

Investigating Source Code Thefts

for presentation at

The 24th Annual National Institute on
White Collar Crime

Miami Beach, FL
February 24-26, 2010

By: Eric Friedberg, Co-President, Stroz Friedberg
James Aquilina, Executive Managing Director, Stroz Friedberg
Matt Friedrich, Partner, Boies, Schiller and Flexner LLP

Investigating Source Code Thefts

By Eric Friedberg¹, James Aquilina², and Matt Friedrich³

Industrial spies, state-sponsored hackers, and defecting employees have targeted company source code either for profit, intelligence value, or in cyber war-game exercises. Source code is comprised of the files and instructions that underlie valuable computer applications. Such proprietary applications are often at the core of a company's business engine. Source code thefts have cut across various industries, involving, for example, genome analysis programs, esoteric airplane radar applications, even high-volume trading technology used by investment firms. When the thieves are corporate insiders, they are often programmers or business principals who have access to source code in the normal course of their duties and copy the code to removable media or use email or file transfer protocols to unlawfully send the code to an external e-mail account or file storage site.

Source code thefts can implicate the civil and criminal provisions of federal statutes prohibiting thefts of trade secrets⁴ and copyright infringement.⁵ If a thief gains unauthorized access to the source code via hacking, that access can be punished under the civil and criminal provisions of the federal Computer Fraud and Abuse Act ("CFAA").⁶ In addition, the CFAA covers corporate employees who "exceed" their authorized access to the code, i.e. personnel who may have authorized access to the code for legitimate corporate purposes, but whose transfer of the code for personal gain exceeds the scope of permitted access.

Upon initial indications of theft, victims face early and critical decisions about how they respond. Properly and expeditiously investigating source code thefts enables companies to make accurate determinations of what was stolen, how, by whom, and when. Proper fact-finding underpins the corporate victim's ability to make an effective criminal referral and to seek civil relief. An effective investigation correctly sequences computer forensic preservation and analysis of electronic data, traditional investigative techniques, liaison with law enforcement, and litigation.

¹ Eric Friedberg is Co-President of Stroz Friedberg, an international consulting and technical services firm focusing on digital forensics, cyber-crime response, investigations, and electronic discovery. He is the former lead computer crimes prosecutor at the U.S. Attorney's Office for the Eastern District of New York.

² James Aquilina is Executive Managing Director at Stroz Friedberg and runs the firm's West Coast offices. He is a former federal computer crimes prosecutor from the U.S. Attorney's Office for the Central District of California, an expert in malware and cyber-crime response, and is co-author of the leading malware treatise, *Malware Forensics: Investigating and Analyzing Malicious Code*, published by Syngress Publishing, Elsevier Science & Technology Books.

³ Matt Friedrich is a partner in the Washington office of Boies, Schiller and Flexner LLP. He is the former acting Assistant Attorney General of the Criminal Division and a former Assistant United States Attorney. Mr. Friedrich's practice includes counseling companies victimized by cybercrime.

⁴ 18 U.S.C. § 1832.

⁵ 17 U.S.C. § 503.

⁶ 18 U.S.C. § 1030.

Alternatively, a failure to confront suspected theft, or to confront it in a thorough and technically competent manner, can exacerbate the consequences to the victim. Failure to protect trade secrets can undermine the protection they are afforded under federal statutes. Of course, allowing one's proprietary source code to create value for a competitor can be economically devastating.

Sequencing the Initial Investigation

In terms of sequence, law enforcement authorities generally look askance at referrals that are made after the matter is bogged down in civil litigation. Accordingly, to the extent that the company wishes to conduct an internal investigation before contacting law enforcement, such investigation must be done very quickly. One advantage of an internal investigation is that the company can preserve all relevant sources of data to perform internal threat assessments. On occasion, when law enforcement is called in immediately, the agents insist on taking the original sources of media, which sometimes leaves the company without its own data, and thus without an ability to analyze that data for the who, what, and when of the attack. Often, an effective strategy for a victim is to preserve and maintain the integrity of relevant data, make an initial assessment, and refer to law enforcement as appropriate if a threshold determination can be made that a serious crime has occurred.

