Cybersecurity and Information Security Newsletter

Issue 18 | May 13, 2022

Table of Contents

Daniel Shin, research scientist with the Commonwealth Cyber Initiative (CCI) Coastal Virginia region, wants to hear from you! Submit any cybersecurity and information security news items or request related topics, via e-mail to dshin01@wm.edu.

This newsletter supports the mission of CCI. To learn more about CCI, including upcoming events, funded research, and news, please visit cyberinitiative.org.

The Ninth Circuit holds that data scraping of publicly available information does not implicate the CFAA

On April 18, 2022, the U.S. Court of Appeals for the Ninth Circuit (Ninth Circuit) ruled on a preliminary injunction case involving data scraping of publicly available information. *HiQ Labs v. LinkedIn*, No. 17-16783, available here. Data scraping refers to the extracting and saving data from output generated from another program. *What is data scraping?*, available here. The court held that hiQ Labs, Inc. (hiQ), a data analytics company, did not violate the Computer Fraud and Abuse Act (CFAA) when it scraped publicly available data from LinkedIn Corp. (LinkedIn), a social media company focusing on professional networking and career development of its users. This ruling was based partially on last year's U.S. Supreme Court decision of *Van Buren v. United States*, 141 S. Ct. 1648 (2021), which limited the scope of criminal violations under the CFAA. *Issue 8: CLCT Cybersecurity and Information Security Newsletter - U.S. Supreme Court limits the scope of criminal violation under the Computer Fraud and Abuse Act*, available here. This decision may hold implications for the question of how to balance data protection against gathering information from online platforms.

Case History

LinkedIn is a social media website that allows its members to post content, resumes, and job listings and create professional networking "connections" with other users. *HiQ Labs v. LinkedIn*, No. 17-16783, *supra* at 8. According to its user agreement, LinkedIn members own the content and information they post on the social media site, and the social media company is granted only a non-exclusive license to use, copy, modify, distribute, publish, and process member posted information. *Id.*

LinkedIn members have control over how certain information is visible to other users of the website. They can limit certain information to be viewable to only direct connections, to a member's "network" (other members that are within three degrees of connectivity), or to the general public (including all members and nonmembers using the site).

HiQ is a data analytics company that focuses on analyzing scraped data from LinkedIn profiles to generate "people analytics," a type of proprietary statistical analysis that gets sold to business clients. It uses automated techniques (bots) to scrape publicly available LinkedIn profiles. According to the court's opinion, hiQ does not seem to have the techniques, nor has it attempted, to harvest information that LinkedIn members have restricted from public view.

LinkedIn has implemented certain technical steps to prevent automated data scraping. For example, the social media network has set explicit instructions via the "robots.txt" file that only allows certain entities, such as a search engine web crawler (an automated program that discovers new webpages and indexes them for a search engine), to access the site. Because "robots.txt" is merely an instructional tool that can be ignored by other automated bots, LinkedIn also utilizes other tools to detect, limit, and block data scraping bots from accessing the social media platform. *HiQ Labs v. LinkedIn*, No. 17-16783, *supra* at 10-11.

On May 23, 2017, LinkedIn sent a cease-and-desist letter to hiQ, alleging that hiQ's data scraping activities violated LinkedIn's user agreement and demanded the company stop accessing LinkedIn servers. The cease-and-desist letter further alleged that hiQ's activities were in violation of the CFAA, the Digital Millennium Copyright Act (DMCA), California Penal

Code § 502(c), and the California common law of trespass. LinkedIn also stated that technical measures were implemented to prevent hiQ from data scraping the social media platform.

HiQ responded to the cease-and-desist letter by demanding that LinkedIn "recognize hiQ's right to access LinkedIn's public pages," or face potential injunction action. *Id.* at 13. The two companies attempted to find an agreement on the matter, but LinkedIn continuously prevented hiQ from accessing its platform and sent another cease-and-desist letter on June 24, 2017. As a result, hiQ filed a lawsuit against LinkedIn at the U.S. District Court for the Northern District of California. It asked the court, among many things, for a preliminary injunction that would restrain LinkedIn from restricting hiQ's access to its social media platform while the litigation is pending.

