Data Hemorrhages in the Health-Care Sector[1]

M. Eric Johnson

Center for Digital Strategies
Tuck School of Business
Dartmouth College, Hanover NH 03755
{M.Eric.Johnson}@darmouth.edu

Abstract. Confidential data hemorrhaging from health-care providers pose financial risks to firms and medical risks to patients. We examine the consequences of data hemorrhages including privacy violations, medical fraud, financial identity theft, and medical identity theft. We also examine the types and sources of data hemorrhages, focusing on inadvertent disclosures. Through an analysis of leaked files, we examine data hemorrhages stemming from inadvertent disclosures on internet-based file sharing networks. We characterize the security risk for a group of health-care organizations using a direct analysis of leaked files. These files contained highly sensitive medical and personal information that could be maliciously exploited by criminals seeking to commit medical and financial identity theft. We also present evidence of the threat by examining user-issued searches. Our analysis demonstrates both the substantial threat and vulnerability for the health-care sector and the unique complexity exhibited by the US health-care system.

Keywords: Health-care information, identity theft, data leaks, security.

1 Introduction

Data breaches and inadvertent disclosures of customer information have plagued sectors from banking to retail. In many of these cases, lost customer information translates directly into financial losses through fraud and identity theft. The health-care sector also suffers such data hemorrhages, with multiple consequences. In some cases, the losses have translated to privacy violations and embarrassment. In other cases, criminals exploit the information to commit fraud or medical identity theft.

[1] Experiments described in this paper were conducted in collaboration with Tiversa who has developed a patent-pending technology that, in real-time, monitors global P2P file sharing networks. The author gratefully acknowledges the assistance of Nicholas Willey and the helpful comments of Lane R. Hatcher. This research was partially supported by the U.S. Department of Homeland Security under Grant Award Number 2006-CS-001-000001, under the auspices of the Institute for Information Infrastructure Protection (I3P). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security, the I3P, or Dartmouth College.

Given the highly fragmented US health-care system, data hemorrhages come from many different sources—ambulatory health-care providers, acute-care hospitals, physician groups, medical laboratories, insurance carriers, back-offices of health maintenance organizations, and outsourced service providers such as billing, collection, and transcription firms.

In this paper we analyze the threats and vulnerabilities to medical data. We first explore the consequences of data hemorrhages, including a look at how criminals exploit medical data, in particular through medical identity theft. Next, we examine types and sources of data hemorrhages through a direct analysis of inadvertent disclosures of medical information on publically available, internet-based file sharing networks. We present an analysis of thousands of files we uncovered. These files were inadvertently published in popular peer-to-peer file sharing networks like Limewire and Bearshare and could be easily downloaded by anyone searching for them. Originating from health-care firms, their suppliers, and patients themselves, the files span everything from sensitive patient correspondence to business documents, spreadsheets, and PowerPoint files. We found multiple files from major health-care firms that contained private employee and patient information for literally tens of thousands of individuals, including addresses, Social Security Numbers, birth dates, and treatment billing information. Disturbingly, we also found private patient information including medical diagnoses and psychiatric evaluations. Finally, we present evidence, from user-issued searches on these networks, that individuals are working to find medical data—likely for malicious exploitation.

The extended enterprises of health-care providers often include many technically unsophisticated partners who are more likely to leak information. As compared with earlier studies we conducted in the banking sector (Johnson 2008), we find that tracking and stopping medical data hemorrhages is more complex and possibly harder to control given the fragmented nature of the US health-care system. We document the risks and call for better control of sensitive health-care information.

2 Consequences of Data Hemorrhages

Data hemorrhages from the health-care sector are diverse, from leaked business information and employee personally identifiable information (PII) to patient protected health information (PHI), which is individually identifiable health information. While some hemorrhages are related to business information, like marketing plans or financial documents, we focus on the more disturbing releases of individually identifiable information and protected health information. In these cases, the consequences range from privacy violations (including violations of both state privacy laws and federal HIPPA standards) to more serious fraud and theft (Figure 1).

