical connections that the backup database has that can be connected to the internet, and only retain connections to the local network within Yahoo's premises.

II. Post-Breach

Even if a data breach has occurred, Yahoo could have still implemented post-breach controls to minimize the blast radius of the compromise. One such control is to track the data breach and find the extent of the damages. Clearly Yahoo's executive team didn't do this as they ignored their CISO's internal reports asking for resources for further investigation [Securities and Exchange Commission, 2018].
This gross negligence by the executive team only brought upon further litigation and regulatory costs down the road.

Furthermore, it should have been standard practice for Yahoo to openly communicate with its user-base in such events. Once the breaches has been identified, it should have been company policy to issue press releases indicating of such an event. The company should also require every user to reset their passwords and encourage the use of multi-factor authentication. Such policies do not cost much to implement, while offers substantial organizational and societal benefits in return.

8. Recommendations and Limitations

The previous few sections detailed Yahoo's shortcomings and what "not to do". Similarly, its comparison with Github detailed a much more effective way of handling the situation. The bullet points below entails a "best practice guide" that should be implemented to prevent or at least reduce the risk of future data breaches and cyber-attacks. The list is itemized from top to bottom by what the authors believe to be the most important and cost effective solutions.

- Timely disclosure of data breaches and honesty with your user-base
- Have a less toxic work environment where the CISO's recommendations are respected and

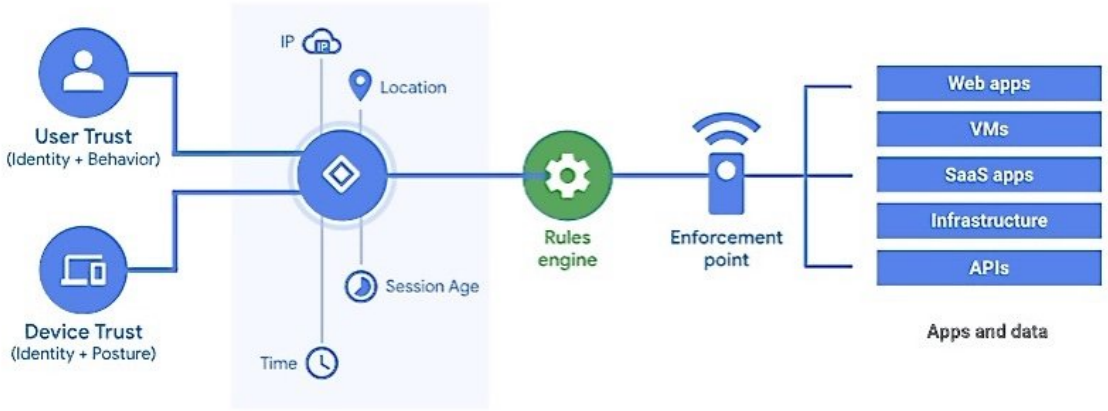

Fig. 6: Zero trust architecture overview [Liderman, 2020]

not ignored

- Keeping backups of important and sensitive information offline and accessible only through the internal network
- Employee training to reduce social engineering and phishing risks
- Enable multi-factor authentication and other low-cost and low-hassle points of identification for both users and employees
- Implementing a segmented network
- Implementing a zero-trust architecture

However, the above recommendations are not without its limitations. Some obvious hurdles preventing the implementation of the above recommendations are monetary and time costs. For example, Yahoo was in a period of decline when the data breaches occurred. It would be improbable that they would have had the necessary financial capabilities to be able to overhaul their entire system and adopt a segmented or zero-trust network.

Similarly, Yahoo would have had to spend considerable time and effort in training employees on spotting phishing attempts and social engineering tricks. It is also doubtful that the organization's culture and work environment could be changed in the short term without a complete reshuffle of the management team. These impose practical limitations on the efficacy of the above recommendations despite their potential benefits on paper.

That being said, it does not relieve Yahoo from their duty of care for their user-base's privacy. The first and third recommendation is surely doable without a substantial commitment in financial costs nor time. Yahoo's inability to implement such cost effective controls is what characterizes its failure in handling the data breach compared to the much more successful and effective controls implemented in Github's case.

9. Conclusion

The 2013 and 2014 Yahoo data breaches in total affected over 3 billion users. The lackluster and negligent response made by Yahoo resulted in the destruction of shareholder wealth and the reputation of the company. Furthermore, it caused widespread societal effects by revealing the identities of over 150,000 highly sensitive government and military employees.

