

The COMPUTER & INTERNET *Lawyer*

Volume 23 ▲ Number 10 ▲ OCTOBER 2006

Privacy/Security

Lost Back-Up Tapes, Stolen Laptops, and Other Tales of Data Breach Woe

By **Eric Friedberg** and **Michael McGowan**

Employees' and executives' laptops are being stolen from conference rooms and left in taxicabs; data storage vendors are losing corporate backup tapes with uncomfortable frequency; overnight delivery firms are losing backup tapes and servers; and law firm information technology (IT) departments are having external hard drives loaded with client data walk out the door over holiday weekends. Many of these lost devices contain protected personal information, including names, credit card numbers, Social Security numbers, and medical information. Almost none of this data is encrypted. These data losses are straining relations with clients, creating regulatory and civil exposure, and causing enormous reputational damage.

Eric Friedberg and **Michael McGowan** are leading experts in cyber-crime response and computer forensics. Mr. Friedberg is Partner and General Counsel at Stroz Friedberg, LLC. Prior to that, Mr. Friedberg served as an Assistant US Attorney in the Eastern District of New York and was the Computer and Telecommunications Coordinator (lead cyber-crime prosecutor). Mr. McGowan is Director, Forensics at Stroz Friedberg. Stroz Friedberg is a consulting and technical services firm specializing in digital forensics, cyber-crime response, electronic discovery, and private investigations.

Investigation into the circumstances of the theft or disappearance and technical analysis of whether the data on these computers is easily readable may be critical to guide the victim and outside counsel on whether reporting obligations have been triggered under the patchwork of state data breach notification statutes.¹ While victims may wish that the lost data were not easily readable, making this evaluation often requires application of a combined forensic, investigative, and legal methodology. Ultimately, these data losses may push companies to consider more carefully their retention and usage policies regarding confidential and personal information and to use technologies that encrypt data on backup tapes, laptop drives, and removable media so that the data are unreadable if lost, regardless of the skill level of the thief or person who finds them.

The Law

California enacted the first data breach notification statute, which has been used as a model by other states. Generally, the statutes require notification to certain individuals and sometimes a state agency if, as the result of a breach in a company's computer security, individu-

Privacy/Security

als’ “personal information” is compromised. The statutes generally share the same key elements, but vary in how those elements are defined, including the definitions of “personal information,” the entities covered by the statute, the kind of breach triggering notification obligations, and the notification procedures required.

California defines personal information as a person’s name plus either a Social Security number, driver’s license number, California identification card number, or an “account number, credit or debit card number in combination with any required security code, access code or password that would permit access to an individual’s financial account.” Most of the other states follow California’s formulation. Arkansas, Florida, Texas, North Dakota, and Maine have broader definitions of personal information to include medical records, unique biometric data, the individual’s digitized or other electronic signature, and unique electronic identification numbers. All of the states, except Rhode Island, exempt personal information that is publicly available from federal, state, or local government records.

In California, data breach notification is triggered if there is an “unauthorized acquisition of computerized data that compromises the security, confidentiality, or integrity of personal information.” Again, most of the states follow California’s formulation. Florida, Montana, Nevada, and Tennessee add the requirement that the security breach be “material,” although it is unclear what materiality means in this context, since a breach of even a few records could lead to identity theft and fraud.

Most statutes exempt reporting if the compromised information is “encrypted,” although the statutes do not set forth the standards for such encryption. Arkansas, Florida, New Jersey, Rhode Island, and Washington exempt reporting if, under all of the circumstances, there is no reasonable likelihood of harm, injury, or fraud to customers. Arkansas requires the company to make “reasonable investigation” before coming to the conclusion that there is no reasonable likelihood of harm.

Notification to the affected customers may ordinarily be made in writing, electronically, or telephonically, or, in the case of large scale breaches, by publication. Many of the statutes require simultaneous notification to a designated state agency. California’s statute requires notification at “the most expedient time possible and without unreasonable delay.”