On one hand, health-care data hemorrhages fuel financial identity theft. This occurs when leaked patient or employee information is used to commit traditional financial fraud. For example, using social security numbers and other identity information to apply for fraudulent loans, take-over bank accounts, or charge purchases to credit cards. On the other hand, PHI is often used by criminals to commit traditional medical fraud, which typically involves billing payers (e.g.,

Medicaid/Medicare or private health-care insurance) for treatment never rendered. The US General Accounting Office estimated that 10% of health expenditure reimbursed by Medicare is paid to fraudsters, including identity thieves and fraudulent health service providers (Bolin and Clark 2004; Lafferty 2007).

PHI can also be very valuable to criminals who are intent on committing medical identity theft. The crime of medical identity theft represents the intersection of medical fraud and identity theft (Figure 1). Like medical fraud, it involves fraudulent charges and like financial identity theft, it involves the theft of identity. It is unique in that it involves a medical identity (patient identification, insurance information, medical histories, prescriptions, test results…) that may be used to obtain medical services or prescription drugs (Ball et al. 2003). Leaked insurance information can be used to fraudulently obtain service, but unlike a credit card the spending limits are much higher—charges can quickly reach tens of thousands or even millions of dollars. And unlike financial credit, there is less monitoring and reporting. Sadly, beyond the financial losses, medical identity theft carries other personal consequences for victims as it often results in erroneous changes to medical records that are difficult and time consuming to correct. Such erroneous information could impact care quality or impede later efforts to obtain medical, life, or disability insurance.

For example, recent medical identity theft cases have involved the sale of health identities to illegal immigrants (Messmer 2008). These forms of theft are a problem impacting payers, patients, and health-care providers. Payers and providers both see financial losses from fraudulent billing. Patients are also harmed when they are billed for services they did not receive, and when erroneous information appears on their medical record.

Between 1998 and 2006, the FTC recorded complaints of over nineteen thousand cases of medical identity theft with rapid growth in the past five years. Many believe these complaints represent the tip of the growing fraud problem, with some estimates showing upwards of a quarter-million cases a year (Dixon 2006, 12-13). Currently, there is no single agency tasked with tracking, investigating, or prosecuting these crimes (Lafferty 2007) so reliable data on the extent of the problem does not exist.

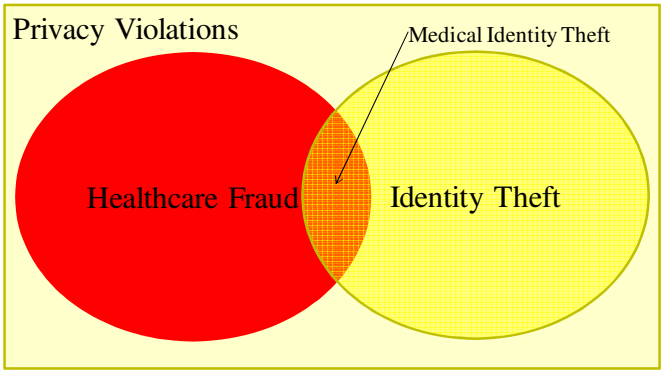


Fig. 1. Consequences of data hemorrhages.

The crime of financial identity theft is well understood with clear underlying motives. A recent FTC survey estimated that 3.7% of Americans were victims of some sort of identity theft (FTC 2007). Significant media coverage has alerted the public of the financial dangers that can arise when a thief assumes your identity. However, the dangers and associated costs of medical identity theft are less well understood and largely overlooked. Of course, PHI (including insurance policy information and government identity numbers) can be fraudulently used for financial gain at the expense of firms and individuals. However, when a medical identity is stolen and used to obtain care, it may also result in life-threatening amendments to a medical file. Any consequential inaccuracies in simple entries, such as allergy diagnoses and blood-typing results, can jeopardize patient lives. Furthermore, like financial identity theft, medical identity theft represents a growing financial burden on the private and public sectors.

