GESMER
UPDEGROVE
LLP
Attorneys At Law

Technology Law Bulletin
Providing comprehensive legal services to high technology companies

Vol. XI, No. 1 Summer 2003

Scrapers, Robots and Spiders: The Battle Over Internet Data Mining

Lee Gesmer
gesmer@lgu.com

When American Airlines sued Farechase, Inc. in federal district court in Texas earlier this year, claiming that Farechase's "screen-scraping" of AA's flight information from AA.com was illegal, it was only the most recent in a series of cases challenging unauthorized data collection from Internet web sites. What practices are encompassed by "screen-scraping?" Is "scraping" really illegal? What does this line of cases mean for your business?

> "While it does appear that the law has thus far favored original content providers, the law...is far from settled."

What Is "Screen - Scraping"? Despite its pejorative title, screen-scraping software simply gathers and aggregates data from other Internet websites for use by the gathering party. Usually, the purpose is to reformat the data and display it for the benefit of the gathering party's customers. Examples of data aggregation range from sites that collect prices on retail sites to companies that aggregate personal financial data on mutual fund and banking web sites, permitting registered users to access information about multiple accounts on a single web site.

The software that performs this function, often referred to as a robot (or "bot"), "spider" or "crawler," automatically searches Internet web sites for specific information. The *Farechase* case provides one example of how this technology works. Farechase's customers are travel agencies. When a customer uses Farechase to research a particular airline, hotel or auto rental fare, Farechase's software will search different airline sites and collect the "webfares" offered. A popular site such as AA.com might be searched thousands of times a day in response to queries initiated by Farechase customers. Farechase's real time search technology is an advance on more traditional data mining, in which companies search sites on a regular basis and maintain a separate database that may be queried by users. In the case of sites selling books or music, a real time search may not be essential, as long as the database is updated frequently. Farechase took this concept one step further by permitting its users to search for fares offered at the very moment the search is conducted, thus guaranteeing that the results would be current.

Needless to say, one's views on this type of data mining depend largely on whether one is the scraper or scrapee. The targets of this practice, such as American Airlines, complain that the constant traffic resulting from scraping puts an extra burden on their Internet servers, slowing down their response times for legitimate users. In the *Farechase* case American Airlines claimed that if left unstopped, Farechase would be performing over 200,000 daily searches by the end of 2003. Moreover, it argued that by permitting customers to access web fares by going directly to the American booking pages at AA.com, American is unable to establish the relationship with its customers that would occur if customers were required to navigate through AA.com's preliminary pages, thereby costing American customer good will. On the other hand, companies like Farechase argue that their service encourages comparison shopping, and that companies that resist it are afraid of the competition (and the lower prices) that result.

Technological Defenses. Before discussing the legalities of screen-scraping, it is worth pointing out that companies who are targeted by this practice and who object to it often undertake a measure of "self-help" before authorizing their lawyers to file suit. Such self-help sometimes leads to a technological battle worthy of a William Gibson novel. The defenders attempt to identify and block the Internet Protocol (IP) addresses of the attackers. The attackers respond by hiding or disguising their scrapers' identities by using fake IP addresses, thereby evading the blocking firewalls. The attackers, not easily discouraged, seem to have a limitless supply of disguises, perpetuating this high tech cat-and-mouse game. In several cases the attackers have prevailed. As a result, several of these disputes have ended up in the courts.

Legal Defenses. When technical defenses fail, screen-scraper targets such as American Airlines have two primary legal weapons to deploy in their defense. The first is to claim breach of a click-wrap or browse-wrap on-line license. The second is to allege a "tort" (or legal wrong), most commonly "trespass to chattel."