Florida and Ohio require notification to be made no later than 45 days following the determination of the breach. Under most state statutes, Illinois being an exception, notification can be delayed if it is determined that the disclosure will impede or compromise a criminal investigation.

Fines for failure to notify can be per day, per violation, per incident, or some combination thereof. For example, Ohio assesses \$1,000 per day; New York assesses the greater of \$5,000 per incident or \$10 per person who should have been but was not notified; and Florida’s fines are a \$500,000 maximum per incident for the first 180 days of non-compliance and an additional \$500,000 for the next 180 days of non-compliance. Rhode Island charges \$100 per person who should have been but was not notified, up to a maximum of \$25,000 per incident.

Technology and Strategy

The technological and strategic analysis for dealing with lost media differs between lost backup tapes and lost computers and hard drives. When backup tapes are misplaced or stolen, the company can avail itself of the statutory exemption that applies if the data on the tapes are encrypted. Unfortunately, most corporate backup tapes are not encrypted, because encrypting tapes can be an expensive, time-consuming, and bandwidth-consuming process.

But what if the data on lost backup tapes, while not encrypted, are so difficult and expensive to restore to a readable form that, as a practical matter, that data are as good as encrypted? While the statutes do not specify an encryption standard, arguably “very difficult and expensive to read” is not encryption. If configured and used in the most secure manner, domestic, commercial encryption products make the data impossible to read without private encryption keys and a user’s or administrator’s pass phrase, regardless of the skill of the thief. By contrast, most unencrypted backup tapes can be read in whole or in part by a skilled adversary with the right hardware and software.

In the few states that exempt reporting when there is no “reasonable likelihood” of harm to consumers, the difficulty of reading data off the lost or stolen backup tapes bears directly on the likelihood of harm, and a technical assessment of that difficulty could well support a decision not to notify consumers.

Assuming “very difficult and expensive” to read provides a legal basis to avoid reporting, a company must use a careful technical and investigative methodology to reach such a conclusion about a backup tape. First, the company should deploy traditional investigative techniques to determine as much as possible about the circumstances of the theft, the motivation and identity of the thieves, and likelihood of recovery of the stolen media. If a Russian organized crime ring specializing in identity theft stole corporate backup tapes with the complicity of a corporate insider, the risk profile would be far different from backup tapes that had fallen off a truck and were picked up off the street by a Mister Softee driver. The organized crime group is likely to devote a much higher level of skills and resources to read data off the tapes. The investigation into whether the disappearance was a result of targeted theft or negligent loss may require examination of chain of custody records, interviews of custodians, background checks on potentially complicit insiders, analysis of videotapes and cardkey logs, and research on similar thefts to identify patterns or organized crime groups that may be targeting similar forms of data across a series of companies.

If someone other than a skilled and motivated thief stole the tapes, it is highly unlikely that the thief could access the data. An average person who finds a lost tape would have no idea what hardware, software, or restoration techniques to use to access the data on the tape. Some risk remains, however, that an unskilled thief could sell or otherwise transfer the stolen media to more skilled persons, who are motivated to make the investment in time and money to recover the data. The question then becomes, as a practical matter, how difficult is it for a skilled and motivated thief to read stolen or found enterprise backup tapes? Is it so difficult that it could be argued that the tapes are *de facto* encrypted? Is it at least sufficiently difficult to read the stolen data to support the argument that there is no reasonable harm to consumers?

The technological and strategic analysis for dealing with lost media differs between lost backup tapes and lost computers and hard drives.

Enterprise backup tapes are often made using expensive hardware devices, called “tape drives,” into which

the tapes are loaded; the tape drives write data to and read data from the tapes. Enterprise hardware solutions can cost many thousands of dollars. Individual tapes can often be read with individual tape drives that are commercially available new for under \$10,000, purchased used for half of that, and rented for as little as \$1,500 per month, however. If an insider is complicit in the theft of the tape, the thief may also be given free access to an appropriate tape drive to read data off the tape. Accordingly, citing the high cost of buying a new, full enterprise backup environment as a barrier to reading a stolen or lost tape is likely an insufficient basis to claim that there is no reasonable harm to consumers from the theft. Rather, skilled thieves would likely be able to obtain the hardware at little or no cost. Medium to low skilled individuals who find the tape would not likely even be able to get that far.