Individuals from several different groups participate in the crime of medical identity theft: the uninsured, hospital employees, organized crime rings, illegal aliens, wanted criminals, and drug abusers. In many cases the theft is driven by greed, but in other case the underlying motive is simply for the uninsured to receive medical care. Without medical insurance, these individuals are unable to obtain the expensive care that they require, such as complicated surgeries or organ transplants. However, if they assume the identity of a well insured individual, hospitals will provide full-service care. For example, Carol Ann Hutchins of Pennsylvania assumed another woman's identity after finding a lost wallet (Wereschagin 2006). With the insurance identification card inside the wallet, Hutchins was able to obtain care and medication on 40 separate occasions at medical facilities across Pennsylvania and Ohio, accumulating a total bill of $16,000. Had it not been for the victim's careful examination of her monthly billing statement, it is likely that Hutchins would have continued to fraudulently receive care undetected. Hutchins served a 3-month jail sentence for her crime, but because of privacy laws and practices, any resulting damage done to the victim's medical record was difficult and costly to erase.

Hospital employees historically comprise the largest known group of individuals involved in traditional medical fraud. They may alter patient records, use patient data to open credit card accounts, overcharge for and falsify services rendered, create phony patients, and more. The crimes committed by hospital employees are often the largest, most intricate, and the most costly.

Take for example the case of Cleveland Clinic front desk clerk coordinator, Isis Machado who sold the medical information of more than 1,100 patients, to her cousin Fernando Ferrer, Jr., the owner of Advanced Medical Claims Inc. of Florida. Fernando then provided the information to others who used the stolen identities to file an estimated $7.1 million in fraudulent claims (USDC 2006).

Individuals abusing prescription drugs also have a motive to commit medical identity theft. Prescription drug addicts can use stolen identities to receive multiple prescriptions at different pharmacies. Drugs obtained through this method may also be resold or traded. Roger Ly, a Nevada pharmacist allegedly filed and filled 55 false prescriptions for Oxycontin and Hydrocondone in the name of customers. Medicare and insurance paid for the drugs that Ly, allegedly, then resold or used recreationally (USA 2007). The total value of drugs sold in the underground prescription market

likely exceeds $1 billion (Peterson 2000). Sometimes, the crimes involving prescription drugs are less serious; a Philadelphia man stole a coworker's insurance identification card to acquire a Viagra prescription, which he filled on 38 separate occasions. The plan finally backfired when the coworker he was posing as attempted to fill his own Viagra prescription and discovered that one had already been filled at another pharmacy. The cost to his company's insurance plan: over $3,000 (PA 2006).

Wanted criminals also have a strong motive to commit medical identity theft. If they check into a hospital under their own name, they might be quickly apprehended by law enforcement. Therefore, career criminals need to design schemes to obtain care. Joe Henslik, a wanted bank robber working as an ad salesman, found it easy to obtain Joe Ryan's Social Security number as part of a routine business transaction (BW 2007). Henslik then went on to receive $41,888 worth of medical care and surgery under Ryan's name. It took Ryan two years to discover that he had been a victim of medical identity theft. Even after discovery, he found it difficult to gain access to his medical records, since his own signature didn't match that of Henslik's forgery.

Anndorie Sachs experienced a similar situation when her medical identity was used to give birth to a drug addicted baby (Reavy 2006). Sachs had lost her purse prior to the incident and had accordingly cancelled her stolen credit cards, but was unaware of the risk of medical ID theft. The baby, which was abandoned at the hospital by the mother, tested positive for illegal drug use, prompting child services to contact Sachs, who had four children of her own. Fortunately, since Sachs did not match the description of the woman who gave birth at the hospital, the problem did not escalate further. If Sachs was not able to prove her identity, she could have lost custody of her children, and been charged with child abuse. Furthermore, before the hospital became aware of the crime, the baby was issued a Social Security number in Sachs name, which could cause complications for the child later in life. Like Sachs, few individuals consider their insurance cards to be as valuable as the other items they carry in their wallet. Moreover, medical transactions appearing on a bill may not be scrutinized as closely as financial transactions with a bank or credit card.

