Hashed & Salted | A Privacy and Data Security Update
May 2022

Ninth Circuit Provides Path Forward for Web Scraping of Public Data

In *hiQ Labs, Inc. v. LinkedIn Corp.*, the Ninth Circuit considered whether the Computer Fraud and Abuse Act (CFAA) could be invoked to preempt state law claims arising out of the web scraping of publicly available data from a website owned by another entity.

The appeal stemmed from hiQ Labs Inc.'s filing of a motion for a preliminary injunction to prevent LinkedIn Corp. from blocking hiQ's web scrapers from harvesting publicly available data from LinkedIn's website. In response, LinkedIn raised several affirmative defenses, including preemption of hiQ's state law claims under the CFAA.

Unlike other circuit courts, the Ninth Circuit took a narrow view of the CFAA, concluding that hiQ "raised serious questions" about whether LinkedIn may invoke the CFAA to preempt hiQ's state law claims.

What Is Web Scraping?

Web scraping or web harvesting is the extraction of data from a website. It is a form of copying in which specific data is located on and then copied from a website. Web pages are built using text-based markup languages such as HTML and often contain useful data in text form. Data that is collected from a web page via scraping is loaded into a database or exported into a format that can be utilized by a user, such as a spreadsheet.

Although web scraping can be done manually by a person via copy and paste, it is generally conducted by an automated tool often referred to as a "web bot" or "bot," especially when large amounts of data are being scraped from the target website. Popular uses of web scraping include, for example, obtaining comparative shopping data, lead generation, real estate listings, brand and reputation monitoring, and industry statistic and insight generation.

Web scraping is accomplished using two tools: a web crawler and a web scraper. The web crawler browses or "crawls" the internet to search for and index content by following various links. A web crawler may look for one specific website or may be used to discover URLs for various web pages, which it then passes on to the web scraper.

The web scraper is a specialized tool designed to quickly and accurately extract data from a web page that has been found by the web crawler. The web scraper may extract all the data from the web page or only certain data specified by the user. Web scrapers vary in design and complexity depending on the nature of the project.

The Computer Fraud and Abuse Act

The CFAA was enacted in 1984 to address unlawful access to government and financial IT systems, and made "unauthorized access" (i.e., hacking) of government computers a felony. In 1996, the CFAA was amended to

Attorney Advertising

extend the prohibition of "unauthorized access" to any "protected computer," not just government computers.

The CFAA states: "Whoever . . . intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer . . . shall be punished" by fine or imprisonment. 18 U.S.C. § 1030(a)(2)(C). A "protected computer" is any computer "used in or affecting interstate or foreign commerce or communication." 18 U.S.C. § 1030(e)(2)(B).

Over the years, companies have attempted to use the CFAA to prohibit web scraping activity, claiming that web scraping violated the "without authorization" clause of the statute, since to gather data a web scraper must access a "protected computer."

hiQ Labs, Inc. v. LinkedIn Corp.

Data analytics company hiQ Labs Inc., founded in 2012, uses an automated web bot to scrape data from publicly available information on LinkedIn Corp.'s website, including names, job titles, work histories and skills. The company then analyzes the harvested data to provide "people analytics" to its clients. At the time hiQ filed suit, hiQ offered two data analytics products: one that identified employees at the greatest risk of being recruited away and another that summarized employees' skills to help employers identify skill gaps so that employers could offer appropriate training to promote internal advancement and minimize external recruitment.

LinkedIn is a professional networking site that enables its members to post resumes and job listings as well as connect with other members. LinkedIn does not own the content and information members submit or post to LinkedIn's website; rather, per LinkedIn's User Agreement, members own their content and information and grant LinkedIn a nonexclusive license to "use, copy, modify, distribute, publish, and process" the information. LinkedIn's User Agreement also prohibits members from scraping or copying data from other member profiles by manual or automated means.

LinkedIn's members can choose from a number of privacy settings and can specify which portions of their profile are visible to the general public (i.e., to members and nonmembers), which portions are visible to all LinkedIn members, and which portions are only visible to direct connections in the member's network. The data at issue in the case was only the information that was made visible to the general public.

The Ninth Circuit noted that LinkedIn institutes numerous tools to protect the data on its website from activity it considers to be misuse or misappropriation. LinkedIn provides instructions in its robots.txt file to prohibit access to LinkedIn servers via automated bots, except for certain entities such as the Google search engine, which has express permission from LinkedIn for bot access. LinkedIn also has systems in place to detect nonhuman activity indicative of web scraping; to slow, limit or block activity from suspicious IP addresses; and to generate a list of known "bad" IP addresses serving as large-scale scrapers. LinkedIn blocks approximately 95 million automated attempts to scrape data every day and has restricted over 11 million accounts suspected of violating its User Agreement through scraping.

LinkedIn was aware of hiQ's use of automated web scraping of LinkedIn's publicly available data at least as early as 2015. LinkedIn representatives attended conferences that were hosted by hiQ in 2015 and 2016 in which hiQ's business model, including the data that was used in its algorithms, was shared and discussed.

In 2017, LinkedIn began exploring ways to monetize the large amounts of data contained in member profiles, and the company launched its own data analytics product in June of that year. A month before the launch, LinkedIn sent hiQ a cease-and-desist letter asserting that hiQ was in violation of LinkedIn's User Agreement and demanded that hiQ stop accessing and copying data from LinkedIn's server. The letter also stated that if hiQ accessed LinkedIn's data in the future, it would be violating state and federal law, including the CFAA, the Digital Millennium Copyright Act (DMCA), California Penal Code Section 502(c) and the California common law of trespass. The letter further stated that LinkedIn had "implemented technical measures to prevent hiQ from accessing and assisting others to access, LinkedIn's site through systems that detect, monitor and block scraping activity."

