The Security of Amazon Web Services

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree Bachelor of Science, School of Engineering

**** ******

Spring 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor Joshua Earle, Department of Engineering and Society

Introduction

Cloud computing is everywhere in our daily lives. Services like Apple's iCloud allow for data to be backed up, Amazon Web Services (AWS) commercials come on during football games, and many large corporations use the cloud every day. Most people have heard the term, and just accept that something called "the cloud" can magically back up their photos or give them more storage. In fact, most people do not really think about what the cloud is, how it works, or if there are any potential downsides to using it (McKendrick, 2012).

The industry of cloud computing is dominated by a few large companies including Amazon, Google, and Microsoft (Cohen, 2021). When a company hosts their website or application in AWS, that just means they are using Amazon's servers instead of their own. Since many businesses and people use cloud services, the security of the data of many is in the hands of just a few. Although each individual AWS user is still responsible for some of their own security, all users would be compromised if Amazon has a major security breach. I did not have much knowledge of this topic until working with the cloud firsthand, and most people still do not (McKendrick, 2012). For clarification, cloud computing is the on-demand renting of server space and services from large datacenters.

In this research paper I investigate the security concerns of AWS and how most people perceive these concerns and AWS. I also apply Actor-Network Theory (ANT) to the security of AWS and its impact on the relevant stakeholders in the methods section. Next, I use literature review to figure out what Amazon says about their security practices and to find out what people know about cloud computing. Furthermore, I include two case studies about major security incidents involving AWS to analyze who is at fault for the breach and what impact it had on those affected through ANT in the analysis section. Finally, in the discussion section I decipher

whether Amazon's dominance in cloud computing is good socially and for security, and what developments I would like to see in the future.

Methods and Frameworks

Applying actor-network theory to my findings as it should help clarify the relationships in this research paper. These "actors" include not just the people, but the technologies and companies involved and how they socially impact one another (Latour, 1992). The actors in this network include Amazon and its place in the cloud market, AWS itself, AWS security principles, attackers, companies that use AWS, those companies' employees that directly and indirectly work with AWS, the customers of the companies that use AWS, and the other cloud competitors. Due to the importance of the nonhuman actors in this network, ANT is the logical choice for my analysis. As Sismondo explains "Representing both human and non-human actors, and treating them in the same relational terms, is one way of prompting full analyses, analyses that do not discriminate against any part of the ecologies of scientific facts" (2010).

Additionally, case studies about AWS and their security practices will be useful to see how other companies view AWS and are affected by its security. The first case study is the leaking of internal information from Capital One in 2019 (Khan and coauthors, 2022). Seeing what caused this leak, how Amazon responded, what the legal ramifications were, and what kinds of data were exposed should provide valuable insight. I will also be doing a similar case study on the Deep Root Analytics data breach of 2017 to gain further info on how AWS users fall victim to these incidents (UpGuard, 2017). I also compare it to the Capital One breach. These studies combined show how Amazon's responses to both incidents have changed over time and if any new measures were or were not enacted to help companies prevent incidents like

these. Applying ANT to my findings should help clarify the relationships in this network. Lastly, I conducted a literature review to find out about Amazon's role in the cloud computing space, how AWS's leadership views security, laws on cloud computing, and more information on AWS data breaches.

Results

Companies must be able to trust the reliability of renting server space and services from Amazon, so they can switch over to the cloud. AWS makes up a little over 30% of the global cloud infrastructure, which makes Amazon the cloud market leader. (Cohen, 2021). This may not seem like a massive market share, but it is nearly 4 times bigger than third place Google. Additionally, over 1.45 million different businesses use AWS in some capacity (Team Intricately, 2022). These businesses have employees that all either work directly with or rely on the cloud, along with even more customers all dependent on AWS. Massive corporations with millions of users such as Airbnb, Moderna, Netflix, State Farm, Starbucks, and Verizon all use AWS (Michalowski, 2023). These and other companies all have lots of sensitive data to store, both for internal operations and personal data of customers. With so many companies and their customers relying upon one large service, it is vital for it to be secure.