Illegal immigrants also represent a block of individuals with a clear motive to commit medical identity theft. In the case of a severe medical emergency, they will not be refused care in most instances, but if an illegal immigrant requires expensive surgery, costly prescriptions, or other non-emergency care, they have few options. One of the most shocking and well documented cases comes from Southern California, where a Mexican resident fooled the state insurance program, Medi-Cal, into believing that he was a resident and therefore entitled to health care coverage (Hanson 1994). Mr. Hermillo Meave, was transferred to California from a Tijuana, Mexico hospital with heart problems, but told the California hospital that he was from San Diego, and provided the hospital with a Medi-Cal ID card and number. Although the circumstances surrounding Mr. Meave's arrival were suspicious, the hospital went ahead and completed a heart transplant on Mr. Meave. The total cost of the operation was an astounding one million dollars. Only after the surgery did the hospital determine that Mr. Meave actually lived and worked in Tijuana and was therefore not entitled to Medi-Cal coverage.

Perhaps emboldened by the success of Hermillo Meave, a family from Mexico sought a heart transplant for a dying relative just three months later at the very same

hospital. This time, fraud investigators were able to discover the plot before the surgery could be completed. While processing the paperwork for the patient who was checked in as Rene Garcia, Medi-Cal authorities found nine other individuals around the state, using the same name and ID number. The hospital had the family arrested and jailed for the attempted fraud, which had cost the hospital $200,000, despite the lack of surgery. The family told investigators that they had paid $75,000 in order to obtain the ID and set up the surgery. The trafficking of identities between Mexico and California is commonplace, but the sale of Medi-Cal identities adds a new dimension to the crime. The disparity in care between California hospitals and Mexican facilities makes the motivation to commit medical identity theft clear: falsified identification is a low-cost ticket to world-class care.

Finally, identity theft criminals often operate in crime rings, sometimes using elaborate ruses to gather the identities of hundreds individuals. In a Houston case, criminals allegedly staged parties in needy areas offering medical deals as well as food and entertainment (USDJ 2007). At the parties, Medicaid numbers of residents were obtained and then used to bill Medicaid for alcohol and substance abuse counseling. The scheme even included fraudulent reports, written by 'certified' counselors. The fraudulent company managed to bill Medicaid for $3.5M worth of services, of which they received $1.8M. In this case, no medical care was actually administered and the medical identity theft was committed purely for financial reasons.

In summary, there are many reasons why individuals engage in medical identity theft, including avoiding law enforcement, obtaining care that they have no way of affording, or simply making themselves rich. Many tactics are used including first hand by physical theft, insiders, and harvesting leaked data. As we saw, PHI can be sold and resold before theft occurs—as in the case of the nine Garcias. The thief may be someone an individual knows well or it could be someone who they've never met.

For health-care providers, the first step in reducing such crime is better protection of PHI by: 1) controlling access within the enterprise to PHI; 2) securing networks and computers from direct intruders; 3) monitoring networks (internal and external) for PII and PHI transmissions and disclosures; 4) avoiding inadvertent disclosures of information. Often loose access and inadvertent disclosures are linked. When access policies allow many individuals to view, move, and store data in portable documents and spreadsheets, the risk of inadvertent disclosure increases.

3 Inadvertent Data Hemorrhages

Despite the much trumpeted enactment of the Health Insurance Portability and Accountability Act (HIPAA), data losses in the health-care sector continue at a dizzying pace. While the original legislation dates back to 1996, the privacy rules regulating the use and disclosure of medical records did not become effective until 2004. Moreover, the related security rules, which mandate computer and building safeguards to secure records, became effective in 2005. While firms and organizations have invested to protect their systems against direct intrusions and hackers, many recent the data hemorrhages have come from inadvertent sources. For

example, laptops at diverse health organizations including Kaiser Permanente (Bosworth 2006), Memorial Hospital (South Bend IN) (Tokars 2008), the U.S. Department of Veterans Administration (Levitz and Hechinger 2006), and National Institutes of Health (Nakashima and Weiss 2008) were lost or stolen—in each case inadvertently disclosing personal and business information.

Organizations have mistakenly posted on the web many different types of sensitive information, from legal to medical to financial. For example, Wuesthoff Medical Center in Florida inadvertently posted names, Social Security numbers and personal medical information of more than 500 patients (WFTV 2008). Insurance and health-care information of 71,000 Georgia residents was accidentally posted on Internet for several days by Tampa-based WellCare Health Plans (Hendrick 2008).

