CFAA Decision May Raise Bar On Scraping Liability

By **James Snell, Nicola Menaldo and Ariel Glickman** (August 7, 2020)

James Snell

Nicola Menaldo

Ariel Glickman

As scraping and crawling of websites becomes more ubiquitous, courts continue to struggle with where to draw the lines regarding what is permissible. This can be a highly fact-intensive inquiry, but a recent case before the U.S. District Court for the District of Columbia provides some important takeaways.

In Sandvig v. Barr, a case involving public interest bots and web crawlers testing for discrimination online, the court held that the Computer Fraud and Abuse Act does not prohibit scraping publicly accessible portions of a website, even when doing so violates the website's terms of service.[1] In issuing this ruling, the court avoided the First Amendment question, raised in this case and others, regarding the constitutional limits on laws that purport to prohibit capturing data from publicly accessible websites.

Sandvig resolves a 2016 complaint brought by the American Civil Liberties Union on behalf of academic researchers and a journalism group that wanted to scrape content and create fake accounts using automated programs (i.e., bots) to investigate websites' business practices, including discrimination on real estate and employment websites.[2] These plaintiffs argued that because the CFAA prohibits anyone from "exceeding authorized access" — which some courts have interpreted as violating a website's terms of service — the researchers and journalism group would risk prosecution were they to engage in such activities.[3]

This decision follows on the heels of the U.S. Court of Appeals for the Ninth Circuit's ruling in hiQ Labs Inc. v. LinkedIn Corp., issued in September 2019, which held that scraping public information without bypassing any "permission requirement" (e.g., a password) does not violate the CFAA.[4] And it suggests that courts may be starting to look more favorably at scraping cases, especially where the fact pattern includes an example of so-called white-hat scraping for the public good.

Below we discuss the hiQ Labs and Sandvig decisions and the impact they may have on scraping law in the coming years.

The hiQ Labs Decision

In hiQ Labs, the Ninth Circuit evaluated, in connection with a challenge to a preliminary injunction, whether defendant social media company LinkedIn could likely prevail with a CFAA claim against a competitor, hiQ, for obtaining and using publicly accessible information from the LinkedIn website. HiQ wanted to scrape public profiles that LinkedIn users had shared and that anyone with a web browser could view.

The Ninth Circuit concluded that LinkedIn would be unlikely to prevent hiQ from obtaining this publicly available information under the CFAA and allowed hiQ's claim of tortious interference with contract against LinkedIn to proceed, subject to all of LinkedIn's defenses.

Importantly, the court was not persuaded that hiQ had obtained information from LinkedIn's

website "without authorization" within the meaning of the CFAA simply because hiQ had scraped the website after receiving a cease-and-desist letter from LinkedIn. Instead, it held that the CFAA's use of the term "without authorization" applied only where "access is not generally available."[5] It determined that permission was not required (and therefore could not be revoked) where the information was accessible to anyone with an internet browser, as opposed to requiring use of a password. Only the latter, according to the court, required permission.[6]

The Sandvig Decision

The Sandvig court reached a similar conclusion. There, academic researchers and one journalism group wished to engage in audit testing, including through web scraping, to investigate websites' compliance with housing, employment, and civil rights laws, and to perform an assessment of business practices. They intended to "create profiles for fictitious job seekers, post fictitious job opportunities, and compare their fictitious users' rankings in a list of candidates for the fictitious jobs," while alerting actual job seekers on the platform that the job seeker is not real, and the postings are fake.[7]

Plaintiffs brought the lawsuit against the government, seeking to enjoin enforcement of the CFAA as a violation of the First Amendment as well as the due process clause of the Fifth Amendment to the U.S. Constitution.[8] The CFAA's "exceeds authorized access" provision in Title 18 of the U.S. Code, Section 1030(a)(2)(C), at issue in the complaint, prohibits (1) "intentionally ... exceed[ing] authorized access" and (2) thereafter obtaining information from a "protected computer."

On March 27, the court dismissed the remaining plaintiffs' and defendant's motions for summary judgment, holding that the CFAA did not criminalize plaintiffs' planned scraping, and thus avoiding the First Amendment question. In reaching this decision, the D.C. district court adopted the parameters set forth in hiQ Labs: Content obtained from public websites without use of a "permission requirement" (like a password) does not trigger criminal liability under the CFAA, whereas content obtained through use of a "permission requirement" may constitute "access ... without authorization" under the statute.[9]

Then, the court analyzed whether terms of service are "permission requirements" such that violating them would constitute accessing a computer without authorization or in "exce[ss] [of] authoriz[ation]" under the CFAA.

First, the court held that terms of service are not permission requirements. It looked at three factors to reach this conclusion: (1) notice, (2) the nondelegation doctrine, and (3) the rule of lenity and canon of avoidance.