However, security is considered a strength of AWS by its leadership. In 2022, AWS CEO Adam Selipsky stated that they had the best cloud security, and that it was vital for companies looking to transition away from traditional servers. Furthermore, he elaborates on how they offer several services to help with security, and how third-party software can be used too (Kapko, 2022). Within AWS, there are many products that can aid with keeping users' safe, like providing

event logs, monitoring threats, safely storing, and encrypting passwords, and much more. Clearly Selipsky believes enough in their security efforts to publicly claim they are the best.

The official AWS website contains more information about the key principles and methods behind their security, like the Shared Responsibility Model. This model clarifies what things Amazon is responsible for securing and what things AWS users and developers are responsible for securing. Amazon must protect the hardware, the networks, and the software that run all of Amazon's prebuilt services, while their customers are supposed to properly use those services to guard sensitive information, ensure proper authentication and authorization, and apps built on AWS (Amazon, n.d.). This means that any data breaches could be caused by a weakness in the underlying infrastructure, or by poor implementation of defense mechanisms by an AWS user.

With millions of users, businesses, and customers of those businesses all relying on AWS, Amazon has a lot of information stored. According to themselves, Amazon complies with numerous international data privacy laws as well as many independent third-party certificates and standards (Amazon, n.d.). In theory, third parties should ensure all personal data stays private, and avoid potential spying or unwanted disclosures. In my research I have been unable to find any privacy violations by AWS, but Amazon at large has several such controversies. The tech giant had to settle lawsuits from the FTC over spying concerns with its Ring and Alexa products. Although Amazon denies breaking any laws, these examples show they are not perfect with regards to privacy (Quiggle, 2023).

Despite the prominence of cloud computing in the business and technology worlds and people's daily lives, the evidence suggests the general public does not really know very much about this field. According to a study from 2012, "Only 16% said they think of a computer

network to store, access and share data from Internet-connected devices [for the cloud]", and "95% [of survey respondents] actually already use cloud in some form" (McKendrick). This study is very revealing, as even when people were using the cloud, they had no idea what it was or that they used it. If people do not fully understand cloud computing, they do not understand what AWS is or how it works. Even though this study is a little dated, there is still a lack of clarity of what the cloud even is. Michael Gibbs, who has worked with cloud technologies since the 1990s, states "…we are seeing an explosion in organizations migrating to cloud computing, building private clouds, and connecting to public clouds. But there's extreme confusion with regards to cloud computing. What is it? How does it work?" (2022). Gibbs points out how different companies use unique terminology for the same technologies in order to confuse people. It is hard to the average person to keep up with most technologies and what they do, let alone an industry that purposely uses elaborate names for its most basic components.

In spite of all these cybersecurity efforts, sometimes applications built on AWS still fall victim to cybercrime. Due to AWS's market dominance, many large and important organizations like to use it. This means that when data breaches occur, massive amounts of private files are leaked or stolen. As of late September 2023, there are 14 known major data breaches that have occurred on AWS starting in 2017 (Heiligenstein). These breaches have combined to impact over 350 million people. The data lost includes names, credit card information, drivers' licenses, social security numbers, bank account numbers, phone numbers, and more (Heiligenstein, 2023). Although the companies are certainly negatively affected by being hacked, it is usually their customers and users that are hurt the most. There are not much these people could have done differently to avoid having their personal information leaked besides being customers of different businesses, yet now they have to worry about identity theft among other things.

Two of these large data leaks deserve special attention. Initially I examine the first known major incident, the Deep Root Analytics breach. This one is the largest attack and has a lot in common with most of the later AWS security failures, making it a good base case. Next, I analyze the Capital One cyberattack, as it was arguably the worst AWS breach so far in terms of outcome and cause.

Case Study 1

Deep Root Analytics is a company that helped the Republican National Convention in the 2016 presidential election. In June of 2017, information on about 198 million voters was available to download by anyone (Heiligenstein, 2023). The exposed material included "names, dates of birth, home addresses, phone numbers, and voter registration details, as well as data described as 'modeled' voter ethnicities and religions" on each person (UpGuard, 2017). As of 2016, there were approximately 200 million registered voters, meaning approximately 99% of them were in this dataset. All of this was unprotected to the public for two weeks, and it is currently unknown whether anyone stole any data before the vulnerability was discovered (Heiligenstein, 2023).

