Portfolio Media, Inc. | 648 Broadway, Suite 200 | New York, NY 10012 | www.law360.com
Phone: +1 212 537 6331 | Fax: +1 212 537 6371 | customerservice@portfoliomedia.com

How To Combat Digital Shoplifting

Law360, New York (November 30, 2009) -- If your Web site is a digital storefront for your company, people are not the only window shoppers. Specialized software applications known as bots almost certainly also are taking a look — and perhaps taking more.

These automated tools act a lot like human Internet users, but they do everything far faster, including submitting queries to sites and extracting data from them.

This makes bots ideal for search engines, which use them to crawl the sprawling and expanding World Wide Web, and to gather copies of the Web sites they find to create the databases the engines use in responding to queries.

Unfortunately, the automation and speed of bots also make them ideal for less commendable uses, such as attacking sites to scrape large volumes of data at the behest of a direct competitor or a company whose business model is based on aggregating data from several sources, including your company's Web site.

Companies with subscription-based sites, sites containing credit card information or e-mail address, or sites offering comprehensive product and pricing information are particularly vulnerable to damage from scraping attacks.

The continued weak economy may motivate some to increase the scope of such attacks, particularly because the popular media often refer to scraping as permissible or, at worst, a "gray area."

If a company determines that bots threaten its Web site and business, the question becomes how best to prevent them from crawling the site, or slow them down once they get there.

There are a host of operational measures a site can take. For example, sites commonly include a special instruction to bots, called a robots.txt or robot exclusion file, that tells bots to stay away altogether or to keep out of certain parts of the site.

Indeed, not using a robots.txt file may cause a court to conclude later that the site owner impliedly licensed at least search engine bots to crawl and copy the site.

Sites also can use CAPTCHAs — those distorted characters embedded within an image that users are required to retype correctly before being allowed to log in and that bots have difficulty reading.

In addition, regularly inspecting server logs for excessive traffic may reveal unwelcomed bots' IP addresses, which then can be blocked.

None of these steps — whether taken individually or in combination, however, likely will be completely effective against unscrupulous bot operaters.

Compliance with robots.txt files is strictly voluntary. The most effective CAPTCHAs are annoying to intended human users and, increasingly, are susceptible to being broken by software. IP addresses can be masked or changed.

If those steps do not stop the bots, there are several well established legal theories that may be deployed successfully against bots and their controllers.

The cornerstone of many of these theories are a site's terms of use. Although bots do not read them, understand them or report on their existence or content to their controllers, courts nonetheless generally are willing to enforce them against the controlling person.

Accordingly, companies need to be sure that their sites' terms of use bar — explicitly — the types of behavior about which they are concerned.

With well-crafted terms of use in force (that is, properly amended if prohibitions are added, and conspicuously displayed, with at least a link to them on each page), a claim for breach of contract arises when a bot engages in prohibited conduct.

Terms of use also set up a potential claim for violation of the federal Computer Fraud and Abuse Act, which always can be asserted in federal court and which may be a more successful route than a breach of contract claim for a preliminary injunction.

A plaintiff pursuing a civil CFAA claim most of the time will have to prove that the defendant either accessed a computer "without authorization" or "exceeds authorized access." The terms of use will prove the absence of authorization.

Bots also may infringe copyrighted material on a Web site by virtue of the fact that bots work by copying much or all of a site for at least a short time while they locate and extract the target information.

In this context, terms of use prohibiting commercial bot use will demonstrate that the copying was outside the scope of any license and is not fair use.

If, for some reason, a site's terms of use cannot be brought to bear, the site owner still may have a legal remedy. For example, depending on which state's law applies, the site owner may be able to bring a claim for the common law tort of trespass to chattels.

Where available, this claim is most likely to succeed if the scraping has damaged the server(s) hosting the site or slowed the site to a crawl.

The site owner also may have a claim under the portion of the Digital Millennium Copyright Act that prohibits circumvention of technological measures that effectively control access to a copyrighted work.

Terms of use or the robots.txt file, which requires voluntary compliance, probably are not the kind of technological measures that the DMCA proscribes evading. However, other technical measures may be a sufficient predicate for a DMCA claim.

As competitors increasingly deploy bots, the cat and mouse game between those controlling the bots and those seeking to keep them out will intensify.

Vigilant use of the most current technical measures will deter many unwelcomed bots. For those that get through a site's defenses, there are effective legal remedies.

--By John F. Zabriskie, Foley & Lardner LLP

John Zabriskie is a partner with Foley & Lardner in the firm's Chicago office.

The opinions expressed are those of the author and do not necessarily reflect the views of Portfolio Media, publisher of Law360.