Atticus Emilsson

CS462

November 24,2024

Cybersecurity Fundamentals

Attack on the Internet Archive

Introduction

The Internet Archive (https://archive.org/) is a 501(c)(3) non-profit organization that aims to preserve digital media/mediums present on the internet. This includes books, movies, television shows, YouTube clips, music videos, and many other forms of digital information. The Internet Archive also hosts a service known as "The Wayback Machine" which enables individuals to archive a website on a specific date, allowing for others to go back and view the archived contents even if the website has been taken down.

The Internet Archive provides various incredibly important services that assist with preserving digital information and making it available to everybody; however, their mission has been controversial with them often being targeted by large corporations who feel that this is a violation of their intellectual property. This has led to multiple legal battles, with _Hachette v. Internet Archive_ being the most notable. The Hachette Book Group is a publishing company, owned by Hachette Livre which is the "third-largest publisher in the world" (Hachette, 2019). The Hachette Book Group alleged that the Internet Archive scanned copyrighted works and publicly redistributed them for free and without permission (Stempel, 2024). On September 4, 2024, the U.S. Court of Appeals maintained their previous stance, which determined that the Internet Archive's actions were indeed damaging to the publishers (Stempel, 2024). The original ruling issued a permanent injunction that rendered the Internet Archive to provide the publisher's works through their CDL (Controlled Digital Lending). These instances show that the Internet Archive is no stranger to legal battles; however, they are now required to divert their attention to incoming cyber-attacks.

Breach 1

In late 2024, The Internet Archive suffered a catastrophic data breach which lead to the compromise of 31 million user records (Abrams, 2024). These compromised records include various sensitive data points such as email addresses, display names, and passwords (Abrams, 2024). Although the passwords were hashed using Bcrypt, this does not mean that they are impenetrable (Abrams, 2024). There are entire communities dedicated to cracking hashed passwords, and whilst some are computationally infeasible, others are not. This is why it is of critical importance to always leverage strong and complex passwords. The compromised data also included timestamps of password-changes, which enabled researchers to determine with confidence that the data breach occurred on September 28th, 2024 (Shaikh, 2024). This is only the first of a series of attacks targeting The Internet Archive.

Unfortunately, as of October 9, 2024, it is unknown how attackers managed to compromise the data of 31 million Internet Archive users (Abrams, 2024). After the initial breach, users who accessed the Internet Archive were presented with a message informing them that the Internet Archive was compromised and the sensitive information of 31 million users was going to be uploaded to HIBP (HaveIBeenPwned), a popular cyber-security tool which enables individuals to determine if their information has been compromised in a data-breach (Abrams, 2024). The best information at the time of this response was that the attacker(s) leveraged a JS (JavaScript) library to deface the website and present the message; otherwise, researchers were in the dark as to how the system was breached (Abrams, 2024).

GitLab is a public, online platform that hosts code repositories, including open-source code for the Internet Archive. The platform serves many purposes alongside the hosting of open-source code, such as Git-based version control, bug tracking, access control, and feature requests. It was later discovered that hackers discovered an exposed configuration file on GitLab, which contained an authentication token that enabled them download the source code of the Internet Archive (Abrams, 2024). There were also additional credentials discovered which granted attackers access to their database management system (Abrams, 2024). This is the avenue attackers used to extract the sensitive user information. Researchers determined that this vulnerability has been open since December 2022 (Abrams, 2024).

Breach 2

As the Internet Archive was actively responding to the aforementioned data-breach involving the compromise of their database management system and 31 million user records, they were also being affected by a DDoS (Distributed Denial-Of Service) attack. A DoS (Denial-Of-Service) attack is a cyber-enabled attack which aims to render networked systems/services unavailable to legitimate users (Cloudflare, 2019). A common method of invoking a DoS attack involves overwhelming servers/services with illegitimate network traffic, which renders web-servers unable to process legitimate traffic (Cloudflare, 2019). An identifying characteristic of a DoS attack (as opposed to DDoS) is the requests are launched from a single system/network. This typically makes DoS attacks much easier to prevent/respond to, as modern network security features are capable of identifying and blocking said machine; however, the issue is much more complex with DDoS attacks which entail requests being launched from many different geographically separate systems. The distributed nature of DDoS attacks make it incredibly difficult to effectively prevent further illegitimate traffic, especially those of massive scales. DDoS attacks are powered by botnets, which are essentially large networks of compromised systems which can be instructed to carry-out certain tasks. One such task is to flood a web-server with illegitimate traffic. Some botnets can be composed of millions of compromised, geographically separated systems, which make it virtually impossible to protect a web-server without essentially blocking all access. This still achieves the same goal of rendering legitimate users to access services; however, may prevent other damages. The Internet Archive responded by disabling access to their services in order to protect content (Abrams, 2024). After this, they slowly rolled out read-only access to services to ensure availability whilst still giving themselves time to properly strengthen their security posture.

