Articles + Publications | May 5, 2022

Ninth Circuit Provides Guidance on Web Scraping

WRITTEN BY

Sadia Mirza | Peter T. Wakiyama | Mary C. Zinsner | Ronald Raether, Jr. | Stephen C. Piepgrass | Robyn W. Lin | Abbey Thornhill | Edgar Vargas

On April 18, the Ninth Circuit issued its opinion in *hiQ Labs, Inc. v. LinkedIn Corporation*[1] in which the court clarified its position on an important topic: whether the common practice of data "web scraping" can create criminal liability under the Computer Fraud and Abuse Act (CFAA). To be clear, the Ninth Circuit was not afforded the opportunity to directly rule on the question of whether web scraping may violate the CFAA, but in reaffirming a district court's grant of a preliminary injunction, the Ninth Circuit strongly indicated that it does not believe the commonly used method of fetching and extracting data from websites violates federal law — even if a website has stated in its terms of service or otherwise that such activity is prohibited. Importantly, the court did not address any other potential claims that might arise from web scraping, such as intellectual property infringement claims.

I. The Computer Fraud and Abuse Act and Accessing Computers "Without Authorization"

Enacted in 1986, the CFAA amended the first federal computer fraud law, with the intent to address and criminalize hacking. The plain language of the law states, "[w]hoever … 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.[2] The statute defines "protected computer" broadly, allowing courts to hold that the statute applies to effectively any computer connected to the internet.[3] As websites are stored on servers — "computers that manage network resources and provide data to other computers"[4] — liability under the CFAA can arise where individuals access a website "without authorization" or in a manner that "exceeds authorized access."

For years, courts have debated the meaning of "without authorization" under the CFAA. The question centers on whether "the CFAA is best understood as an anti-intrusion statute" or a "misappropriation statute."[5] Is liability limited to traditional hackers who infiltrate a computer's security systems to gain access to information unavailable to the public? Or can "without authorization" apply to circumstances where someone has been granted access to a computer but then uses the computer in a manner not "authorized" by the owner's rules? For example, if an employee is given a laptop for "work purposes only," does the user access the computer "without authorization" when she browses a website for a new pair of shoes or tunes into a basketball game in the middle of March Madness? The latter interpretation may seem absurd, but in relying on the plain text of Section 1030(a)(2)(C), many circuits have held that even violation of an employer's internal rules or a website's terms of service, user agreements, confidentiality agreements, or other similar contractual terms could constitute a violation of the CFAA.[6]

In *hiQ Labs v. LinkedIn*, the Ninth Circuit faced a new variation of this question. HiQ is a data analytics company

that “scrapes” information from LinkedIn, a web-based social media service, for information included on public LinkedIn profiles. The company collects information like names, job titles, work history, and skills and filters that information through a predictive algorithm to sell clients information regarding, for example, which employees may be most likely to take a new position or what skills their workforce seems to be lacking.

In 2017, LinkedIn sent hiQ a cease-and-desist letter, asserting that hiQ violated LinkedIn’s user agreement and demanded that hiQ stop accessing and copying data from LinkedIn’s servers. LinkedIn’s letter further stated that continued access by hiQ would violate state and federal law, including California Penal Code 502(c), the CFAA, common law of trespass, and the Digital Millennium Copyright Act. LinkedIn also claimed that the company’s continued access to its information would violate the CFAA and announced its intention to implement certain technical measures to prevent hiQ from accessing LinkedIn’s website. HiQ then moved for a preliminary injunction against LinkedIn, asking a district court to enjoin LinkedIn’s implementation of the protective technical measures on the grounds that the conduct constituted a tortious interference with its contracts with its paying customers.[7] To hold in favor of hiQ, the district court had to find that hiQ was “likely to succeed on the merits” on its claim against LinkedIn.[8] This necessarily involved the court’s consideration of whether hiQ’s conduct violated the CFAA. If so, hiQ’s tortious interference contract claim must fail: hiQ could not base a lawsuit on LinkedIn’s decision to block it from conducting an *illegal* activity.

Upon remand from the U.S. Supreme Court,[9] the Ninth Circuit issued a lengthy and detailed opinion, upholding the district court’s decision granting the injunction and found that hiQ had made the required showing by raising at least a “serious question”[10] as to whether web scraping constitutes unauthorized access of a website in violation of the CFAA.[11] In doing so, the Ninth Circuit reaffirmed its position on the side of the circuit split that considers CFAA to be an anti-intrusion statute that prohibits “breaking and entering” or what is traditionally thought of as “hacking.” In the court’s eyes, the operative question under the CFAA is whether the computer’s gates are up or down.[12] Unauthorized access occurs when an individual breaks past the gates when they are “up.” Public websites like LinkedIn, however, have no gates to data accessibility, and in such context, the CFAA simply does not apply.[13] As the court concluded:

[I]t appears that the 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. It is likely that when a computer network generally permits public access to its data, a user’s accessing that publicly available data will not constitute access without authorization under the CFAA.[14]

Again, to be clear, in reaching this conclusion, the Ninth Circuit did *not* officially declare web scraping to be legal, and LinkedIn has stated it will continue to litigate the merits of this case. The court’s opinion, however, is certainly notable and provides a roadmap for those evaluating compliance with the CFAA. If and when another court is afforded the opportunity to rule on this question, surely the Ninth Circuit’s well-reasoned and detailed opinion will provide crucial guidance and carry great persuasive weight.

