Data scraping and the implications of the latest LinkedIn-hiQ court ruling

Sep 20, 2019 Save This ()

Rita Heimes, CIPP/E, CIPP/US, CIPM
IAPP Staff Contributor

Information posted to social networks that are publicly accessible may be scraped and aggregated by third parties regardless of the social media sites' terms and conditions or even technical means taken to prevent data mining, according to the U.S. Court of Appeals for the 9th Circuit.

In an opinion (http://cdn.ca9.uscourts.gov/datastore/opinions/2019/09/09/17-16783.pdf) issued Monday, Sept. 9, the federal appellate court affirmed a lower court's opinion granting a preliminary injunction against professional social networking site LinkedIn that prevents the company from blocking access to automated bots deployed by hiQ, a data aggregator.

The case opens opportunities for companies that collect personal data from sites that do not restrict public access, discourages claims that social media sites have property rights in their users' data, and limits the scope of the Computer Fraud and Abuse Act as a means for preventing automated bots from scraping publicly visible data.

Although the case is only at the preliminary injunction stage and has yet to go to trial on the merits, the opinion has significant implications for the personal data marketplace.

Background

LinkedIn allows people to create professional profiles, post articles and comments, search for jobs, and connect to others using the site to grow their professional networks. Members may specify which portions of their profiles are visible to the general public and which are limited to their network, as well as switch on a "Do Not Broadcast" option that prevents notification to their network when their profiles are updated.

In addition to prohibiting data scraping or copying in its User Agreement, LinkedIn works to prevent access to its servers by unauthorized automated bots and uses other technical systems to detect non-human activity indicative of scraping and to block suspicious or disfavored IP addresses.

Among the users LinkedIn has terminated for allegedly violating its User Agreement is hiQ Labs, a data analytics company founded in 2012 that scrapes information from LinkedIn users' public profiles (including name, job title, work history and skills) and sells that information to business clients, such as eBay, Capital One and GoDaddy. HiQ's analytics are designed to identify employees at risk of being recruited away or identify skills gaps in employers' workforces so they can offer internal training and mobility.

In 2017, LinkedIn sent a cease-and-desist letter to hiQ asserting that hiQ's use of scraping bots violated LinkedIn's User Agreement and the CFAA, among other laws. HiQ responded by filing a lawsuit seeking a declaration that it was not violating any law and an injunction preventing LinkedIn for blocking its access to users' data.

The district court granted hiQ's motion and ordered LinkedIn to remove any technical barriers to hiQ's access to public profile information. LinkedIn filed an appeal.

Limitations on LinkedIn's ownership and control of users' data

The 9th Circuit found that hiQ's business model depended on access to LinkedIn's publicly accessible data and rejected LinkedIn's arguments that hiQ could gather workforce data from other means. It also rejected LinkedIn's arguments that allowing hiQ to scrape LinkedIn's site threatened its users' privacy and put at risk LinkedIn's goodwill with its members.

The court found "doubtful" that LinkedIn users had any expectation of privacy with respect to information they post publicly, even pointing to LinkedIn's own privacy policy that notifies users that their information can be seen by others. The court also found that LinkedIn's own products (enabling subscribers to get alerts and export data for recruiting and marketing purposes) further undercut its arguments about user privacy.

Regarding LinkedIn's economic interests — avoiding competition from third parties that also want to profit from selling its users' data — the court found that LinkedIn "has no protected property interest in the data contributed by its users, as the users retain ownership over their profiles." Users, moreover, "quite evidently" intend their profile data to be accessed by others, "including for commercial purposes."

HiQ's lawsuit against LinkedIn claimed that, by blocking hiQ's access to LinkedIn data, LinkedIn had tortiously interfered with hiQ's contractual relationships with third parties and thereby harmed hiQ's business. Rather than challenging the facts underlying this claim, LinkedIn instead sought (at this stage of the litigation) to justify its actions on the basis of its "legitimate business interests." Although this issue will be further explored at trial, the 9th Circuit struggled to find LinkedIn's technical practices for blocking hiQ's scraping activity were "recognized trade practices" — they were not "similar to trade practices heretofore recognized as acceptable justifications for contractual interference."