SN_BlackMeta, an alleged pro-Palestinian group, claimed responsibility of the DDoS attacks that hammered Internet Archive systems (Abrams, 2024).

Breach 3

After the compromise which leveraged exposed authentication tokens, the Internet Archive suffered a third breach which involved their Zendesk ticketing system (Abrams, 2024). Similarly, this breach involved an exposed token within a configuration file (Abrams, 2024). This granted attackers access to over 800,000 support tickets sent to the Internet Archive (Abrams, 2024). Although sorting through these tickets may be much more difficult than processing the structured data breach containing 31 million user records, these tickets can certainly contain incredibly sensitive information which may be damaging to the individuals whose information had been leaked. One such example is individuals who were required to upload their ID in order to successfully request removal of information from the Internet Archive (Abrams, 2024). This not only binds an individual to this content, but the information coupled with leaked credentials could further enable attacks such as identity theft and fraud. Many online services, especially banking, gambling, and adult content providers, require that users upload their ID in order to verify their identity and/or age. Adversaries can leverage this information to register/authenticate as another individual. This third incident shows not only a failure to secure sensitive credentials, but also shows a failure to properly respond to previous incidents. The Internet Archive had time to rotate their API keys and further secure their systems; however, they did not act accordingly which led to the further compromise of sensitive information and user data.

Conclusion

The Internet Archive, a 501(c)(3) non-profit organization was victim of a series of cyber-attacks which led to the compromise of sensitive user information and the disruption of services. It is unknown exactly why these attacks took place. SN_BlackMeta, an alleged pro-Palestinian group, claimed responsibility for the DDoS attacks which led to the Internet Archive disabling access to their services; however, the two data breaches have not been confidently attributed to any one group (Abrams, 2024). Conspiracies certainly exist as to why some of the attacks occurred. SN_BlackMeta claims that they attacked the Internet Archive services because they have roots in the United States, a country which supports the "genocide that is being carried out by the terrorist state of Israel" (Petkauskas 2024). Any other attempts at attributing the attack to a group or determining the motive behind the attacks are simply guesswork. Given the nature of the breach, it is quite likely that the attacks were opportunistic in nature.

DDoS attacks, as aforementioned, are incredibly hard to defend against. This is why the Internet Archive eventually decided to disable access to their services. This time enabled them to take a step back, observe the attack, and implement a plan of action. Unfortunately, their response to the data breaches involving exposed authentication tokens was less than sufficient. They did not properly, nor timely, rotate their API keys even after being made aware of the existing vulnerabilities. This led to unnecessary compromise which further damaged their reputation, and the confidentiality of individual information. This does however, serve as an incredibly valuable learning experience that emphasize the importance of timely response to vulnerability disclosures. The necessity for sufficient response is further exacerbated by the threat presented towards critical information and sensitive credentials.

Works Cited

"About Hachette Book Group." *Hachette Book Group*, Hachette, 16 Jan. 2019, www.hachettebookgroup.com/landing-page/about-hachette-book-group-2/. Accessed 24 Nov. 2024.

Abrams, Lawrence. "Internet Archive Breached Again through Stolen Access Tokens." *BleepingComputer*, 20 Oct. 2024, www.bleepingcomputer.com/news/security/internet-archive-breached-again-through-stolen-access-tokens/. Accessed 24 Nov. 2024.

---. "Internet Archive Hacked, Data Breach Impacts 31 Million Users." *BleepingComputer*, 9 Oct. 2024, www.bleepingcomputer.com/news/security/internet-archive-hacked-data-breach-impacts-31-million-users/. Accessed 24 Nov. 2024.

Cloudflare. "What Is a Denial-of-Service (DoS) Attack?" *Cloudflare*, Cloudflare, 2019, www.cloudflare.com/learning/ddos/glossary/denial-of-service/. Accessed 24 Nov. 2024.

Internet Archive. "About the Internet Archive." *Archive.org*, 2009, archive.org/about/. Accessed 24 Nov. 2024.

Petkauskas, Vilius. "Internet Archive Hacking Drama: Why Did They Do It?" *Cybernews*, 11 Oct. 2024, cybernews.com/editorial/internet-archive-hack-drama-explained/. Accessed 24 Nov. 2024.

Shaikh, Roshan Ashraf. "Internet Archive Hacked and 31 Million User Accounts Leaked — Hacking Group "SN_Blackmeta" Claims Responsibility." *Tom's Hardware*, 10 Oct. 2024, www.tomshardware.com/tech-industry/cyber-security/internet-archive-hacked-and-31-million-user-accounts-leaked-hacking-group-sn-blackmeta-claims-responsibility. Accessed 24 Nov. 2024.

Stempel, Jonathan. “Major Book Publishers Defeat Internet Archive Appeal over Digital Scanning.” *Reuters*, 4 Sept. 2024, www.reuters.com/legal/major-book-publishers-defeat-internet-archive-appeal-over-digital-scanning-2024-09-04/. Accessed 24 Nov. 2024.