So how was so much sensitive information left available so easily? Amazon offers a service called Simple Storage Service (S3), which allows for entire objects to be stored in one of the countless AWS servers withing a region. These objects can be anything, including csv files, images, pdfs and more. As long as a business is willing to pay, practically any amount of data can be uploaded to the cloud. These S3 buckets have several security options, and it is up to each business to ensure the correct one is applied. In the Shared Responsibility Model, it is Amazon's job to make sure each security configuration works properly, and each user's job to use the right

one. Deep Root Analytics mistakenly was set to public, meaning “anyone with an internet connection” could have found and stolen from this S3 bucket by “navigating to a six-character Amazon subdomain: ‘dra-dw’” (UpGuard, 2017). In short, the difference between exposing 198 million peoples’ information and keeping it secure was simply not hitting the private button in a settings window. Unfortunately, this is a common occurrence as at least eight of the other thirteen major AWS breaches were caused by issues with S3 bucket configurations.

Case Study 2

The Capital One data breach was one of the worst known data breaches. The financial services giant had “Social Security Numbers, bank account numbers, credit card transaction records, credit scores, and more” stolen from over 100 million customers in 2019 (Heiligenstein, 2023). Perhaps most concerningly of all, the attack was carried out by a former AWS employee named Paige Thompson, who was convicted in 2022 for her crimes (Heiligenstein, 2023). Having such important financial information stolen is deeply concerning and warrants further discussion.

Capital One was an early adopter of AWS and made a big deal of switching to the cloud. This was part of a successful effort to switch to becoming more of a technology company. In spite of this push, they ultimately fell victim to relatively simple and “well-understood vulnerabilities” (Khan and coauthors, 2022). Thompson was able to exploit a misconfiguration in the Web Application Firewall (WAF), which was supposed to protect Capital One’s infrastructure from unwanted outside influence. The WAF was not made by Amazon but instead was from a third party (Novaes Neto and coauthors, 2020). The exact method used to get passed the WAF is still not clear, but it was likely one of few common attack types. Regardless of technique, the attacker

was then able to find temporary credentials attached to an AWS service known as EC2 that was helping to host the WAF. The specifics of EC2 are not important, but these credentials had access to far more than they should have and allowed Thompson to access a private S3 bucket containing the data (Khan and coauthors, 2022). To what extent her knowledge from her time working for Amazon helped Thompson remains unclear, but it is hard to imagine it did not help in some way. It is easy to blame the data breach Capital One for misconfiguring the WAF and having a credential with too much access. But Amazon still bears some responsibility. They could have built in better protections for the WAF attack, as the attack type was a well-known technique (Khan et al., 2022).

Analysis

The main actors are Amazon, AWS and its status in the cloud market, the other cloud competitors, AWS security principles, attackers, companies that use AWS, those companies' employees that directly and indirectly work with AWS, and the customers of the companies that use AWS. The most obvious connection is between Amazon and AWS. The technology giant owns AWS, which is and has been their most important profit generator. According to reports, AWS made "$62 billion in revenue in 2021 and $18.5 billion in net profits" in 2021 and "$80 billion in revenue in 2022 and close to $23 billion in operating profits" in 2022 (Michalowski, 2023). The cloud is still a growing part of Amazon's business, and it is very important that it stays that way. Likewise, even though AWS has its own CEO, its decisions and strategies are controlled by its parent company, and its primary goal is to make Amazon as much profit as possible.

Amazon's position in the cloud market and its competitors is also a key component of this network. The battle for supremacy against Microsoft, Google, and Alibaba likely has

impacted AWS's security. Amazon's cloud offering is still the market leader at 32% market share, but Microsoft's Azure at 20% market share, Google's cloud at 9% market share, and Alibaba cloud at 6% market share all would love to become the premier cloud provider (Cohen, 2021). Amazon is doing everything it can to keep its place atop this valuable market. The AWS CEO bragging about their security relative to their competitors makes more sense in this context (Kapko, 2022). Seeming like the saftest cloud platform regardless of the number of data breaches that have occurred is a good way to stay ahead in this arms race.