The University of Pittsburgh Medical Center inadvertently posted patient information of nearly 80 individuals including names and medical images. In one case, a patient's radiology image was posted along with his Social Security number, insurance information, medications, and with information on previous medical screenings and procedures (Twedt, 2007). Harvard University and its pharmacy partner, PharmaCare (now part of CVS Caremark), experienced a similar embarrassment when students showed they could easily gain access to lists of prescription drugs bought by Harvard students (Russell 2005). Even technology firms like Google and AOL have suffered the embarrassment of inadvertent web posting of sensitive information (Claburn 2007, Olson 2006)—in their cases, customer information. Still other firms have seen their internal information and intellectual property appear on music file-sharing networks (DeAvila 2007), blogs, YouTube, and MySpace (Totty 2007). In each case, the result was the same: sensitive information inadvertently leaked creating embarrassment, vulnerabilities, and financial losses for the firm, its investors, and customers. In a recent data loss, Pfizer faces a class action suit from angry employees who had their personal information inadvertently disclosed on a popular music network (Vijayan 2007). In this paper we examine health-care leaks from a common, but widely misunderstood source of inadvertent disclosure: peer-to-peer file-sharing networks.

In our past research, we showed that peer-to-peer (P2P) file-sharing networks represented a significant security risk to firms operating within the banking sector (Johnson and Dynes, 2007; Johnson 2008). File sharing became popular during the late 1990s with rise of Napster. In just two years before its court-ordered closure in 2001, Napster enabled tens of millions of users to share MP3-formatted song files. Through its demise, it opened the door for many new P2P file-sharing networks such as Gnutella, FastTrack, e-donkey, and Bittorrent, with related software clients such as Limewire, KaZaA, Morpheus, eMule, and BearShare. Today P2P traffic levels are still growing with as many as ten million simultaneous users (Mennecke 2006). P2P clients allow users to place shared files in a particular folder that is open for other users to search. However, there are many ways that other confidential files become exposed to the network (see Johnson et al. 2008 for a detailed discussion). For example a user: 1) accidentally shares folders containing the information—in some cases confusing client interface designs can facilitate such accidents (Good and Krekelberg (2003)); 2) stores music and other data in the same folder that is shared—this can happen by mistake or because of poor file organization; 3) downloads

malware that, when executed, exposes files; or 4) installs sharing client software that has bugs, resulting in unintentional sharing of file directories.

While these networks are most popularly used to trade copyrighted material, such as music and video, any material can be exposed and searched for including databases, spreadsheets, Microsoft Word documents, and other common corporate file formats. The original exposure of this material over P2P networks is most likely done by accident rather than maliciously, but the impact of a single exposure can quickly balloon. After a sensitive file has been exposed, it can be copied many times by virtually anonymous P2P users, as they copy the file from one another and expose the file to more peers. Criminals are known to engage in the sale and trafficking of valuable information and data. In earlier studies using "honeypot" experiments (experiments that expose data for the purpose of observing how it is stolen), we showed how criminals steal and use both consumer data and corporate information (Johnson et al. 2008). When this leaked information happens to be private customer information, organizations are faced with costly and painful consequences resulting from fraud, customer notification, and consumer backlash.

Ironically, individuals who experience identity theft often never realize how their data was stolen. While there are many ways personal health-care data can be exposed, we will show in the next section how data hemorrhages in P2P networks represent a missing link in the "causality chain." Far worse than losing a laptop or a storage device with patient data (Robenstein 2008), inadvertent disclosures on P2P networks allow many criminals access to the information, each with different levels of sophistication and ability to exploit the information. And unlike an inadvertent web posting, the disclosures are far less likely to be noticed and corrected (since few organizations monitor P2P and the networks are constantly changing making a file intermittently available to a subset of users). Clearly, such hemorrhages violate the privacy and security rules of HIPAA, which call for health-care organizations to ensure implementation of administrative safeguards (in the form of technical safeguards and policies, personnel and physical safeguards) to monitor and control intra and inter-organizational information access.

4 Research Method and Analysis

To explore the vulnerability and threat of medical information leakage, we examined health-care data disclosures and search activity in peer-to-peer file sharing networks. To collect a sample of leaked data, we initially focused on Fortune Magazine's list of the top ten publically traded health-care firms (Fortune Magazine (Useem 2007)). Together those firms represented nearly $70B in US health-care spending (Figure 2).