A preliminary injunction is a court order that requires a party to do or cease doing a specific action to preserve the status quo before final judgment. *Preliminary injunction*, available here; *see also Injunction*, available here. A preliminary injunction motion is an extraordinary remedy that courts are hesitant to grant unless the plaintiff meets the high legal burden of "by a clear showing." *Mazurek v. Armstrong*, 520 U.S. 968, 972 (1997), available here. To prevail on its motion for a preliminary injunction against LinkedIn, hiQ had to establish several factors, including that hiQ is likely to succeed on the merits, is likely to suffer irreparable harm in the absence of preliminary relief by the court, the balance of equities tips in its favor, and the preliminary injunction motion is in the public interest. *Winter v. Nat. Res. Def. Council*, 555 U.S. 7, 20 (2008), available here.

On August 14, 2017, the district court granted hiQ's motion for a preliminary injunction against LinkedIn. *HiQ Labs v. LinkedIn*, 273 F. Supp. 3d 1099 (N.D. Cal. 2017), available here. The court determined that hiQ had argued successfully the necessary elements required for a preliminary injunction against LinkedIn. Specifically, the court found that:

1. hiQ raised serious questions about the applicability of the CFAA to its data scraping operation;
2. hiQ would face irreparable harm in the absence of a preliminary injunction, because it will likely be driven out of business; and
3. the preliminary injunction motion is in the public interest, because giving private entities like LinkedIn "the blanket authority to block viewers from accessing information publicly available on its website for any reason, backed by sanctions of the CFAA, could pose an ominous threat to public discourse and the free flow of information promised by the Internet." *HiQ Labs v. LinkedIn*, 273 F. Supp. 3d 1099, 1119, *supra*.

In the preliminary injunction order, the court instructed LinkedIn to stop (1) preventing hiQ's access, copying, or use of public profiles on LinkedIn's website, and (2) implementing any legal or technical mechanisms that prevent hiQ from accessing public information on the platform. The court also ordered LinkedIn to withdraw the cease-and-desist letters to hiQ and refrain from issuing further such letters during the trial.

LinkedIn appealed the district court's order to the Ninth Circuit. After reviewing the district court's decision, on September 9, 2019, the Ninth Circuit affirmed the ruling. *HiQ Labs v. LinkedIn*, 938 F. 3d 985 (9th Cir. 2019), available here. On March 9, 2020, LinkedIn filed a petition for a writ of certiorari to the U.S. Supreme Court to appeal the Ninth Circuit's decision. The petition asked the Supreme Court to consider the following legal issue:

> "Whether a company that deploys anonymous computer 'bots' to circumvent technical barriers and harvest millions of individuals' personal data from computer servers that host public-facing websites—even after the computer servers' owner has expressly denied permission to access the data—'intentionally access a computer without authorization' in violation of the Computer Fraud and Abuse Act." *On Petition for a Writ of Certiorari to the United States Court of Appeals for the Ninth Circuit (19-1116)*, available here.

On June 14, 2021, the Supreme Court granted LinkedIn's petition and vacated the Ninth Circuit's decision and remanded the case for further consideration in light of the recently ruled *Van Buren v. United States*. *U.S. Supreme Court Order List, Monday, June 14, 2021,* available here. (Note, the Supreme Court did not conduct a public hearing when determining LinkedIn's petition.)

On April 18, 2022, the Ninth Circuit issued its second ruling in light of *Van Buren*, and it affirmed the district court's order for the preliminary injunction. The court noted that the Supreme Court in *Van Buren* introduced a "gates-up-or-down inquiry—one either can or cannot access a computer system, and one either can or cannot access certain areas within the system." Meaning, *Van Buren* only focuses on whether an individual was granted access to a particular system, not the subjective intention behind granting such access when applying liability under the CFAA.

The Ninth Circuit determined that *Van Buren's* "gates-up-or-down inquiry" reinforced its previous determination that hiQ's data scraping techniques fail to implicate the CFAA. The court notes that "a user with authorization is not subject to limitations on access, whether those limitations are code-based or contained in contracts or policies." *HiQ Labs v. LinkedIn*, No. 17-16783, *supra* at 35. If a computer resource requires authorization and such authorization has been given to a user, then the gates are up; under the same scenario, if authorization has not been given, then the gates are down. *Id.* at 36.