In its case against Farechase, American Airlines attempted to fire both barrels at its opponent, but its opening salvo was weak. First, American claimed that Farechase violated American's "browsewrap" agreement. By its use of the term "browsewrap" American was referring to an online agreement which appears on the site (usually under the terms and conditions link), but does not require the user to click on or express consent to the agreement before proceeding to use the site. By contrast, the better known (and far more effective) "click-wrap" agreement requires the first-time user to click on a word or symbol to express acceptance of a site's licensing terms before gaining access to the site. While the user of a properly implemented click-wrap agreement can expect enforcement, no court has yet enforced a browsewrap agreement, and the only two courts that have considered the issue at all have expressed doubts as to the enforceability of such an agreement. However, the ability to protect a web site with nothing more than an explicit statement on the website restricting access received a potential boost in a recent decision by the First Circuit Court of Appeals in Boston. That court suggested that screen-scraping may violate the Computer Fraud and Abuse Act (the "CFAA"), and that a restrictive warning of the sort used in browsewrap agreements may be enough to invoke the CFAA.

The second barrel of American's gun was loaded with more powerful munitions, in the form of its claim that Farechase had violated the law of "trespass to chattels" (i.e. goods). While the English law of trespass as applied to chattels can be traced back hundreds of years, it has shown a surprising ability to adapt itself to the law of the Internet. Most courts that have considered the applicability of trespass law to data scrapers have ruled in favor of the complaining party. The best known of these cases, *eBay, Inc. v. Bidder's Edge, Inc.*, resulted in an injunction ordering Bidder's Edge to stop data mining from the eBay website. Moreover, in several of these cases the courts have not required proof that the scrapers caused any measurable harm, or caused any specific injury, to the sites they were data mining.

Not surprisingly, based on the above record, American Airlines was successful in obtaining an injunction against Farechase. While Farechase is still in business, its searches no longer include American web fares.

TLB Comment: *Based on this state of the law, can data miners expect to build a business based on unauthorized screen-scraping? Somewhat surprisingly, the outlook may be better than it appears. First, many companies do utilize this form of data mining without objection from the owners of the sites they are crawling. The reasons are economic, not legal. In some industries screen-scraping has become an accepted method of business. Further, the vast majority of companies are willing to provide access to their sites when they are approached cooperatively. The fact that some percentage of their capacity is being used by a scraper is not a deterrent, as long as the scraper's customers ultimately are referred to the vendor's site to make the purchase.*

Second, while the law thus far has favored original content providers, the law on electronic trespass to chattels is far from settled. Just before this article went to press the California Supreme Court issued a decision in Intel v. Hamadi, rejecting Intel's attempt to prevent a former employee from sending mass e-mails to Intel employees. In that case the court held that electronic trespass to chattels is not actionable under California law unless it involves "actual or threatened injury to the personal property or the possessor's legally protected interest in the personal property." Since Hamadi's e-mails (numbering in the hundreds of thousands) to Intel employees caused no such harm, the court refused to order Hamadi to cease communications. Although this case was not a screen-scraping case, the issues implicated are essentially the same (Intel relied heavily on the scraper cases), and therefore Hamadi may be an important defensive tool for scrapers to use in the future.

The DMCA and Interoperability: A Troubling Legal Strategy in the Aftermarket Industries

Peter Moldave
moldave@lgu.com

The Digital Millennium Copyright Act (DMCA) prohibits technological devices that assist in the circumvention of copyright access controls. This "anti-circumvention" prohibition was enacted to prevent the manufacture of devices that could be used to circumvent "digital locks" on copyrighted materials such as books, films and music that are sold in digital form. However, a case decided early this year raises significant concerns about the ability of companies to use the DMCA not just to prevent the copying of music or other copyrighted media, but to use technology to lock-in consumers in order to prevent aftermarket competition.

In the case, *Lexmark International, Inc. v. Static Control Components, Inc. (SCC),* Lexmark used the DMCA and copyright laws to obtain a preliminary injunction, preventing SCC from selling less expensive toner cartridges for use with Lexmark's printers.

The *Lexmark* case is based on the economics of the low-end printer industry. Like video console manufacturers, the base equipment is sold at a very low price, perhaps even at a loss. But just as game cartridges are the profit center for console makers, toner cartridges are the profit center for printer companies.

