LinkedIn Data Breaches

Prasanthi Desiraju, Vittorio Pepe, and Will Shin

MSDS 485 Data Governance, Ethics & Law

May 2, 2021

INTRODUCTION

LinkedIn, a social media network, is the de facto unique player in the professional networking arena. Started in 2003 as a platform where users could post their CVs and employees post job openings, it quickly became a sales tool and a means to keep business connections. It allows both free and the premium user and as of February 2021 has over 740 million registered members. LinkedIn is available as a website, desktop version, and mobile device app. It offers a variety of services, including messaging, emails, and more recently, a learning platform that provides a variety of professional courses.

BACKGROUND

There is currently very little regulation concerning data privacy issues. The most prominent ones that affect social media platforms such as Facebook or LinkedIn are the California Breach Notification Law ("CalBNL") introduced in 2004 and modified in 2019, and the European General Data Protection Regulation (GDPR) issued in 2018. Both regulations are intended for residents in the respective states or nations. Still, they have provisions to enforce public disclosure of the accidents to the advantage of all the users affected.

ISSUES - CASE STUDY 1

LinkedIn has been the involuntary protagonist of several data breaches, one as recent as April 6, 2021. One of the most infamous and the first is the one that occurred on June 5, 2012. Since the breach object of this study was relative to the year 2012 and 2016, the only applicable regulation was the Californian CalBNL.

The initial report was that 6.5 million usernames and passwords were hacked, and the user lost access to their accounts.[1] The company quickly acted, and on June 6, the CEO Vicente Silveira posted on the LinkedIn official blog confirmed that the data breach occurred and

stated that the affected user would receive an email to reset the password and noted that the breach affected only the user that did not recently update the password.[2] The users who changed the password benefited from increased protection given by the recent introduction of higher security protocols to store the passwords.

The issue seemed contained, resolved, and adequately managed following the mandates of the California Breach Notification Law ("CalBNL", 2003). Nevertheless, in May 2016, a large dataset containing more 117 mln usernames and passwords of LinkedIn users surfaced for sale on the dark web.[3] The affected users had passwords that could be decoded and had been retrieved from a larger dataset containing 167 mln user data, that was originated from the breach that happened in 2012.

LinkedIn again swiftly posted an announcement acknowledging that the recently resurfaced data belonged to the 2012 breach and got in contact with all the users having an account in 2012 to reset the password and suggested to take advantage of two-factor authentication, to not reuse passwords between different websites and update the passwords often.[4]

The issue originated from the fact that although Linkedin provided hashing of the passwords, it did not provide it with salting prior to the breach. The hash algorithms encode the passwords, but they are deterministic algorithms. If a unique key (salt) is not used for each user, the hashing can be reversed using lookup tables, much like the Rosetta Stone.

At the moment of the breach, LinkedIn was aware of this problem and had just released new security features like advanced hashing and two-factor authentication. In fact, any user that had changed the password after the release of the new security features was not affected.

The main issue in this case is not the communication part of the compliance. Linkedin in both cases posted timely announcements, and, in the case of the 2016 occurrence, regular updates on their findings and countermeasures that were adopted.

Linkedin failed in estimating which users were affected, greatly reducing the effectiveness of their response. If all the users would have been notified at the first occurrence, the largest part of the user data would have been not useful to the hackers since the passwords would have been updated soon after the 2012 occurrence.

ISSUES - CASE STUDY 2

The most recent data breach involves selling data of 500 million LinkedIn profiles on a popular cybercriminal forum that transpired in April 2021.[5] LinkedIn alleges that this data is an aggregation from different websites and includes publicly viewable member data that might have been scraped from the website. LinkedIn also stated that it was not a data breach as no private information was leaked. While the statements make sense and technically correct, this leads to some open questions which are to be addressed in order to avoid such future hacks as we progress steep into the digital era.