As to notice, the court deemed a website's terms of service inadequate in notifying users for purposes of criminal liability, as they are often lengthy and difficult to understand, and can be changed.[10] Users are also not always required to view the terms before using a website, and the terms may be in fine print or only referenced in a link at the bottom of the website, which, the court stated, is not significant enough to allow for criminal liability.[11]

As to the nondelegation doctrine, the court determined that the CFAA does not allow website owners to define the scope of criminal liability, which would turn websites into their "own criminal jurisdiction and each webmaster its own legislature."[12] Finally, addressing the rule of lenity and canon of avoidance, the court found that because neither the statutory history nor legislative history provided a satisfactory definition of "access[ing] a computer without authorization," it was compelled to narrowly interpret the phrase as not including

terms violations.[13]

Second, the court reasoned that to "exceed[] authorized access" means to first have passed through a permission requirement.[14] Since terms of service are not permission requirements, it followed that violating a website's terms of service does not constitute "exceed[ing] authorized access" under the CFAA.[15]

Key Takeaways

The Sandvig decision adds to the authority that would limit the bases for which website owners can police access to their platforms under the CFAA. It also reflects what may be a trend among courts to favor scrapers over website owners under certain factual circumstances, at least as to publicly accessible content.[16]

But while Sandvig may represent additional authority that makes it difficult for website owners to raise CFAA claims for scraping of public content, it does not foreclose the possibility of website owners raising CFAA claims involving terms violations. For instance, in some cases, terms of service may be claimed to qualify as "permission requirements" if, for example, an individual accesses a password-protected website using another's credentials and the terms prohibit doing so. This may be alleged to be true even if the credentials were freely given to the individual using them by the credential owner. Neither hiQ Labs nor Sandvig considered this scenario.

Additionally, even if a website owner faces more difficulty in claiming CFAA liability based on terms-of-use violations, scraping can still lead to claims under various other theories. Other types of claims include:

- Breach of contract: If scraping would violate an agreement between the website owner and the scraper, the website owner could raise this claim as a violation of that contract. Although Sandvig holds that terms-of-service violations are not permission requirements that can lead to CFAA liability, the decision does not prevent website owners from pursuing claims for breach of contract under those terms.

- Copyright infringement: Website owners may claim copyright infringement against scrapers that copy photographs, original text and other copyrightable material from their sites. These types of claims may be subject to a fair use analysis.[17]

- Trespass to chattels: Website owners may claim trespass to chattels if automated access to their website constitutes an "intentional interference with the possession of personal property ... [that] proximately cause[s] injury."[18] To raise this claim, websites need to show how the access negatively impacts their use of the platform, such as by affecting the utility of their servers.

- Hot news misappropriation: There is some precedent for website owners raising claims under this tort where the cause of action exists in Illinois, Pennsylvania, New York, California and Missouri. This claim can apply where a scraper has reproduced

both factual and time-sensitive information that was obtained through the effort of the website owner and thus republication of the information was at its expense.[19] This claim is often difficult to make if the republication of the information includes unique analysis and not mere repackaging of the information.

Given the prevalence of scraping, we can expect additional cases to be filed in the near future, requiring courts to grapple with efforts to prevent unauthorized automated access on websites.

James Snell and Nicola Menaldo are partners, and Ariel Glickman is an associate, at Perkins Coie LLP.

[1] See Sandvig v. Barr, No. 16-1368 (JDB) (D.D.C.) (Dkt. No. 67).

[2] Compl. ¶¶ 7, 75–132.

[3] Id. ¶ 4.

[4] hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985 (9th Cir. 2019).

[5] Id. at 999–1000.

[6] See id. at 1000–01 (citing H.R. Rep. No. 98-894, at 20).

[7] Sandvig v. Barr Mem. Op. at 2.

[8] Sandvig v. Barr Compl. ¶¶ 9, 21.

[9] Sandvig v. Barr Mem. Op. at 19–20.

[10] Id. at 20.

[11] Id.

[12] Id. at 20–21.

[13] Id. at 22.

[14] Id. at 24.

[15] Id.

[16] Jim Snell and Derek Care, Use of Online Data in the Big Data Era: Legal Issues Raised By the Use of Web Crawling and Scraping Tools for Analytics Purposes, Bloomberg Law (Dec. 24, 2013, 12:00 AM), https://news.bloomberglaw.com/us-law-week/use-of-online-

data-in-the-big-data-era-legal-issues-raised-by-the-use-of-web-crawling-and-scraping-tools-for-analytics-purposes; James Snell and Nicola Menaldo, Web Scraping in an Era of Big Data 2.0, Bloomberg Law (June 1, 2016, 7:40 PM), https://news.bloomberglaw.com/tech-and-telecom-law/web-scraping-in-an-era-of-big-data-20.

[17] See Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1154–55, 1165, 1168 (9th Cir. 2007) (holding that Google's display of low-resolution thumbnail images in its search results was a fair use of plaintiff's copyrighted photographs); Associated Press v. Meltwater U.S. Holdings, Inc., 931 F. Supp. 2d 537, 552 (S.D.N.Y. 2013) (holding that the news aggregator's "use[] [of] its computer programs to automatically capture and republish designated segments of text from news articles, without adding any commentary or insight in its News Reports" constitutes "undiluted use" of the Associated Press's copyrighted articles).

[18] eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1069 (N.D. Cal. 2000) (quoting Thrifty–Tel v. Bezenek, 46 Cal. App. 4th 1559, 1566 (1996)).

[19] See, e.g., Associated Press v. All Headline News Corp., 608 F. Supp. 2d 454, 460–61 (S.D.N.Y. 2009) (finding that Associated Press adequately stated claim for hot news misappropriation against defendant who rewrote news articles published by Associated Press and passed them off as articles reported by defendant).