Companies also sometimes reason that lost backup tapes are unreadable as a practical matter because the thief likely cannot afford to buy the expensive software with which the backup is made. Licenses for the expensive software, such as Veritas or ArcServe, used to create the backup tapes can easily exceed \$10,000. Normally, corporate environment backup tapes are restored to readable form using the same software. Nevertheless, the retail cost of that backup software may not be a barrier to retrieving the data off the tape since copies of this software can sometimes be illegally downloaded from warez sites.

In making an assessment that such factors would render critical data on a backup tape *de facto* unreadable, the better practice is to perform an actual test to determine whether a skilled forensic examiner can unscramble the data.

In addition, it is possible to read raw data directly off the tapes without the software used to create the backup. When raw data are dumped from a tape to a hard drive, some or much of that data may not be in a proprietary format and thus may be readable without the original software. In other words, some or much of the data may be in clear text that is searchable with standard search utilities. Some backup systems, such as

IBM's Tivoli system, make it infeasible to read data on individual tapes. The Tivoli backup system relies on a tape index to reconstitute complete files and compresses data in a proprietary format, rendering data effectively inaccessible without the Tivoli back-up software.

A company with a data loss problem should conduct a technical investigation to determine how much and what kind of information is on the tape "in the clear." In some cases, credit card numbers, Social Security numbers, and dates of birth may exist on the tape in the clear. Although they may be interspersed in large swaths of unreadable, proprietary data, it is easy for a skilled attacker to search for the valuable identity information using "grep" searches. A grep search is a search through the data for numbers or characters that have a specified logic. For example, a thief could execute a grep search for all 16 digit numbers beginning with 5424, 5466, 5473, or any of the other MasterCard prefixes or for four digit numbers ending in 06, 07, 09, or 10, representing expiration dates. The search could further specify 100 characters on either side of one of these grep searches be retrieved, in the hope that within those 100 characters might be names associated with credit card numbers or expiration dates. These grep searches can return results against enormous quantities of data in a matter of hours. Thus, the total volume of data on a lost backup tape does not act as much of a speed bump for a thief who has the skill to run the appropriate searches: For the skilled thief, the wheat is easily separated from the chaff.

The Readability of the Data

Sometimes, however, the critical data on a backup tape is in a proprietary format that makes the data extremely difficult or impossible to read without the application or hardware that created the data. Further difficulty in reading data from a tape can occur when larger files are stored in non-contiguous "blocks" (discrete storage areas) on the tape, "striped" across tapes (meaning when the backup system puts different parts of the same file on multiple tapes, so that the file cannot be read linearly off a single tape), or "compressed" (meaning that an algorithm has been used to pack more data into the limited space on the tape). In making an assessment that such factors would render critical data on a backup tape *de facto* unreadable, the better practice is to perform an actual test to determine whether a skilled forensic

examiner can unscramble the data rather than to rely solely on theoretical pronouncements. The tests should be performed on the most representative sample of the lost data. This may be the backup tape made nearest in time to the lost tape or a current backup of the contents of the server for which the backup tape was lost.

There are a number of ways to encrypt data as it is stored on servers or when it is backed up to tape. Enterprise database applications such as SQL and Oracle can encrypt database tables as they reside on their servers. When the servers are backed up to tape, the data is already encrypted. One added advantage to these technologies is that they protect against live hacks of database servers. If the hacker is able to compromise the server, the tables that s/he accesses are unreadable due to the encryption. The disadvantage to these technologies is that they degrade the performance (speed) of the server, since every time data is read it must be decrypted, which takes processing time.