Applying to websites, the court observed that publicly available webpages are "open to anyone with a web browser," where there are "no gates to lift or lower in the first place." *Id.* Subsequently, the court concluded that the CFAA is not implicated when a user accesses data that is hosted on a computer network that generally permits public access to its data. *Id.* at 40. Because LinkedIn public pages permit public access without any authorization mechanisms (e.g., user account login), the CFAA "gates" contemplated by both the U.S. Supreme Court and the Ninth Circuit do not apply to hiQ's data scraping operations. Finding those other elements for a preliminary injunction had been met, the Ninth Circuit affirmed the district court's ruling the second time, allowing hiQ to continue its data scraping operations during its trial.

Analysis

Data scraping operations targeting social media platforms have undergone controversy because they involve personally identifiable information of users. For example, on July 24, 2019, the Federal Trade Commission (FTC) filed an administrative complaint against Cambridge Analytica, a now-defunct data analytics company, for allegedly employing "deceptive tactics to harvest personal information from tens of millions of Facebook users for

voter profiling and targeting." *FTC Sues Cambridge Analytica, Settles with Former CEO and App Developer [FTC]*, available here.

According to the FTC's complaint, Cambridge Analytica partnered with a Facebook app developer to repurpose a survey app, "thisisyourdigitallife," to collect the personal information of not only the consenting app users but also the unconsenting friends of users. The survey app paid Facebook users for completing surveys, but almost half of the app users refused to share their Facebook profile information with the app. According to the FTC, the app falsely told its users that it would not "download your name or any other identifiable information." Relying on the app's statement, users who granted the app access to their Facebook profile had personally identifiable information harvested not only from their profiles but also from their friend's profiles.

Cambridge Analytica was mainly about a case of deceptive business practices that ended up in a settlement with the FTC on December 18, 2019. *FTC Grants Final Approval to Settlement with Former Cambridge Analytica CEO, App Developer over Allegations they Deceived Consumers over Collection of Facebook Data*, available here. This incident also highlighted Facebook's poor data control mechanisms, which allowed permissioned Facebook apps to access the data of users' friends if the user granted the app access to their profile.

Although Facebook announced stricter data control mechanisms, the Cambridge Analytica incident highlights the risk of having poor data access control mechanisms, especially with systems hosting personally identifiable information. In fact, in June 2020, hackers released details of 533 million Facebook user accounts, which included names, phone numbers, emails and other personally identifiable information. *533 million Facebook users' phone numbers leaked on hacker forum*, available here. These data may have been compiled by using data scraping techniques targeting public Facebook profiles and the now-patched "Add Friend" bug that allowed threat actors to gain access to the target user's phone number without gaining proper authorization.

The ongoing case between hiQ and LinkedIn demonstrates the importance of implementing proper data access controls from both data platforms (e.g., social media websites) and users. LinkedIn may have implemented reasonable technical barriers to prevent third-party data scraping operations, but LinkedIn users with public profiles may not have realized that their public data could have been subject to data archival by another party. Given that LinkedIn is prohibited from preventing hiQ's data scraping operations, users should consider implementing stricter access controls using LinkedIn's profile management tool to limit what information is publicly available.

In its opinion's footnote, the Ninth Circuit mentions that LinkedIn's cease-and-desist letter asserted California's common law of trespass to chattels, which refers to the use or interfering of property without permission from the owner. *HiQ Labs v. LinkedIn*, No. 17-16783, *supra* at 41 n.21; *Trespass to Chattels*, available here. The court speculates that data scraping may exceed the scope of the website owner's consent of access, which may give rise to a common law tort claim for trespass to chattels *if LinkedIn is able to demonstrate harm*. Given that LinkedIn does not own information posted publicly by its members, it may be difficult for the company to demonstrate harm from hiQ's data scraping operations.

This case demonstrates the legal difficulty of prohibiting data archival operations targeting publicly available websites. As of this newsletter publication, LinkedIn is prohibited from stopping hiQ's data scraping operations both using legal and technical means. Organizations and individuals should be aware that there may be limited legal options available in case data, including data involving personally identifiable information, were harvested because it was made available on a public website.

To the extent that it is commercially viable, LinkedIn could consider pivoting away from displaying public user information to the open web and instead implementing a registration requirement for all users seeking to access its platform to address the data scraping issue. It is important that users are educated to the fact that publicly displaying information is likely to be subject to data scraping by third parties. Implementing a much stricter data access control policy may have to become the norm, especially if the data processor has no possessory right over the data.