Simplified somewhat, the facts of the *Lexmark* case are as follows. Lexmark printers and toner cartridges are manufactured in such a way that a software-based "authentication sequence" is resident on the toner cartridge. This sequence requires the toner to send an encrypted code to the Lexmark printer before the printer will work with the toner. SCC, a competitor in the toner market, reverse engineered the system to gain access to the authentication sequence, and copied the sequence onto a microchip embedded in its own cartridges. SCC advertised the fact that its microchip circumvented Lexmark's "secret code" and sent "the *right* messages" to Lexmark printers. Lexmark sued for violation of the DMCA and copyright infringement, seeking a preliminary injunction.

In addition to the unusual fact that the court found copyright infringement based on very minimal copying (approximately 50 bytes - this finding alone may be the basis for a successful appeal), the court's application of the DMCA is of great interest. The DMCA claim was based on the fact that the microchip embedded in the SCC toner cartridge was designed to circumvent the secret code that restricted access to the software loaded on the Lexmark printer. The court held that by creating a microchip that circumvented Lexmark's authentication sequence, SCC had violated the DMCA, which bans such circumvention devices. Moreover, the DMCA's "interoperability" exemption for reverse engineering was not available to SCC, since it had copied software contained on Lexmark's toner cartridge.

The implications of *Lexmark* may be far-reaching. Given the amount of embedded software in devices, it could provide new levels of protection for all sorts of aftermarket repair, accessory and replacement industries. We can foresee a variety of products, from electronic auto repair equipment (forcing consumers to use only authorized repair shops) to household appliances that utilize replacement components, being outfitted with conventional hardware/software or even encrypted radio frequency identification (RFID) chips to prevent the use of unauthorized aftermarket components. In the software industry, operating system manufacturers could put software locks on their systems, requiring licenses from application developers. The interoperability exemption, while in theory a way around these issues, may prove difficult and expensive to take advantage of in practice, due to the complexity of the technology involved and the ability of manufacturers to hide or disguise their locks.

TLB Comment: *In the case of manufacturers with monopoly power, these types of practices may lead to antitrust challenges, but monopoly power is rare, and most firms might be unaffected by antitrust concerns. While Lexmark is one of the first cases of this sort to reach the courts, in the absence of a congressional amendment to the DMCA we expect significant litigation in this area in coming years.*

GESMER UPDEGROVE • LLP •

Attorneys At Law

Gesmer Updegrove LLP
40 Broad Street
Boston, MA 02109

Address Correction Requested

Events & Announcements

September 9, 2003 - *Ken Appleby* will be presenting at a seminar entitled *LLCs: Advising Small Business Start-Ups and Larger Companies* sponsored by Lorman Education Services held in Cambridge, MA. His presentation will focus on current "hot topics" in the area of partnership taxation. More information is available at http://www.lorman.com

September 16, 2003 - Gesmer Updegrove will kick-off its *Technology Leadership Series* with the *Entrepreneur's Roadmap for Fundraising*, an interactive program featuring veteran entrepreneurs, venture capitalist and angel investors. The *Technology Leadership Series* is a new series of educational and networking events hosted by the firm. More information will be available later this summer at http://www.gesmer.com

October 1, 2003 - *Chris Dahl* will run *Angel and VC Financing*, the 3rd program in the Access to Capital Series of the Massachusetts Software and Internet Council sponsored by Gesmer Updegrove. Registration information will be available later this summer at http://www.msicouncil.org/calendar/listup.stm

October 22, 2003 - *Chris Dahl* will organize and moderate a panel discussion analyzing *Current Trends in Venture Investing (a Look at Markets and Deal Terms)* at a CapitalVenue meeting at the downtown Harvard Club in Boston. Registration information will be available later this summer at http://www.capitalvenue.com/events.htm

November 12, 2003 - *Ken Appleby* will be presenting at a seminar entitled *Partnerships, LLCs and LLPs: Organization and Operation in Massachusetts* sponsored by Lorman Education Services held in Peabody, MA. His presentation will focus on the basics of partnership taxation. More information is available at http://www.lorman.com

November 14, 2003 - *Bill Contente* will be speaking on *Financing Trends in the Medical Device Industry* at the Fifth Annual MassMEDIC MedTech Investors Conference held at the Park Plaza Hotel in Boston. Registration information is available at http://www.massmedic.com