Dominic Drury

06/06/2024

CS-405

5-1 Case Study: Triple A and Defense in Depth

**Introduction:**

The name of this case is Details On 700 Million LinkedIn Users For Sale On Notorious Hacking Forum ([link to article](https://www.forbes.com/sites/leemathews/2021/06/29/details-on-700-million-linkedin-users-for-sale-on-notorious-hacking-forum/?sh=5e61571b34a4)) and it occurred in June of 2021. This made the news because it impacted 700 million users and was just 2 months after the data of 500 million users had their data scraped and put up for sale (Mathews, 2021).

**Describe the breach:**

The data leak was caused by the misuse of LinkedIn’s API, which allowed user data to be collected (Baek, 2023). LinkedIn claims that this was not a data breach, but rather a data scrape that collected public data on its users and that technically no private user data was exposed but some information that is not normally available publicly, like some of the user email addresses, were a part of the data collected and put up for sale (Cimpanu, 2021). I believe that LinkedIn was targets solely because they had a vast user database and their API did not prevent the data scraping from occurring.

**Identify the threat(s):**

The immediate threat is that the user data was put up for sale on a notorious data marketplace and was available for purchase from people aiming to create sophisticated phishing attacks or identity theft (Mathews, 2021; Baek, 2023). The potential threat is that this process of scraping data from sites is just a potential violation of the sites Terms of Service and that sifting through public information will make attackers far more intelligent than can be kept up with. As of April 2022, the Ninth Circuit court of appeals has rulled that scraping public information is within the legal rights of companies (POLSINELLI, 2023).

**What could a developer have done to prevent this breach?**

This question is tricky, but some things that can help prevent this kind of attack are limiting the rate that a client can make requests to slow down the rate of data collection or throw an error if the limit is exceeded, protect your site from automated traffic to stop automatic scrapping, blacklist malicious users that use known malicious IP’s or known scrapping libraries, and using a technique called honey-potting where you have buttons not rendered by the webpage, so they can only be accessed by bots, in order to root out non-human users (Jsamuelbaron, 2020).

**Summarize the case by explaining the role of best practices, Triple A and defense in depth in preventing future attacks.**

The best way to prevent this type of attack is to ensure that all users are human users and not bots scaping your database for user information (public or not). You need to authenticate the users on your site to ensure that they are in fact who they say they are, ensure that the users only have access to the systems they are authorized to use, and you need to make sure that you keep a record of the data that is accessed so if an attack does occurred you can inform the appropriate parties, or even prevent an attack by observing the beginning signs of a scraping attack. You need to have these security measures be at multiple layers of your security and be independent of each other so that if an attacker were to bypass one factor, you can catch them at another.

References:

Mathews, L. (2021, June 29). *Details On 700 Million LinkedIn Users For Sale On Notorious Hacking Forum*. Https://www.Forbes.com/. Retrieved June 6, 2024, from <https://www.forbes.com/sites/leemathews/2021/06/29/details-on-700-million-linkedin-users-for-sale-on-notorious-hacking-forum/?sh=5e61571b34a4>

Baek, D. S. (2023, December 10). *LinkedIn Data Breach (700M) in 2021*. Https://www.Linkedin.com/. Retrieved June 6, 2024, from <https://www.linkedin.com/pulse/linkedin-data-breach-700m-2021-david-sehyeon-baek--ex85c/>

Cimpanu, C. (2021, September 21). *Hackers leak LinkedIn 700 million data scrape*. Https://Therecord.Media/. Retrieved June 6, 2024, from <https://therecord.media/hackers-leak-linkedin-700-million-data-scrape>

POLSINELLI (2023, March 9). *Data Scraping Update: ‘LinkedIn v. HiQ’ Answers Some Questions but Leaves Many More Open*. Https://www.Polsinelli.com/. Retrieved June 6, 2024, from <https://www.polsinelli.com/publications/data-scraping-update-linkedin-v-hiq-answers-some-questions-but-leaves-many-more-open#:~:text=In%20June%202021%2C%20the%20Court,593%20U.S.%20__%20(2021>).

Jsamuelbaron (2020, December 14). *How to Protect Your Web Application From Web Scraping.* Https://Medium.com/. Retrieved June 6, 2024, from <https://medium.com/swlh/how-to-protect-your-web-application-from-web-scraping-cc01ec7ddadd#id_token=eyJhbGciOiJSUzI1NiIsImtpZCI6IjY3NGRiYmE4ZmFlZTY5YWNhZTFiYzFiZTE5MDQ1MzY3OGY0NzI4MDMiLCJ0eXAiOiJKV1QifQ..b07LdR6R3mD7JcwyCdlkVCPYdXX8TvQ-elaFdX8fiAZ1KIPVpkwAnARAdkbj80X15XLeCWUvB7MdBo19i9MM3C_-y23bp4y45q4Y8chQc2_CLnD3lv0qNkfd97MChRRTLbGsoc0d9CnvqSHEFOwhmF7OghaoAT4pAjb-im7Cs61FY1x_1RUAlqf2GOlvzKc4GQij37x0UjPUB7wqXwAOjsazN9UYot4cEHc1cKtAnGoq1zsETgOPmLD1_VdeYbEQMSb5SHQVT4VXYjF7RbfJ3fEMG7F-2lOvNk0ZaMmNpuS1hoj4DdWoxBgDaOQe15jR84WN0ItL6m3cL4L-DbrtvA>