After receiving the letter, hiQ filed an action seeking injunctive relief based on California law and a declaratory judgment that LinkedIn could not lawfully invoke against hiQ the CFAA, the DMCA, California Penal Code Section

502(c) or the common law of trespass. The company also filed a request for a temporary restraining order, which was converted into a motion for a preliminary injunction.

The district court granted hiQ's motion and ordered LinkedIn to withdraw its letter and remove any existing technical barriers to hiQ's access of public profiles, and to refrain from putting in place any measures that would block hiQ's access to public profiles. LinkedIn appealed.

The Ninth Circuit upheld the preliminary injunction, and LinkedIn filed for a petition for writ of certiorari to the Supreme Court. The Supreme Court granted the petition, vacated the Ninth Circuit's judgment, and remanded for further consideration in view of Van Buren v. United States, which addressed the "exceeds authorized access" clause of Section 1030(a)(2) of the CFAA.

The Ninth Circuit's Analysis

On remand, the Ninth Circuit went through the preliminary injunction factors: 1) that plaintiff needs to establish that they are likely to succeed on the merits, 2) that plaintiff is likely to suffer irreparable harm absent the preliminary relief, 3) that the balance of equities tips in plaintiff's favor and 4) that an injunction is in the public's interest.

As to the second factor, given that hiQ's whole business model was dependent on LinkedIn's public profile data, the Ninth Circuit found that the district court did not abuse its discretion in finding that hiQ demonstrated that it had a likelihood of irreparable harm if the preliminary injunction was not granted. The Ninth Circuit did not find persuasive LinkedIn's arguments that hiQ could use alternative sources such as employee surveys to obtain the information it gets from LinkedIn's public profile data.

The Ninth Circuit also found the balance of equities to be in hiQ's favor. The court found hiQ's interest in staying in business was stronger than LinkedIn's alleged interest in maintaining some privacy with respect to its users' public data. The court discounted LinkedIn's argument that LinkedIn will be harmed by "free riders" who use the profiles for commercial purposes in view of the fact that members chose to make certain information public and because LinkedIn had no protected property interest in its members' data, since members maintained ownership of the data per LinkedIn's User Agreement.

The Ninth Circuit next considered the likelihood of hiQ succeeding on the merits on the specific issues presented before it. On appeal, hiQ's claim for preliminary injunctive relief was considered only on the basis of its claim of intentional interference with contract or unfair competition under California's Unfair Competition Law. Likewise, the court only considered LinkedIn's affirmative defense under the CFAA.

After finding that hiQ made a sufficient showing of its likelihood to succeed on the tortious interference claim, the Ninth Circuitconsidered LinkedIn's affirmative defense under CFAA, which, if it applied, would preempt all of hiQ's state law causes of action.

According to the Ninth Circuit, "the pivotal CFAA question here is whether once hiQ received the cease-and-desist letter, any further scraping and use of LinkedIn's data was 'without authorization' within the meaning of the CFAA and thus a violation of the statute." If so, LinkedIn asserted, hiQ would have no legal right of access to LinkedIn's data and so could not succeed on any of its state law claims, including tortious interference with contract claims.

In evaluating the CFAA, the Ninth Circuit analyzed the language of the statute, its prior interpretation of the statute, legislative history and the Supreme Court's decision in Van Buren. The Ninth Circuit ultimately determined that hiQ raised serious questions about whether LinkedIn may invoke the CFAA, finding that:

CFAA's prohibition on accessing a computer "without authorization" is violated when a person circumvents a computer's generally applicable rules regarding access permissions, such as username and password requirements, to gain access to a computer. However, when a computer network generally permits public access to its data, a user's accessing that publicly available data will likely not constitute access without authorization under the CFAA. The data hiQ seeks to access is not owned by LinkedIn and has not been demarcated by LinkedIn as private using such an authorization system. HiQ has therefore raised serious questions about whether LinkedIn may invoke the CFAA to preempt hiQ's possibly meritorious tortious interference claim.

Finally, the Ninth Circuit found that the public's interest also weighed in hiQ's favor. LinkedIn argued that the preliminary injunction is against the public interest because it will invite malicious actors to access and attack LinkedIn's computers and servers, which in turn will force LinkedIn and companies like it to choose between leaving their servers vulnerable to such attacks and protecting their websites with passwords, causing them to be cut off from public view. Although the court acknowledged that there is a significant public interest in LinkedIn's position, it found that the district court properly determined that, on balance, the public interest favors hiQ's position:

> *We agree with the district court that giving companies like LinkedIn free rein to decide, on any basis, who can collect and use data—data that the companies do not own, that they otherwise make publicly available to viewers, and that the companies themselves collect and use—risks the possible creation of information monopolies that would disserve the public interest.*

Key Takeaways

- Publicly available information (i.e., information that can be accessed without payment or logging into or creating a password protected account) may be susceptible to legal web scraping.
- The Ninth Circuit's distinction between public and privately owned data will need to be reevaluated. Companies may think about whether they want to give users more education and more control over what is made public.
- Other restrictions may apply. Companies that use or rely on web scraping to obtain data should consider whether federal IP laws or state laws restrict their ability to use data scraped from other websites, even if the CFAA does not provide a barrier to doing so.

Related Professional

Melaina D. Jobs . . . . mjobs@loeb.com

This is a publication of Loeb & Loeb and is intended to provide information on recent legal developments. This publication does not create or continue an attorney client relationship nor should it be construed as legal advice or an opinion on specific situations.

6978 REV1 05-26-2022