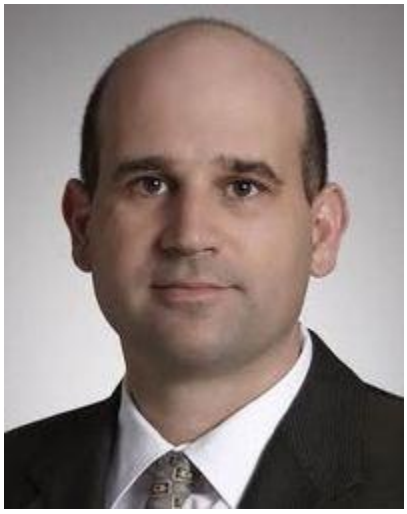
Portfolio Media. Inc. | 860 Broadway, 6th Floor | New York, NY 10003 | www.law360.com
Phone: +1 646 783 7100 | Fax: +1 646 783 7161 | customerservice@law360.com

How Website Operators Use CFAA To Combat Data-Scraping

Law360, New York (August 25, 2014, 10:01 AM ET) --



Aaron Rubin

"Web scraping" or "Web harvesting" — the practice of extracting large amounts of data from publicly available websites using automated "bots" or "spiders" — accounted for 18 percent of site visitors and 23 percent of all Internet traffic in 2013. Websites targeted by scrapers may incur damages resulting from, among other things, increased bandwidth usage, network crashes, the need to employ anti-spam and filtering technology, user complaints, reputational damage, and costs of mitigation that may be incurred when scrapers spam users, or worse, steal their personal data.

Though sometimes difficult to combat, scraping is quite easy to perform. A simple online search will return a large number of scraping programs, both proprietary and open source, as well as DIY tutorials. Of course, scraping can be beneficial in some cases. Companies with limited resources may use scraping to access large amounts of data, spurring innovation and allowing such companies to identify and fill areas of consumer demand. For example, Mint.com reportedly used screen scraping to aggregate information from bank websites, which allowed users to track their spending and finances.

Unfortunately, not all scrapers use their powers for good. In one case, the operators of the website Jerk.com allegedly scraped personal information from Facebook to create profiles labeling people "Jerk" or "not a Jerk." According to the Federal Trade Commission, over 73 million victims, including children, were falsely told that they could revise their profiles by paying $30 to the website.

Website operators have asserted various claims against scrapers, including copyright claims, trespass to chattels claims and contract claims based on allegations that scrapers violated the websites' terms of use. This article, however, focuses on another tool that website operators have used to combat scraping: the federal Computer Fraud and Abuse Act.

The CFAA imposes liability on "whoever … intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains … information from any protected computer." While the CFAA is primarily a criminal statute, it also provides for a civil remedy where a plaintiff suffers more than $5,000 in aggregate losses during any one-year period arising from a violation of the CFAA. For large website operators asserting CFAA claims against scrapers, the $5,000 damages requirement has not

proven to be a difficult obstacle to overcome.

For example, in CollegeSource Inc. v. AcademyOne Inc., the District Court for the Eastern District of Pennsylvania found that the plaintiff's cost of initiating an internal investigation of the defendant's website, hiring a computer expert to analyze the scope of the defendant's actions and implementing increased security measures were well in excess of $5,000. Similarly, in Facebook Inc. v. Power Ventures Inc., the District Court for the Northern District of California found that the plaintiff's expenditures made in response to defendant's specific acts, which included three to four days of engineering time, $75,000 in outside counsel costs and the costs of responding to a minimum of 60,000 instances of spamming by defendant, were well in excess of the statutory threshold. The more difficult question is whether scraping violates the CFAA at all.

The CFAA was originally intended as an anti-hacking statute, and its application to scraping — which, after all, usually involves accessing publicly available data on a publicly available website — is not always a foregone conclusion. Does a scraper access a website "without authorization" or "exceed authorized access" when it harvests publicly available data on a publicly available website? Plaintiffs often argue that scrapers act without authorization because the websites' online terms of use prohibit scraping and/or prohibit the scrapers' use of the data that they harvest. As discussed below, such claims have met with success in some cases, but courts have been less willing to find a CFAA violation in other scraping cases.

In Cvent Inc. v. Eventbrite, Cvent sued Eventbrite for scraping Cvent's website to obtain venue information and using the information in Eventbrite's "Venue Directory." Cvent claimed that this was a violation of the CFAA because Cvent's terms of use specifically stated that such activities were unauthorized. The District Court for the Eastern District of Virginia held that Eventbrite's actions did not constitute "hacking" in violation of the CFAA because the information was publicly available, Cvent's website did not require any login, password or other individualized grant of access, and Cvent's terms of use were difficult to locate. Therefore, the court granted Eventbrite's motion to dismiss, concluding that Eventbrite was authorized to access the information on Cvent's website, and that the mere allegation that Eventbrite used the information inappropriately was not grounds for relief under the CFAA.