Forensic Preservation and Analysis

When source code is stolen, it is important to preserve all digital evidence from which the identity of the attacker and the nature of the attack can be determined. Preservation, or the failure to do so, has important consequences in terms of the ability to find electronic evidence, analyze it, and, where appropriate, introduce it in court. The analysis should be conducted by credentialed digital forensic examiners in secure digital forensics laboratories who have the expertise to reconstruct such thefts by analyzing the relevant digital media and data. Such reconstruction involves identifying digital logs, artifacts, and other evidence that the attacker's actions leave on the relevant computers and media. At a minimum, the forensic investigation should include the preservation and analysis of:

1. The desktop and laptop computers of any internal suspects. By forensically analyzing such media, it is often possible to reconstruct malicious actions, such as transfers of the source code to an external account, even if the suspect used his or her personal webmail account. The forensic examiner can also see what files were recently accessed by the computer, and what connections to outside websites or file servers were made, as the source code may have been transferred to such remote locations.
2. The server on which the source code resided. By examining the "last accessed" dates and times of the source code files as they reside on the server it may be possible to match those dates and times to the relevant

dates and times of actions on the suspects' laptops. If the source code was stolen from a connection from outside the corporate environment, the investigator still must attempt to construct a timeline of all of the malicious actions taken in accessing the source code. The metadata on the file server is critical to that effort.

3. The software development and web servers used to track and implement changes to source code. Programmers often use version control systems as repositories to edit, store, and maintain current and historical versions of code files. Access to these repositories is often managed by web server clients. Both may contain log data or configuration, authorization, or authentication files that prove valuable in reconstructing who accessed what and when.
4. Network devices containing logs of connections into the corporate network at or about the time of the theft. Such logs may include firewall logs; Virtual Private Network or "VPN" logs; DHCP logs, which contain records of the internal IP addresses that the network assigns to its own computers; and command history or "bash" logs, which record commands run against certain machines. Often, an inside or remote attacker's malicious commands (e.g. commanding a server to transfer the source code files to an external address) will be recorded in a command history.
5. Email servers and other devices which store email communications from both corporate and personal accounts. Internal investigative efforts should endeavor to preserve email and attachments containing source code or other proprietary information, as well as correspondence with potential competitors.
6. Other digital evidence demonstrating the victim's efforts to protect the information as a trade secret or as confidential or proprietary. Such information might include login or port banners, access control logs, employee use and IP agreements, and policies on need-to-know access rights. Such evidence can support the victim's right to rely on federal statutes protecting against theft of trade secrets.

If the victim is fortunate, the forensic analysis shows clearly how, when, and to where the stolen source was transferred. If an investigation warrants and allows an interview of an employee or ex-employee who is suspected of source code theft, that interview should take place only after the forensic analysis has been done. Most interviewees have a difficult time explaining away a trail of digital bread crumbs that lead right back to their work or personal computers.

Proactive Measures

If strong and clear enough, the forensic analysis can then be used in an immediate and proactive fashion. In criminal cases, forensic evidence, when shared with authorities, may sometimes lead to a search warrant. In civil cases, it may be used as part of the showing necessary to obtain a seizure order – either *ex parte* or on notice – allowing the victim to seize evidence of the theft at the location to which the data was transferred.⁷ In seizure orders or requests for inspection, the term “source code” can be defined broadly to include not just the plain text files containing the alphanumeric programming language (typically C, C++, Cobol, Fortran, Java, Perl, PHP, Python or Tcl/Tk), but also any related comments or notes, earlier versions or historical modifications, and associated log files relating to the development, storage of, and access to the code.

Due to the element of surprise, search warrants and *ex parte* seizure orders can result in the discovery of highly incriminating evidence. While reporting evidence of a suspected crime is cloaked with qualified immunity, proceeding via a civil *ex parte* order bears a number of risks. Showing up at the defendant’s premises unannounced, accompanied by a Marshal,⁸ and seizing or forensically imaging an ongoing business’ computers and media dramatically raises the temperature of the litigation. The opportunities and costs of such aggressive measures must be weighed carefully. A party subject to such an order may respond with spurious but hard-to-disprove claims that the adverse forensic imaging resulted in destruction to its systems or data. Magistrates and Judges have been known to have second thoughts after the seizure, especially once the

⁷ See 17 U.S.C. § 503 (providing authority for courts to order the impoundment of certain materials in copyright infringement cases); 15 U.S.C. § 1116(d)(1)(a) (providing authority for courts, in civil actions involving counterfeit goods, to issue an order upon *ex parte* application for “seizure of goods and counterfeit marks involved in such violation and the means of making such marks, and records documenting the manufacture, sale, or receipt of things involved in such violation.”).

Such seizure orders should be pursued only after careful analysis of the underlying facts and law, and even then only with great care. See Jay Dratler, Jr. and Stephen M. McJohn, *Intellectual Property Law: Commercial Creative and Industrial Property* §13.03 (recognizing that “such orders are fraught with danger both to innocent defendants and to the integrity of the judicial process.”); see also *Paramount Pictures Corp. v. Doe*, 821 F. Supp. 82 (E.D.N.Y. 1993) (noting that Supreme Court’s Copyright Rules on impoundment may not comply with due process); *Warner Bros., Inc. v. Dae Rim Trading, Inc.*, 677 F.Supp. 740, 765-67 (S.D.N.Y. 1988) (noting narrow scope of Copyright Rules on impoundment), *aff’d in relevant part and rev’d on other grounds*, 877 F.2d 1120, 1123-26 (2d Cir. 1989). Compare *Vuitton v. White*, 945 F.2d 569, 574-76 (3d Cir. 1991) (finding abuse of discretion in district court failure to grant seizure order as to alleged counterfeit goods held by street vendors) with *Lorillard Tobacco Co. v. Bisan Food Corp.*, 377 F.3d 313, 319-22 (3d Cir. 2004) (affirming district court’s refusal to issue seizure orders against three retail cigarette vendors with fixed addresses).