II. Takeaways

Private Vs Public Ownership

In its decision, the Ninth Circuit made an important distinction between information that is "private" and information that is "public." According to the court, the legislative history of Section 1030 of the CFAA "makes clear that the prohibition on unauthorized access is properly understood to apply only to private information."[15] By this, the court did not mean privately *owned* information — the question is not whether the information is owned by a private individual or business or by the government.[16] Instead, the question is whether a website has erected any gates. Information may be "delineated as private" and accordingly protected by the CFAA "through use of a permission requirement of some sort."[17]

Presumably, this means that any private individual or government entity can erect technological barriers to protect its information from access by certain individuals, including web scrapers. Of course, government agencies are subject to the Freedom of Information Act (FOIA) and state equivalents, meaning there are limits on the types of information the government may keep confidential.[18] But the government, just like any other entity, may create websites with gates that prevent web scrapers from fetching and aggregating data by closing such information off from the general public with password protections or other technological restrictions.

Issues Still Remaining

While it does offer insight into the Ninth Circuit's perspective, the *hiQ Labs* decision also leaves open a number of important questions. For example, the court did not address the question of whether, or the extent to which, government bodies may erect barriers to access otherwise public information. The court also did not consider whether if a government or private entity seeks to restrict access to information — by password or through "softer" restrictions like implementation of a CAPTCHA meant to prevent easy access by bots — the development of workarounds to facilitate scraping of the information sought to be restricted would be a violation of CFAA. Nor did the court address whether the answer to this second question would be different depending on if the entity seeking to restrict access is a government body or a private business, particularly given the well-established public interest in access to government records.

Beyond the issues presented by technological gates and barriers, other statutes may determine the permissibility of access to data and permissibility of blocking web scraping. For example, intellectual property statutes, such as the Copyright Act of 1976,[19] the Lanham Act[20], state and federal trade secret laws, and state unfair competition laws, can provide protection to information. Copyright protects original works of creative expression fixed in a tangible medium of expression. This can include websites, mobile applications, compilations of data, and the software that enables websites and the website functionality used to query, process, and display data. Meanwhile, trademarks protect the identification of a good or service and can provide legal protection for brands. Both the Copyright Act and the Lanham Act provide litigants with a private right of action through which individuals or businesses may seek injunctive relief and/or monetary damages for infringement of their intellectual property rights. Additionally, copyright registrations for works of authorship may provide significant benefits to the registrant in litigation, including the ability to recover attorneys' fees and seek statutory damages. Therefore, intellectual property protection for data assets should be carefully considered by any web-based business that permits access to its data or engages in web scraping.

What *hiQ* Means for Privacy Litigation

HiQ is not the only company that "scrapes" data as a business model. Another example is Clearview AI, a facial

recognition company that has scraped billions of pictures from the internet to create a massive facial recognition database. Clearview AI's database includes more than three billion photos of individuals — all scraped from social media sites, including Facebook, Twitter, Instagram, and Google. Clearview AI contracts to provide facial recognition services for law enforcement and other agencies.

While the Ninth Circuit's holding in *hiQ* indicates that Clearview AI and other companies that have adopted scraping data as a business model can avoid criminal liability under CFAA, web scrapers may still face legal challenges under state law regimes and intellectual property laws. For example, Clearview AI currently faces litigation in Illinois under the state's Biometric Privacy Information Act (BIPA). Web scrapers may also see legal challenges based on state unfair competition and unfair and deceptive acts and practices (UDAAP) statutes, which have been adopted in various forms in all 50 states. The remedies available under state UDAAP laws vary but often include injunctive relief, as well as civil penalties and compensatory damages and attorneys' fees. Further heightening the stakes, state attorneys general also can enforce these statutes.

Conclusion

While the Ninth Circuit's decision in *hiQ* is significant, the fight over the legality of web scraping is far from over. The *hiQ* decision may provide a roadmap for where the Ninth Circuit and some other courts are headed on the issue of its legality under the CFAA, but there remain courts that have treated the CFAA as a misappropriation statute. Those courts may be more willing to side with companies like LinkedIn that hope to block the practice when they violate the terms of service or other contractual agreements. Further, even if all courts follow the Ninth Circuit, data scraping may still face legal challenges under other state and federal statutes.