Many of the major enterprise backup technologies, including CA ARCserve, IBM Tivoli Storage Manager (TSM), and VERITAS NetBackup, have the ability to encrypt data during the backup process, although these solutions present overhead and encryption key management issues. On the other hand, these technologies are simply a part of existing backup software infrastructures and thus present no additional cost. Technologies also exist, such as MediaMerge/TapeSecure from eMag Solutions, that can impose a layer of encryption when copying a tape for shipment offsite. Obviously, such a solution requires a second set of tapes, which can be very expensive. Finally, encryption of a redundant copy can be achieved using a hardware storage appliance that sits in the path of the data, does not degrade server performance, and has centralized encryption key management. The principal disadvantage to such a solution is cost.

Executives are going to have to assess the extra cost in tapes and system resources in deploying such technologies and to weigh that cost against complying with data breach notification statutes, possible fines, and damage to reputation when unencrypted data are lost. In a recent case involving the theft of a single server that housed medical information tied to Social Security numbers, the victim spent hundreds of thousands of dollars in tape restoration, forensic, and database costs to identify the hundreds of thousands of affected consumers and then to establish their current addresses for notification purposes.

Privacy/Security

Among the advantages of disk encryption is that not only do all the traditional files within the encrypted disk get encrypted, but so do the temporary files, deleted files, and remnants of merely viewing files.

When it comes to lost or stolen laptops, desktops, or external hard drives, the technical and investigative questions are somewhat different from those involving backup tapes. If a third party gains possession of a computer that has no Windows password, s/he will be able to read the data on the computer simply by booting it up. A Windows password-protected computer will be harder to access, although even a moderately skilled thief can install utilities to bypass the Windows password or simply take the hard drive out and read it with forensic software. In this way, a thief can view the contents of any unencrypted files stored on the hard drive in the same way as one would browse the contents of a CD-ROM.

Sometimes, individual files or email attachments, such as Word documents or Excel spreadsheets, can be password-protected. Password-cracking utilities are commercially available that can be used to attempt to crack such passwords by “brute force,” that is, by successively trying millions of combinations of letters and numbers. This effort is not trivial, however, and could take a week of processing time on a very fast computer to crack even a single, moderately strong password. If the user has protected a document with a strong password that incorporates 10 numbers, letters, and symbols, some upper case and some lower, the file may be effectively unbreakable by even a motivated attacker, absent the harnessing of massive computing power capable of generating billions of combinations in an acceptable period of time. The use of either strong or weak file-protecting passwords would probably count as “encryption” within the meaning of the data breach notification exemptions, since most file-based password technologies scramble the data in the file rather than just prevent the opening of the file.

There are a host of true encryption programs that can be used to protect data on laptop, desktop, and external hard drives. These include PGP (Pretty Good Privacy), Whole Disk Encryption, and EFS encryption technology, which is included with Windows XP. These technologies can be used to encrypt an entire hard drive, a volume (a smaller, defined area on a hard drive),

folders, or individual files. The loss of any lost data in an encrypted disk, volume, folder, or file does not trigger data breach notification under the above statutes, but would be exempt. Among the advantages of disk encryption is that not only do all the traditional files within the encrypted disk get encrypted, but so do the temporary files, deleted files, and remnants of merely viewing files. A highly motivated thief might attempt to use forensic techniques to recover these collateral materials, which can contain just as much valuable information as the original file itself. Disk encryption prevents such forensic recovery, unless a thief gains live access to the disk while it is being used in a decrypted state.

Some email programs, such as Lotus Notes, have built-in encryption that can be enabled, so that, if a computer is stolen or lost, the email residing on the computer will be encrypted.