Also, this struggle to remain the dominant market leader incentivizes Amazon to add as many new services to AWS as possible. Although AWS at its core is a cloud computing company that can host websites, databases, and applications, it has countless additional services of variable nicheness. The cloud leader has focused on adding as many new capabilities as possible: "When AWS began offering its cloud infrastructure platform in 2006, it only released a handful of services; but each year, the pace of releasing new services and features has increased exponentially" (Khan et al., 2022). This strategy has definitely been very successful at making AWS the most used cloud platform, but it is not the best for ensuring secure applications. Risk evaluator and cybersecurity worker Stephen Harris, a man with over 30 years of experience in his field, states that "I really am not a fan of the AWS security model; there's far far too many knobs and controls, and it's not clear how the interact with each other. It can be hard to even know something simple … because of how configurations interact" (2019). The push for more and more new offerings has led to AWS having confusing and complicated security. The company's top priority is making as much money for its parent, which requires staying ahead of its competitors in the cloud marketplace. Even though having a great cyber-defense is important, pumping out as many shiny new features as possible to increase revenue is the number one goal.

The connections between Amazon, AWS, the cloud market, and competitors have a great effect on the companies and employees who use these web services. As of 2022, 1.45 million businesses use AWS (Team Intricately). Almost all of them will need to rely on the cloud platform for some amount of security needs. Additionally, most of these corporations will likely need more than one service, even more so if they are going to or already have fully transitioned to the cloud. The more services used, the more complicated it becomes to secure the entire system. Employees at companies who use AWS that do not directly work with it can be easily enticed by new technologies being available that could save time and money, which puts a lot of stress on those directly working with AWS to make everything work properly, let alone work properly while following best security practices.

The data breaches covered in my two case studies this analysis. The Deep Root Analytics exposure was caused by an unprotected S3 bucket (UpGuard, 2017). S3 buckets are one of the most used AWS offerings, yet they are the most common issue in data breaches (Michalowski, 2023). There is plenty of documentation on the best security principles on how to secure these buckets. In spite of this, nine of the fourteen major AWS leaks had misconfigured S3 settings. I suspect that S3 itself is not the issue, but the bloat of systems around it. S3 literally has the word simple it its acronym, it seems very unlikely that it is impossible to set up correctly. Rather the constant addition of new services and products has put so much burden on cybersecurity teams and developers that it becomes extremely difficult to keep track of everything, especially the most basic of practices. Furthermore, the Shared Responsibility Model makes it even easier for AWS to add complications without taking on blame. Under this methodology, Amazon is responsible for making sure each of their components functions properly and safely, while it is their customers to ensure the safety of their whole system working together (Amazon, n.d.).

When a company fails to properly use their services in harmony, Amazon can just put all of the culpability on that company. This occurred after the Capital One data breach, where “Amazon denied any responsibility, stating that their systems weren’t at fault” (Heiligenstein, 2023). This stance purposely fails to acknowledge how AWS’s bloat of services makes proper cybersecurity very difficult and paints the tech giant in a friendly light.

The last set of connections I am focusing on is that of the attackers and the customers of AWS powered businesses. From an attacker’s perspective, Amazon overreliance of the Shared Responsibility Model suggests that not much will change and that their current tactics will continue to work. Afterall, if nine data breaches have been due to unprotected S3 buckets, then that incentives them to keep looking for that vulnerability. Unfortunately, another large exposure from S3 seems probable, which means more people will have sensitive information stolen from them. I am positive more than 16% of people have a decent understanding of the cloud in 2024, but still doubt most people could explain it coherently (McKendrick, 2012). And even in recent years people have expressed confusion over confusing terminology in the field (Gibbs, 2022). This means people are having their personal data exposed by technology they do not really understand or know about. They are influenced by AWS and these companies all while mostly being in the dark.

Discussion

Amazon has unquestionably made a very useful product out of AWS. Millions of businesses use it and are happy to do so given the growth of AWS’s profits (Michalowski, 2023). Almost all of these customers never experienced any massive information leaks, and most probably never will. Amazon deserves a lot of credit for making fundamentally sound services

that have yet to be directly hacked into. But these large data breaches are still unacceptable, and Amazon deserves its share of the blame. While no system will ever be 100% secure, having hundreds of millions of people's private information is a serious matter where improvements must be made.