Indeed, the court held, "if companies like LinkedIn, whose servers hold vast amounts of public data, are permitted selectively to ban only potential competitors from accessing and using that otherwise public data, the result — complete exclusion of the original innovator in aggregating and analyzing the public information — may well be considered unfair competition under California law."

> In sum, this case supports the business model of scraping personal data from websites that allow personal data to be visible to the general public, regardless of the websites' terms of use or other efforts a business may have to control access to or monetize for its own purposes that data.

Finally, the court rejected LinkedIn's arguments that it was protecting its members' data and enforcing its User Agreement, emphasizing again that LinkedIn "has only a non-exclusive license to the data shared on its platform, not an ownership interest." The court identified LinkedIn's core business model as providing a platform for professionals to share their information with each other that could continue to exist even if third parties use that information for commercial gain. The fact that LinkedIn had developed its own data analytics tool to generate revenue from its users' data only served to support the court's position that LinkedIn didn't have "its members' privacy interests in mind."

control access to or monetize for its own purposes that data.

CFAA not applicable, enacted to prevent "hacking"

The CFAA forbids access to protected computers "without authorization" or in a manner that exceeds authorization. The 9th Circuit held that LinkedIn's servers were "protected computers." Key to deciding this case was the court's interpretation of the term "authorization." In short, the 9th Circuit was not willing to find that hiQ's scraping bots were unauthorized or that they exceeded authorized access.

LinkedIn attempted to prevent hiQ's automated systems from gaining access to users' personal data by deploying blocking technology and by communicating — at least through a cease-and-desist letter — that it disapproved of hiQ's practices.

The court focused on the public accessibility of LinkedIn's site in general. Because LinkedIn's site is accessible to anyone who visits the site, it is by default freely accessible — everyone is authorized, the court noted. Thus, denial of access is a ban, not withdrawal of authorization. By contrast, the court found that the phrase "access … without authorization" in the CFAA applies more appropriately when permission is typically required of everyone, when access is generally restricted only to those "specially recognized or admitted."

> This case has significant implications for privacy. It sets a precedent that data entered by users to a social media website does not belong to (but rather is merely licensed to) the site owner.

To justify its reasoning, the court opined that the CFAA is a "computer hacking" law, citing legislative history that references the prohibited conduct of "breaking and entering." Accordingly, the court found that informing someone via contract (or notice) that their conduct on the site isn't welcome or permissible does not, should they access the site anyway, constitute conduct "without authorization." It even cited a law review article by Prof. Orin Kerr that an "authentication requirement, such as a password gate, is needed to create the necessary barrier that divides open spaces from closed spaces on the web" to support the notion that authorization for data scraping is only required when sites are password protected or otherwise not visible to the public.

Finally, the court interpreted the CFAA to divide the information universe into three categories:

- Information for which access is open to the general public and permission is not required.
- Information for which authorization is required and has been given.
- Information for which authorization is required and has not been given (or, in the case of the prohibition on exceeding authorized access, has not been given for the part of the system accessed).

Importantly, the court distinguished the Facebook, Inc. v. Power Ventures, Inc. case, in which Facebook had successfully prevented a social networking aggregator from accessing Facebook users' data, on the grounds that Facebook had "tried to limit and control access to its website" and required "its users to register with a unique username and password." Here, the court found, hiQ was scraping data from LinkedIn that "was available to anyone with a web browser."

Conclusion

This case has significant implications for privacy. It sets a precedent that data entered by users to a social media website does not belong to (but rather is merely licensed to) the site owner. It undermines the significance of user agreements to set the terms for non-users who might, in contradiction to the agreements' terms, collect and use data made available on those sites. It also narrows the definition of "authorization" in the context of websites that collect and host personal data to those sites that require usernames and passwords and increases the responsibility on such websites to inform users of the benefits of privacy settings.

The case now returns to the district court for a trial on the merits, provided there is no settlement. But lawyers, privacy professionals and privacy scholars will no doubt be pouring over the implications of this case — including the novel interpretation of the CFAA — for months and years to come.

Photo by Kevin Ku (https://unsplash.com/@ikukevk?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) on Unsplash (https://unsplash.com/s/photos/tech?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText)

Pease International Tradeport, 75 Rochester Ave.
Portsmouth, NH 03801 USA • +1 603.427.9200