To gather relevant files, we developed a digital footprint for each health-care institution. A digital footprint represents key terms that are related to the firm—for example names of the affiliated hospitals, clinics, key brands, etc. Searching the internet with Google or P2P networks using those terms will often find files related to those institutions. With the help of Tiversa Inc., we searched P2P networks using our digital signature over a 2-week period (in January, 2008) and randomly gathered a sample of shared files related to health care and these institutions. Tiversa's servers

and software allowed us to sample in the four most popular networks (each of which supports the most popular clients) including Gnutella (e.g., Limewire, BearShare), FastTrack (e.g., KaZaA, Grokster), Aries (Aries Galaxy), and e-donkey (e.g., eMule, EDonkey2K). Files containing any one or combination of these terms in our digital footprint were captured. We focused on files from the Microsoft Office Suite (Word, Powerpoint, Excel, and Access). Of course, increasing the number of terms included in the digital footprint increases the number file matches found, but also increases false positives—files captured that have nothing to do with the institution in question. Given the large number of hospitals within these ten organizations (more than 500), our goal was to gather a sample of files to characterize the ongoing data hemorrhage. Since users randomly join P2P networks to get and share media (and then depart), the network is constantly changing. By randomly sampling over a 14-day period, we collected 3,328 files for further (manual) analysis.

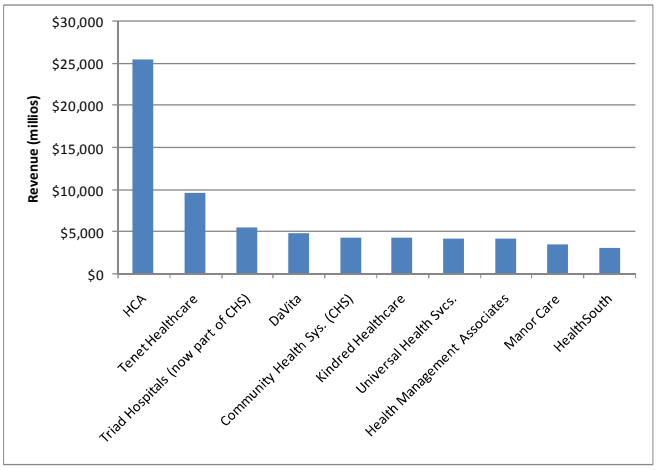


Fig. 2. Revenue of the top ten US health-care firms (Useem 2007).

Of 3,328 documents in our sample, 50.3% could be immediately identified as duplicate copies of the same file (same hash) that had spread or were on multiple IP addresses, leaving us with 1,654 documents to categorize. While duplicate files were not downloaded from the same IP address, duplicate files were collected when a target file had spread to multiple sharing clients. They were also collected from users who joined the network at different IP addresses (what we call an IP shift). Through a manual analysis of the remaining 1,654 files, we found that 71% were not relevant to health care or the organizations under consideration and were downloaded because our search terms overlapped with other subject matter. This was the result of the size and quality of our digital footprint. By casting a large net, we found more files but also many that were not related to the health-care sector. Of the remaining 475 documents, 86 were manually evaluated as duplicate files. With this cross section of

data associated with the health-care organizations, we categorized each file evaluating the dangers associated with it. Figure 3 shows a categorization of the 389 unique, relevant files.

The most common type of files found were newspaper and journal articles, followed by documents associated with students studying medicine. This should not come as a surprise as many P2P users are students. Interestingly, we found entire medical texts being shared. We also found many documents dealing directly with medical issues, such as billings, letters to hospitals, and insurance claims. Many of these documents were leaked by patients themselves. For example, we found several patient-generated spreadsheets containing details of medical treatments and costs—likely for tax purposes. Other documents discovered included hospital brochures and flyers, which were intended for public consumption. Finally there were job listings, cover letters, and résumés, all likely saved on computers of job-seekers. The lack interest in sharing these files for a typical P2P user makes it readily apparent that they were likely shared by mistake. However, all of the files weren't so innocuous. After categorizing the files, we found that about 5% of the files recovered by our loosely tuned search were sensitive or could be used to commit medical or financial identity theft.