1. Web scraping is ruled as legal in many instances when using public information. The LinkedIn-hiQ dispute is one of the most important web scraping cases to interpret the scope of Computer Fraud and Abuse Act (CFAA) liability. hiQ, an AI-based data analytics start-up, defeated LinkedIn's attempt to block its access to the public webpage in Ninth Circuit and the case is pending with the supreme court. Comparing this incident to the recent data breach, one could notice hiQ scraped the data for the development of its

sound products, and the recent data breach highlights the potential unethical use of data scraping, where an unknown party scraped the data and posted it for sale.[5]

2. Data Scraping techniques, though they make use of public information, may pose a serious threat to Organization's data privacy if it targets a significant chunk of the total user base.[6]
3. Once data is scraped, the end-user has nothing in his control to correct the situation and escape from the phishing attacks, which will eventually have a negative impact on the brand's reputation.

There are laws preventing the scraping of private data or copyrighted information; the law is unsettled for unethical use of public data, and there are no clear directions to address the implications of such data leaks. However, we could limit the information (PII) retrieved through web scraping by enforcing data minimization and laws introduced to protect user's privacy. General Data Protection Regulation (GDPR - For EU residents), California Consumer Privacy Act (CCPA) grants users control over their personal information and sets strict guidelines to Organizations on how they handle user data. LinkedIn is taking a global approach to GDPR to help ensure all members benefit from increased control and clarity irrespective of the location.[7] For example, in November 2018, LinkedIn added a new feature called 'Who can see my email address,' which users can use to hide their email address from everyone on their profile if they choose to (defaults to 1st connections).[8]

Despite implementing high security protocols, adhering to the regulatory requirements and constantly innovating on protecting user’s data privacy there are data breaches occurring time and again, through different techniques. LinkedIn is in the right path of applying these

measures and focus on public data minimization, user education implying the potential effects of marking his/her profile as public, and proactively updating the compliance/regulatory frameworks can help address the issues at least until laws around public data scraping become more coherent.

REFERENCES

1. "More than 6 million LinkedIn passwords stolen", 2012, accessed May 1 https://money.cnn.com/2012/06/06/technology/linkedin-password-hack/index.htm
2. "An Update on LinkedIn Member Passwords Compromised." 2021. *Recent Posts*. Accessed May 1. https://blog.linkedin.com/2012/06/06/linkedin-member-passwords-compromised.
3. "2012 Linkedin Breach had 117 Million Emails and Passwords Stolen, Not 6.5M", accessed May 1 https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/2012-linkedin-breach-117-million-emails-and-passwords-stolen-not-6-5m
4. "Protecting Our Members." 2021. *Recent Posts*. Accessed May 1. https://blog.linkedin.com/2016/05/18/protecting-our-members?version=meter+at+1&module=meter-Links&pgtype=article&contentId=&mediaId=&referrer=http%3A%2F%2Fnews.google.com%2F&priority=true&action=click&contentCollection=meter-links-click
5. Neuburger, Jeffrey D. 2021. "Trove Of Online LinkedIn User Data Fuels LinkedIn's Anti-Scraping Position - Privacy - United States." *Welcome to Mondaq*. Proskauer Rose LLP. April 14. https://www.mondaq.com/unitedstates/privacy-protection/1057886/trove-of-online-linkedin-user-data-fuels-linkedin39s-anti-scraping-position.
6. Canales, Katie. 2021. "Hackers Scraped Data from 500 Million LinkedIn Users - about Two-Thirds of the Platform's Userbase - and Have Posted It for Sale Online." *Business Insider*. Business Insider. April 8. https://www.businessinsider.com/linkedin-data-scraped-500-million-users-for-sale-online-2021-4?r=DE&IR=T.
7. "GDPR." 2021. *Privacy*. Accessed May 1. https://privacy.linkedin.com/gdpr.
8. "Why LinkedIn Is the Ray of Sunshine in the GDPR Aftermath for B2B Companies - E-Business Promotion." 2019. *E*. April 5. https://www.e-businesspromotion.co.uk/why-linkedin-is-the-ray-of-sunshine-in-the-gdpr-aftermath-for-b2b-companies.