⁸ See, e.g. GREGORY J. BATTERSBY & CHARLES W. GRIMES, *LAW OF MERCHANDISE AND CHARACTER LICENSING* § 13:10 (“In accordance with Rule 3 of the Rules of Practice governing impoundment procedures under the Copyright Statute, the plaintiff may, at any time after the filing of an action, seek an impoundment order by filing with the clerk of the court an affidavit stating the number, location, and value of the allegedly infringing articles and posting a bond of not less than twice the value of the articles. The clerk of the court, upon approval by the court, is then empowered to issue a writ directing the United States Marshal to seize and hold those items identified in the affidavit.”) (citation omitted) (emphasis added).

supporting affidavits are disclosed and challenged, which can result in changes to the Court's willingness to allow the plaintiff to fully inspect the seized data.

If the plaintiff does not proceed by way of seizure order, it can also initiate a suit and request that its code be produced by the defendant pursuant to a request for inspection or document request.

Source Code Review Procedures in Civil Litigation

When sued, the defendant often demurs that it did not steal the plaintiff's source code but arrived at its competing software application by independent efforts. The defendant also typically protests that any inspection of the defendant's source code would reveal trade secrets to a direct competitor. To balance inspection rights with protection of independent intellectual property, the parties normally will agree to an inspection protocol that is typically so-ordered by the Court. These protocols are normally comprised of complex stipulations and orders that take weeks to negotiate. In setting forth where the source code review takes place, and what security precautions govern the review, protocols often contemplate the following:

1. The Room. Often, a trustworthy third party administers the security at the source code room. While the room may be located at the defendant's law firm, some protocols provide that the third party will have exclusive access to the room, i.e. that it will be separately keyed by the third party. In other cases, the defendant's lawyers administer the room, although in bitterly-fought cases this often results in squabbles over scheduling and whether the details of the inspection protocol are being followed. Protocols sometimes require video surveillance, logging access to the room, and documenting all actions taken in the room with respect to the specific data that is the subject of inspection.
2. Encryption. The source code in question is often kept on encrypted media which only the third party can unlock, and the Internet access and USB slots of the computer on which the source code reside are often disabled.
3. Review Media. Locked down, standalone computers are often used by the parties inside the room to access the encrypted media housing the source code in order to conduct the agreed upon analysis. Administrative privileges on all examining machines are typically disabled, and the installation of forensic software on review machines is typically limited, supervised, and agreed upon in advance. Notes, analyses, and findings can be stored locally on the standalone review machine.
4. Attorney's Eyes Only. In virtually all cases, the defendant's code is designated as Attorneys Eyes Only: while plaintiff's expert may disclose to plaintiff's attorneys portions of the code necessary for expert analysis, the plaintiff itself is prohibited from reviewing any part of the code.

5. Access. Rules regarding when plaintiff's expert may inspect the source code are often established. Some protocols simply provide for "reasonable access," whereas others set forth particulars, especially when the source code review is expedited and experts require access to the room after hours and on the weekends.
6. Copying. In some cases, nothing leaves the room other than hard copy printouts of notes, work product, and relevant code snippets. In others, electronic versions of the same may be copied to encrypted media and removed from the room by those who have agreed to the non-disclosure provisions of the protocol. Because the protocol is normally so-ordered, intentional violation can result in an order of contempt.
7. Destruction. Most protocols provide for the permanent deletion of the source code and the forensic wiping of all media at the end of the matter, however that end is defined.

In cases where the defendant asserts a defense that it independently developed its own source code for a product that competes with plaintiff's, the inspection can be bilateral, as the defendant's experts need access to the plaintiff's source code to assess whether the defendant's end-product incorporates actual pieces or structures from plaintiff's code. In all source code matters where a hearing is likely, especially code comparison cases, the parties are well served to utilize witnesses who have sufficient experience and credentials reading and writing code in the applicable programming language, and/or to utilize forensic experts who can validate with mathematical tools the wholesale copying of sections of the code.

Conclusion

If not redressed, source code thefts can seriously undermine a company's competitive advantages and the validity of its intellectual property protection. A robust digital forensic investigation, accompanied by the appropriate civil litigation, a source code review under a proper protocol, and possibly a law enforcement referral, can provide that redress.