Comparing this with Github's 2018 DDoS attack response, it was evident what had failed in Yahoo's case. Github's response was swift and efficient, characterized by a highly sophisticated active anomaly detection system and structured internal controls. Github was also highly transparent with its user-base and promised to further strengthen its cyber-security.

On the other hand, Yahoo was opaque and negligent, claiming to investigators and shareholders intentionally misleading information. As a result, Yahoo was fined and faced lawsuits from both the government and affected users. Such problems were non-existent for Github due to their efficient

and transparent response.

To mitigate and decrease the blast radius of future data breaches, several recommendations were given. The few most important include being honest with your user-base, respecting and implementing the CISO's recommendations, and keeping backups of sensitive information offline. Hopefully with these recommendations in place, future cyber-attacks would be not be as damaging and destructive to the organization and society as the one faced by Yahoo in 2013 and 2014.

REFERENCES

[Dreyfus, 2022] Dreyfus, S. (2022). Information security and privacy lecture 3a: Emerging trends in cybersecurity.

[Fiegerman, 2017] Fiegerman, S. (2017). Verizon cuts yahoo deal price by $350 million. Available at https://money.cnn.com/2017/02/21/technology/yahoo-verizon-deal/index.html.

[Gagliordi, 2015] Gagliordi, N. (2015). Yahoo hires new ciso, now the third exec in role in 6 months. Available at https://www.zdnet.com/article/yahoo-hires-new-ciso-now-the-third-exec-in-role-in-6-months/.

[Jackson, 2016] Jackson, G. (2016). Russia? china? who hacked yahoo, and why? Available at https://phys.org/news/2016-09-russia-china-hacked-yahoo.html.

[Kottler, 2018] Kottler, S. (2018). February 28th ddos incident report. Available at https://github.blog/2018-03-01-ddos-incident-report/.

[Liderman, 2020] Liderman, E. (2020). How android enterprise supports a zero trust security model. Available at https://blog.google/products/android-enterprise/zero-trust/.

[May, 2016] May, P. (2016). How a super cybersleuth helped crack the huge yahoo hack. Available at https://www.mercurynews.com/2016/12/15/how-a-super-cyber-sleuth-helped-crack-the-latest-yahoo-hack/.

[National Cyber Security Centre, 2017] National Cyber Security Centre (2017). Yahoo data breach: Ncsc response. Available at https://www.ncsc.gov.uk/news/yahoo-data-breach-ncsc-response.

[Ng, 2017] Ng, A. (2017). Fixing yahoo cybersecurity when they're 'really out to get you'. Available at https://www.cnet.com/news/privacy/yahoo-cybersecurity-bob-lord-what-its-like-putting-out-the-fires/.

[Perlroth, 2017] Perlroth, N. (2017). All 3 billion yahoo accounts were affected by 2013 attack. *The New York Times*. Available at https://www.nytimes.com/2017/10/03/technology/yahoo-hack-3-billion-users.html.

[Richter, 2017] Richter, F. (2017). 57 million users affected by data breach at uber. Available at https://www.statista.com/chart/5983/data-breaches/.

[Samuel and Kalluvila, 2017] Samuel, M. and Kalluvila, S. (2017). Verizon executive says company unsure about yahoo deal. Available at https://www.reuters.com/article/us-yahoo-m-a-verizon-idUSKBN14P215.

[Sat, 2020] Sat, G. (2020). 3 mega-breaches and how they could have been prevented. Available at https://medium.com/@sat_g/3-mega-breaches-and-how-they-could-have-been-prevented-c35f29873b3e.

[Securities and Exchange Commission, 2018] Securities and Exchange Commission (2018). Yahoo cease-and-desist proceedings.

[Stamos, 2022] Stamos, A. (2022). Alex stamos. Available at https://cisac.fsi.stanford.edu/people/alex-stamos.

[Tsukayama et al., 2016] Tsukayama, H., Timberg, C., and Fung, B. (2016). Yahoo data breach casts 'cloud' over verizon deal. Available at https://www.washingtonpost.com/news/the-switch/wp/2016/09/22/report-yahoo-to-confirm-data-breach-affecting-hundreds-of-millions-of-accounts/.

[United States San Jose District Court, 2020] United States San Jose District Court (2020). Yahoo! inc. customer data security breach litigation settlement. Available at https://yahoodatabreachsettlement.com/.