Proskauer

Researchers May Challenge the Constitutionality of the CFAA "Access" Provision as Applied to Web Scraping

New Media & Technology Law Blog on **April 27, 2018**

Such Scraping "Plausibly Falls within the Ambit of the First Amendment"

The Ninth Circuit is currently considering the appeal of the landmark *hiQ* decision, where a lower court had granted an injunction that limited the applicability of the federal Computer Fraud and Abuse Act (CFAA) to the blocking of an entity engaging in commercial data scraping of a public website. While we wait for that decision, there has been another fascinating development regarding scraping, this time involving a challenge to the CFAA brought by academic researchers. In *Sandvig v. Sessions*, No. 16-1368 (D.D.C. Mar. 30, 2018), a group of professors and a media organization, which are conducting research into whether the use of algorithms by various housing and employment websites to automate decisions produces discriminatory effects, brought a constitutional challenge alleging that the potential threat of criminal prosecution under the CFAA for accessing a website "without authorization" (based upon the researchers' data scraping done in violation of the site's terms of use) violates their First Amendment rights.

In a preliminary decision, a district court held that the plaintiffs have standing and allowed their as-applied constitutional challenge to the CFAA to go forward with regard to the activity of creating fictitious accounts on web services for research purposes. The decision contains vivid language on the nature of the public internet as well as how the plaintiffs' automated collection and use of publicly available web data would not violate the CFAA's "access" provision even if a website's terms of service prohibits such automated access (at least with respect to the facts of this case, which involves academic or journalistic research as opposed to commercial or competitive activities).

The Plaintiffs' Action for Declaratory Relief

The CFAA was enacted in 1984 to enhance the government's ability to prosecute computer crimes and target hackers. The CFAA, 18 U.S.C. §1030, prohibits a number of different computer crimes, the majority of which involve accessing computers without authorization or in excess of authorization, and then taking specified forbidden actions, ranging from obtaining information to damaging a computer or computer data. The statute also provides for a civil right of action for violations, and such a claim is regularly pled by website owners against unwanted data scrapers and by employers against departing employees who access proprietary company data for improper purposes.

The plaintiffs in this action directly challenge the so-called "Access Provision," 18 U.S.C. §1030(a)(2), which provides for criminal penalties: "[w]hoever . . . intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer . . . shall be punished as provided in subsection (c) of this section." [The statute does not define the phrase "without authorization"].

To conduct their audit testing to determine if the use of algorithms and data tracking technologies results in discriminatory effects in the housing and employment sectors, the plaintiffs stated they plan on creating artificial user profiles (and even fictitious job listings) to measure how websites process users with different demographics and also deploy bots to visit a host of websites to determine how the websites display content to users profiled with varying web browsing habits. The researchers will also necessarily use automated scrapers to record the data and eventually publish their findings. They are aware that such activities will violate certain websites' terms of use, as most websites prohibit scraping and the use of bots and also the creation of false profiles (or "sock puppets"). In their suit, the plaintiffs claim that due to the threat of the CFAA, they must either refrain from conducting their website testing and academic research (which they claim are protected free speech) or else risk criminal prosecution. As a result, the plaintiffs brought an action seeking a declaratory judgment to, among other things, enjoin the enforcement of the Access Provision, as applied to them.

In finding that the plaintiffs have standing (e.g., showing a credible threat of prosecution), the court made some noteworthy findings regarding whether the plaintiffs' scraping actions are protected activity. The court found that the First Amendment generally protects the gathering and creation of information, and, as such, "scraping plausibly falls within the ambit of the First Amendment." Going further, the court added, "scraping is merely a technological advantage that makes information collection easier." The court also found that, for standing purposes, plaintiffs have a First Amendment interest in sock puppeting, or harmless misrepresenting their identities in order to audit test websites. Lastly, the court stated that the application of criminal penalties for "publishing original material that user publicly available information…triggers First Amendment scrutiny." In rejecting the Government's argument that the First Amendment does not create a right to acquire information in whatever manner one desires, the court noted that the plaintiffs are not attempting to compel private websites to provide information others cannot get, but only to prevent the government from prosecuting them from obtaining or using publicly available data that the general public can access. While the court recognized the DOJ policy of discouraging CFAA prosecutions based upon harmless terms of use violations, the court found that plaintiffs established a credible threat of prosecution for standing purposes based on past prosecutions where the Government has read the CFAA Access Provision to include terms of use violations. (See generally *United States v. Drew*, 259 F.R.D. 449 (C.D. Cal. 2009) (involving defendant charged with violating sections §§1030(a)(2)(C) and 1030(c)(2)(B)(ii) for creating a fictitious profile on a social networking website and then using the account to cyberbully a teenager in violation of the website's terms of service; court overturned the conviction and rejected the interpretation of "unauthorized access" under the CFAA for the terms of service violation in this case as void under the vagueness doctrine).