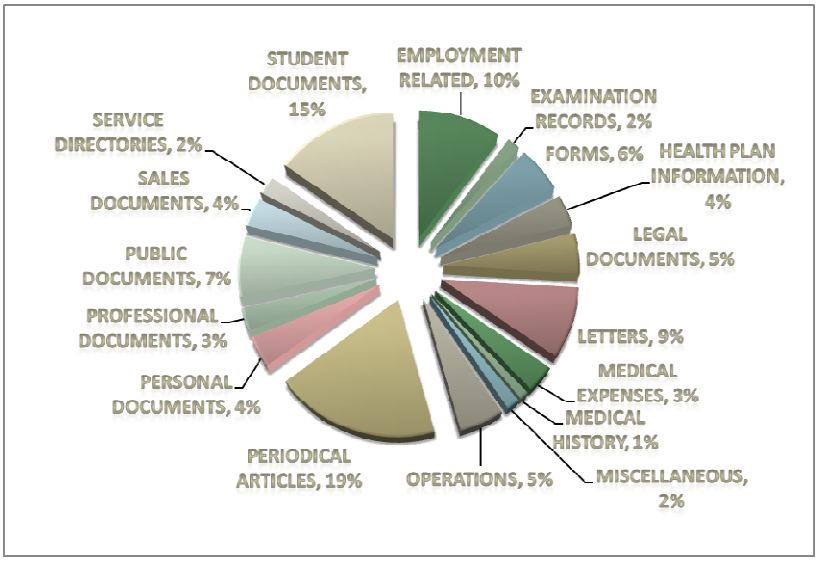


Fig. 3. Summary of unique relevant files.

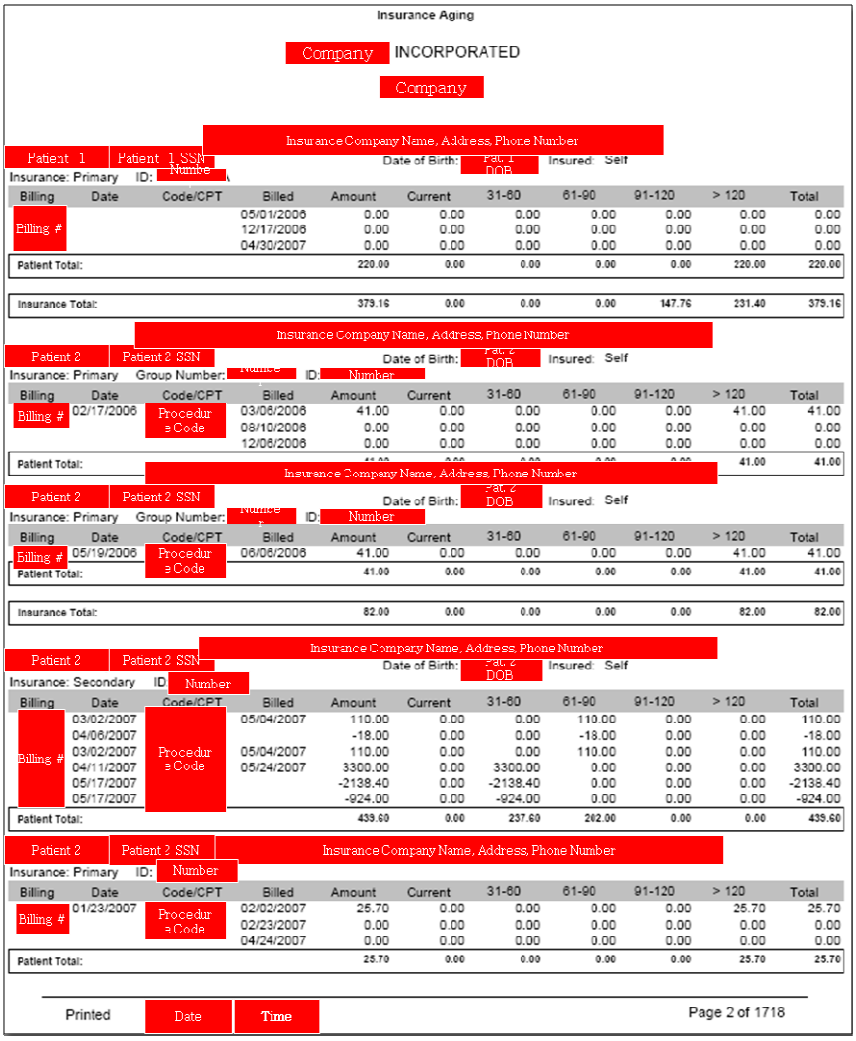
The set of dangerous documents discovered contained several files that would facilitate medical identity theft. One such document was a government application for employment asking for detailed background information. The document contained the individual's Social Security number, full name, date of birth, place of