If Amazon was the only cloud provider, none of their complication issues would ever get solved due to complacency, and the same attacks will continue to be effective. Similarly, if they maintain their market share of around 30%, they will continue to prioritize new services over simplifying what they already have (Cohen, 2021). I think the best-case scenario is for the cloud market to equalize, with each large corporation having an approximately equal share of the pie. Due to the need for large datacenters across the globe, cloud computing will have to be done by those with ample resources. If every major player is unable to substantially differentiate their total offerings from their competitors, hopefully they will focus on refining what they have to gain a slight edge. Ideally a more standardized terminology for their services to make things less confusing to the general public. Ultimately, the future of cloud computing is still up in the air, and I remain optimistic about the security of Amazon Web Services.

References

Amazon. (n.d.). *Data Privacy FAQ*. Amazon. https://aws.amazon.com/compliance/data-privacy-faq/

Amazon. (n.d.). *Shared responsibility model*. Amazon Web Services. https://aws.amazon.com/compliance/shared-responsibility-model/

Callon, M. (1984). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *The sociological review*, *32*(1_suppl), 196-233.

Cohen, J. (2021, December 29). 4 companies control 67% of the world's cloud infrastructure. PCMAG. https://www.pcmag.com/news/four-companies-control-67-of-the-worlds-cloud-infrastructure

Gibbs, M. (2022, February 18). *The Hidden Truth Behind Cloud Computing*. TechRadar. https://www.techradar.com/features/the-hidden-truth-behind-cloud-computing

Harris, S. (2019, August 21). *Capital One Breach · Ramblings of a Unix Geek*. SWEHarris. https://www.sweharris.org/post/2019-08-21-capital-one/

Heiligenstein, M. X. (2023, October 5). *Amazon Web Services (AWS) data breaches: Full timeline through 2023*. Firewall Times. https://firewalltimes.com/amazon-web-services-data-breach-timeline/

Kapko, M. (2022, November 30). *AWS CEO stresses the core elements of cloud security*. Cybersecurity Dive. https://www.cybersecuritydive.com/news/aws-ceo-cloud-security/637623/

Khan, S., Kabanov, I., Hua, Y., & Madnick, S. (2022). *A systematic analysis of the Capital One Data Breach: Critical Lessons Learned*. ACM Transactions on Privacy and Security, 26(1), 1–29. https://doi.org/10.1145/3546068

Latour, B. (1992) *'Where are the missing masses? The sociology of a few mundane artifacts'*, in Bijker, W. E. and Law, J. (eds) *Shaping Technology/Building Society: Studies in Sociotechnical Change*, Cambridge, MA, MIT Press, pp. 225-58

McKendrick, J. (2012, August 29). *Most Americans don't understand cloud computing: Does it really matter?* Forbes. https://www.forbes.com/sites/joemckendrick/2012/08/29/most-americans-dont-understand-cloud-computing-does-it-really-matter/?sh=7b2e2b874ef7

Michalowski, M. (2023, July 24). *Who's using Amazon Web Services? [2023]*. Spacelift. https://spacelift.io/blog/who-is-using-aws

Novaes Neto, N., Madnick, S., de Paula, M. G., & Malara Borges, N. (2020). *A case study of the capital one data breach*. Stuart E. and Moraes G. de Paula, Anchises and Malara Borges, Natasha, A Case Study of the Capital One Data Breach (January 1, 2020).

Quiggle, J. (2023, June 5). *Amazon Has Explicit Words On Concerns It Broke Privacy Laws*. The Street. https://www.thestreet.com/investing/stocks/amazon-has-explicit-words-on-concerns-it-broke-privacy-laws

Sismondo, S. (2010). *An introduction to science and technology studies*. Wiley-Blackwell.

Team Intricately. (2022, May 12). *Introducing the AWS ecosystem in 2022 report*. Intricately. https://blog.intricately.com/hg-insights-intricately-aws-ecosystem-report

UpGuard Team. (2017, January 19). *The RNC files: Inside the largest US voter data leak:*

Upguard. UpGuard. https://www.upguard.com/breaches/the-rnc-files