Ironically, while victims who use encryption technology can clearly invoke the encryption exemption to avoid reporting the loss of encrypted data, in many situations it is not the encryption program itself, but the strength of the user’s pass phrase, that controls whether a thief can decrypt the data. For example, a number of encryption programs use a dual key system. In such systems, an encryption algorithm² is used to generate a key pair, including a private key and a public key. The user publishes his public key to others, who may then use that key to encrypt messages to send to the user. Those messages can be decrypted only with the user’s private key. The real strength of the strongest, current encryption programs is that intruders can not decrypt messages if they are intercepted in transit over the Internet. The algorithms are so strong that they do not create any patterns that can be reversed into plain text without use of the private key, and the private key can not be deduced from the public key.

If a disk encryption product or an email encryption product is configured on a hard drive, however, as they often are, so that public and private keys are on the hard drive, then a thief who comes into possession of the hard drive can attempt to decrypt the data on the drive by using the private key. The private key is normally protected with a pass phrase. If the pass phrase is weak (*e.g.*, three lower case letters like “abc”), a brute force attack on the pass phrase can be successful, leading to a compromise of the data. That is why it is a more secure practice, though highly inconvenient, to keep one’s private key on a separate device that is attached to the

computer only when decryption is desired. That way, if the computer is lost or stolen, the attacker is faced with the impossible or near-impossible task of attempting to decrypt the data by breaking the algorithm (by detecting patterns in the encrypted data and reversing them into clear text), since the private key is unavailable.

Prevention

In tandem with deploying encryption technologies, companies can eliminate wide swaths of risk through changes in corporate governance, user behavior, and data retention/recycling policies. Most companies keep enormous stores of unnecessary data whose only effect is to increase risk. Companies keep weeks, months, and years of backup tapes pursuant to misguided disaster recovery plans when there is no legal obligation to do so; retail credit card databases keep highly sensitive data that is key to identity theft, such as CV2 records (those three numbers on the back of your credit card), in violation of their agreements with the card issuers; companies, law firms, and consulting firms fail to physically secure laptops, desktops, and external hard drives from common theft by cleaning personnel, messengers, and vendors who steal the devices not for the data but for the hardware; and companies let outdated lists of identity information relating to employees, customers, and would-be customers languish for years on servers. Pri-

vacuity audits can help companies take stock of and mitigate the risk presented by their data retention profiles.

Conclusion

Once a company loses or steals unencrypted data, it will incur costs in investigating the loss, determining the number and identities of affected customers, assessing their data breach notification obligations across many state statutes, and then complying with those obligations. Encryption technologies at the desktop, server, and backup tape level provide excellent security and provide complete exemption from notifying consumers. Companies should employ technological solutions as feasible, in tandem with company-wide efforts lawfully to reduce the kinds and amounts of legacy, identity data that provide marginal business benefit but enormous risk.

Notes

1. 2005 Ark. Acts 1526; Cal. Civ. Code § 1798.82; 2005 Conn. Pub. Acts 148; 75 Del. Laws 61 (2005); 2005 Fla. Laws ch. 229; 2005 Ill. Laws 36; 2005 Ind. Acts 91; 2005 Ga. Laws 163; 2005 La. Acts 499; 2005 Me. Laws 379; 2005 Minn. Laws 167; 2005 Mt. Laws 518; 2005 Nev. Stat. 485; 2005 N.J. Laws 226; N.Y. Gen. Bus. Law § 899-aa (2006); 2005 N.C. Sess. Laws 414; 2005 N.D. Laws 447; H.B. 104, 126th Gen. Assem., Reg. Sess. (Oh. 2005); 2005 Pa. Laws 94; 2005 R.I. Pub. Laws 225; 2005 Tenn. Pub. Acts 473; 2005 Tex. Gen. Laws 294; 2005 Wash. Laws 368.
2. Some leading encryption algorithms are RSA, Triple DES, Blowfish, and IDEA.

Reprinted from *The Computer & Internet Lawyer*, Volume 23, Number 10, October 2006, pages 6-10, with permission from Aspen Publishers, a WoltersKluwer Company, New York, NY
1-800-638-8437, www.aspenpublishers.com