Power Ventures, the defendant in Facebook Inc. v. Power Ventures Inc., operated a social media account integration site. As part of a promotion to gain new members, Power Ventures provided users with a list of their Facebook friends, which Power Ventures obtained through scraping the Facebook website, and asked users to select friends to invite to use the Power Ventures site. Facebook notified Power Ventures that its access was unauthorized and blocked Power Ventures' IP addresses. However, Power Ventures' scraping technology was designed to circumvent such technological measures and the scraping continued. The District Court for the Northern District of California held that Power Ventures' accessing of Facebook was without authorization and violated the CFAA and accordingly, granted summary judgment to Facebook on the CFAA claim.

CollegeSource, the plaintiff in CollegeSource Inc. v. AcademyOne Inc., maintained an archive of college course catalogs in PDF format and a hyperlink service called CataLink, both of which it made available to paying subscribers. AcademyOne, a CollegeSource subscriber, hired a third party to download college catalogs directly from college websites in order to compile a course description database. However, the third party instead copied some of the PDF documents from CollegeSource through CataLink. AcademyOne removed the CollegeSource documents from its system after receiving a cease-and-desist letter from CollegeSource, but CollegeSource nonetheless proceeded to bring a number of claims against AcademyOne, including CFAA claims based on the argument that AcademyOne accessed the

documents without authorization and exceeded authorized access.

The court held, however, that AcademyOne did not access the documents without authorization because those documents were available to the general public. CollegeSource's argument that AcademyOne exceeded authorized access was based on AcademyOne's alleged violation of CollegeSource's terms of use. The court acknowledged that accessing a website in violation of the applicable terms of use has been held to support a CFAA claim in some cases, but was unconvinced by CollegeSource's argument here because CollegeSource's subscription agreement did not cover CataLink. Accordingly, the court granted summary judgment to AcademyOne on the CFAA claims.

In Craigslist Inc. v. 3Taps Inc., 3Taps allegedly scraped Craigslist's website and republished Craigslist ads on its own site, craiggers.com. In response, Craigslist sent 3Taps a cease-and-desist letter revoking 3Taps's authorization to access Craigslist's website for any purpose, and reconfigured the website to block 3Taps. When 3Taps allegedly continued its scraping activities by using different Internet Protocol addresses and proxy servers to conceal its identity, Craigslist brought suit under the CFAA.

Even though Craigslist's website was publicly available, the District Court for the Northern District of California declined to grant 3Taps's motion to dismiss the CFAA claim. According to the court, while Craigslist may have granted the world permission to access its website, it retained the power to revoke that permission on a case-by-case basis, a power it exercised when it sent the cease-and-desist letter and blocked 3Taps's IP addresses. Therefore, 3Taps's continued access was without authorization.

The court also rejected 3Taps's attempt to invoke the Ninth Circuit's decision in United States v. Nosal. In Nosal, the Ninth Circuit had held that an employee's use of information in violation of an employer's policies did not constitute a CFAA violation where the employee's initial access to the employer's computer system was authorized. The court in 3Tap's concluded, however, that the "calculus is different where a user is altogether banned from accessing a website," as was the case with 3Taps.

Fidlar, the plaintiff in Fidlar Technologies v. LPS Real Estate Data Solutions Inc., provides its Laredo program to governmental agencies, such as county clerks' offices, which use Laredo to make public records available for viewing over the Internet. Laredo prevents users from downloading or electronically capturing the documents they view. Users who want a copy of a public record must pay the county a print fee. LPS, a real estate analytics company, contracted with many counties to access their public records using Laredo, but used a scraping program to capture documents electronically without paying any fees.

Fidlar sued LPS for violating section 1030(a)(5)(A) of the CFAA, which imposes liability on anyone who "knowingly causes the transmission of a program, code, or command, and as a result … intentionally causes damage without authorization, to a protected computer." The District Court for the Central District of Illinois denied LPS's motion to dismiss the CFAA claim, holding that Fidlar's complaint properly alleged that LPS undertook intentional actions that, among other elements of damage, compromised the integrity of Laredo.

In light of the cases discussed above, it seems that plaintiffs are likely to have more success asserting CFAA claims against scrapers where they clearly and unambiguously revoke authorization to access their websites and take affirmative steps to block the scrapers, as in 3Taps and Power Ventures.

In contrast, when the scraper ceases scraping after access is revoked and takes remedial action, as in CollegeSource, courts may be less willing to impose CFAA liability. As seen in Cvent, a mere terms of use

violation, particularly where the scraper may not have actual notice of the terms of use, may not support a CFAA claim. Whether the scraper is simply using software to collect publicly available information more efficiently or to do something else — such as to avoid paying fees for the information, as seen in Fidlar — may also be relevant.

In any event, in an era when data is expensive to collect, valuable to have and cheap to take, the CFAA, when properly used, remains a viable tool to combat scrapers.

—By Aaron Rubin and Tiffany Hu, Morrison & Foerster LLP

Aaron Rubin is a partner and Tiffany Hu is a summer associate in Morrison & Foerster's San Francisco office.