For companies and government agencies looking to prevent data scraping on their websites, the *hiQ* decision suggests they should consider launching websites with technological barriers to entry, such as passwords or paywalls, that prevent general public access to their information. Additionally, if companies believe they may have information protected under copyright or trademark law, they should discuss with counsel the possibility of registering their intellectual property for added protection in addition to using appropriate intellectual property markings, such as copyright and trademark notices.

For companies using web scraping methods, the *hiQ* decision suggests they may be able to avoid liability under the CFAA for their actions. Web scrapers should remain aware of other potential legal challenges, which could result in significant monetary penalties or injunctive constraints.

If you have any questions, please contact Stephen Piepgrass (Regulatory, Strategy + Enforcement), Ronald Raether, Jr. (Cybersecurity, Information Governance + Privacy), Peter Wakiyama (IP, Technology, Data Privacy and Security), Mary Zinsner (Consumer Financial Services), Robyn Lin (Cybersecurity, Information Governance + Privacy), Sadia Mirza (Cybersecurity, Information Governance + Privacy), Abbey Thornhill (Regulatory, Strategy + Enforcement), and Edgar Vargas (Cybersecurity, Information Governance + Privacy).

---

[1] — F.4th –, 2022 WL 1132814 (9th Cir. Apr. 18, 2022).

[2] 18 U.S.C. § 1030(a)(2)(C).

[3] *See* 18 U.S.C. § 1030(e)(2)(B) (defining protected computer as any "used in or affecting interstate or foreign commerce"); *United States v. Nosal* (Nosal II), 844 F.3d 1024, 1050 (9th Cir. 2016), cert. denied 138 S. Ct. 314 (2017).

[4] *HiQ Labs*, 2022 WL 1132814, at *11.

[5] *Id.* at *13 (citing *United States v. Nosal* (Nosal I), 676 F.3d 854, 857 – 58 (9th Cir. 2012)).

[6] *See, e.g.*, *EF Cultural Travel BV v. Explorica, Inc.*, 274 F. 3d 577, 583 – 84 (1st Cir. 2001); *United States v. Rodriguez*, 638 F.3d 1258, 1263 (11th Cir. 2010).

[7] *Id.* at *21.

[8] "'A plaintiff seeking a preliminary injunction must establish that he is likely to succeed on the merits, that he is likely to suffer irreparable harm in the absence of preliminary relief, that the balances of equities tips in his favor, and that an injunction is in the public interest.'" *Id.* at *5 (*Winter v. Nat. Res. Def. Council, Inc.*, 555 U.S. 7, 20 (2008)).

[9] The district court granted the injunction, *hiQ Labs, Inc. v. LinkedIn Corp.*, 273 F. Supp. 3d 1099 (N.D. Cal. 2017), and the Ninth Circuit affirmed, *hiQ Labs, Inc. v. LinkedIn, Corp.*, 938 F.3d 985 (9th Cir. 2019). LinkedIn then filed a petition for writ of certiorari to the Supreme Court in March 2020, and the Supreme Court granted the petition but immediately sent the case back to the Ninth Circuit for further consideration in light of the Supreme Court's decision in *Van Buren v. United States*, 141 S. Ct. 1648 (2021). *See LinkedIn Corp. v. hiQ Labs, Inc.*, 141 S. Ct. 2752 (2021). In *Van Buren*, the Supreme Court held that a police officer who used his valid credentials to run a license plate in exchange for money in violation of department policy did *not* violate the CFAA.

[10] The Ninth Circuit found that hiQ had made a strong showing of irreparable harm and that, accordingly, hiQ "need demonstrate only 'serious questions going to the merits.'" *Id.*

[11] *Id.* at *14.

[12] *Id.*

[13] *Id.* For this reason, whether a tortious interference claim could be brought by hiQ if it had affirmatively agreed not to use data obtained from LinkedIn violated that agreement, and LinkedIn then sought to cut hiQ off from its data is a question not addressed in this analysis.

[14] *Id.* at *16.

[15] *Id.* at *13.

[16] Whether the government "owns" public records or simply holds them on behalf of the public is a separate question outside the scope of this article.

[17] *Id.*

[18] Examples of information that must be made public would include most court records, property records, records of public bodies not held in closed session, and law enforcement reports once investigations have concluded. Of course, government records containing personal identifying information and other details regarding minors may be redacted or withheld.

[19] 17 U.S.C. § 101 et seq.

[20] 15 U.S.C. § 1051 et seq.

RELATED INDUSTRIES + PRACTICES

- Public Records/FOIA
- State Attorneys General