In looking at the application of the CFAA's Access Provision to the plaintiffs' activities, the court noted that the law is uncertain regarding whether violating a website's terms of service "exceeds authorized access" under the statute. Importantly, with regard to web scraping, the court declined the invitation to carve out an exception from the CFAA for harmless terms of service violations, suggesting that such a construction is untenable. After a deep examination into how the statutory terms "access without authorization" and "exceeds authorized access" have been construed by various courts, the court espoused a narrow interpretation of the Access Provision and noted the risks of enforcement: "By incorporating ToS that purport to prohibit the purposes for which one accesses a website or the uses to which one can put information obtained there, the CFAA threatens to burden a great deal of expressive activity, even on publicly accessible websites—which brings the First Amendment into play."

Ultimately, the court noted that much of the plaintiffs' proposed activities fall outside the CFAA's reach and that the CFAA "prohibits far less than the parties claim (or fear) it does."

> "Scraping or otherwise recording data from a site that is accessible to the public is merely a particular use of information that plaintiffs are entitled to see. The same goes for speaking about, or publishing documents using, publicly available data on the targeted websites. [….] Employing a bot to crawl a website or apply for jobs may run afoul of a website's ToS, but it does not constitute an access violation when the human who creates the bot is otherwise allowed to read and interact with that site. [citation omitted]"

Thus, out of all the plaintiff's proposed activities, the court held that only the researchers' plans to create fictitious user accounts on employment sites would violate the CFAA because such activities do not occur on portions of websites that anyone can view, but on pages that are limited to "those who meet the owners' chosen authentication requirements and targeted to the particular preferences of the user." At this stage, the court allowed the plaintiffs' as-applied constitutional challenge to go forward based on such potential sock puppeting activities, because, absent any evidence that the speech would be used to gain a material advantage, such false speech retains First Amendment protection and "rendering it criminal does not appear to advance the government's proffered interests."

Implications for Data Scraping

What are the implications of the *Sandvig* ruling for data scraping and the availability of a civil CFAA cause of action based upon violations of a website terms of use?

The *Sandvig* case primarily presents constitutional questions, but the opinion offers some relevant language with respect to how to interpret terms of use violations under the CFAA. Given that the plaintiffs' academic scraping activities differ greatly from the typical commercial data scraping scenario, the opinion's importance is limited (but helpful) – the decision contains some language that narrowly interprets the scope of the CFAA in the criminal context for terms of service violations for research activities, but the decision certainly falls short of giving the all-clear signal to data scraping of public websites when considering the potential for civil liability. Indeed, the opinion does not really address the hot issue before the Ninth Circuit in the *hiQ* appeal which concerns whether a public website can invoke the CFAA to block unwanted scraping activities after having expressly revoked access to its site to the unwanted user.

Some additional considerations:

- The plaintiffs' non-commercial activities for the purpose of academic research are much different than data scraping performed by an upstart competitor seeking to collect website content for data aggregation or a completely new service, or by an investor scraping available web data to gain knowledge on market conditions. Presumably, the court would have reached a different result if the plaintiffs were not professors or big data researchers test auditing websites as opposed to data scrapers seeking commercial gain or otherwise engaging in what a website owner might deem "free riding." Secondly, the plaintiffs' activities were also not so extensive as to affect the various websites' server loads or provision of service to real customers, so there was no threat of irreparable harm to the websites' operations in this case from the plaintiff's "unauthorized access."
- The *Sandvig* decision is focused on the criminal CFAA issue and does not address the availability of civil causes of action for bypassing robots.txt, CAPTCHAs or other technical measures (e.g., common law trespass or the DMCA anti-circumvention provisions, if the content protected by a technological measure was copyrighted).
- The decision also does not address civil liability for breach of contract based upon a violations of a site's terms, though certain dicta in the opinion suggest that the court believes publicly available web data presents different issues under the CFAA than data behind a paywall or authentication scheme. ("The First Amendment does

not give someone the right to breach a paywall on a news website any more than it gives someone the right to steal a newspaper. But simply placing contractual conditions on accounts that anyone can create, as social media and many other sites do, does not remove a website from the First Amendment protections of the public Internet").

- Entities engaging in web scraping will certainly be buoyed by other dicta in the opinion that characterizes the scraping of public website data for research purposes as implicating First Amendment interests or merely "a technological advantage that makes information collection easier" (the latter characterization brings to mind a recent Ninth Circuit decision interpreting automated scraping activities with respect to the California state computer access law). While courts have considered scraping activities under various legal causes of action over the past decade, rarely has a court made such sweeping pronouncements about the utility and expressive nature of web scraping and the openness of the public web. Moreover, the court ruled that certain of the plaintiff's methods of access (e.g., scraping publicly accessible web data, publishing academic content using such data, or employing a bot to interact with a website when such activity is allowed for a human user) fell outside the Access Provision of the CFAA. Still, the court did not consider the same activities in the civil context, or if such interpretation would change if a website operator has expressly revoked a user's access to a public website in a cease and desist letter, an issue that the Ninth Circuit is poised to resolve.

View Original

Related Professionals

- **Jeffrey D. Neuburger**
  Partner