birth, mother's maiden name, history of residence and acquaintances, schooling history, and employment history (the individual had worked at one of the hospitals under study). Despite the document's three-page forward highlighting the privacy act measures undertaken by the government to protect the information in the document, and the secure Data Hash code stamped at the bottom of every page along with the bolded text 'PRIVACY ACT INFORMATION', this document somehow ended up on to a P2P network.

More disturbing, we found a hospital-generated spreadsheet of personally identifiable information on recently-hired employees including Social Security numbers, contact information, job category etc. Another particularly sensitive document was an Acrobat form used for creating patient prescriptions. The scanned blank document was signed by a physician and allowed for anyone to fill in the patient's name and prescription information. This document could be used for medical fraud by prescription drug dealers and abusers. Additionally, the doctor's own personal information was included in the document, giving criminals the opportunity to forge other documents in his name. Finally, another example we found was a young individual's medical card. This person was suffering from various ailments and was required to keep a card detailing his prescription information. The card included his doctor's name, parent's names, address, and other personal information. A person with a copy of this identification card could potentially pose as the patient and attempt to procure prescription drugs. All of these dangerous files were found with a relatively simple sample of files published for anyone to find.

As a second stage of our analysis, we then moved from sampling with a large net to more specific and intentional searches. Using information from the first sampling, we examined shared files on hosts where we had found other dangerous data. One of the features enabled by Limewire and other sharing clients is the ability to examine all the shared files of a particular user (sometimes called "browse host"). Over the next six months, we periodically examined hosts that appeared promising for shared files.

Using this approach, we uncovered far more disturbing files. For a medical testing laboratory, we found a 1,718-page document containing patient Social Security numbers, insurance information, and treatment codes for thousands of patients. Figure 4 shows a redacted excerpt of just a single page of the insurance aging report containing patient name, Social Security number, date of birth, insurer, group number, and identification number. All together, almost 9,000 patient identities were exposed in a single file, easily downloaded from a P2P network.



Insurance Aging

Company INCORPORATED

Company

Insurance Company Name, Address, Phone Number

Patient 1 | Patient 1 SSN | Date of Birth: Pat. 1 DOB | Insured: Self

Insurance: Primary ID: Number

Billing	Date	Code/CPT	Billed	Amount	Current	31-60	61-90	91-120	> 120	Total
Billing #			05/01/2006	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			12/17/2006	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			04/30/2007	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Patient Total:				220.00	0.00	0.00	0.00	0.00	220.00	220.00
Insurance Total:				379.16	0.00	0.00	0.00	147.76	231.40	379.16

Insurance Company Name, Address, Phone Number

Patient 2 | Patient 2 SSN | Date of Birth: Pat. 2 DOB | Insured: Self

Insurance: Primary Group Number: Name ID: Number

Billing	Date	Code/CPT	Billed	Amount	Current	31-60	61-90	91-120	> 120	Total
Billing #	02/17/2006	Procedure Code	03/06/2006	41.00	0.00	0.00	0.00	0.00	41.00	41.00
			08/10/2006	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			12/06/2006	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Patient Total:									41.00	41.00

Insurance Company Name, Address, Phone Number

Patient 2 | Patient 2 SSN | Date of Birth: Pat. 2 DOB | Insured: Self

Insurance: Primary Group Number: Name ID: Number

Billing	Date	Code/CPT	Billed	Amount	Current	31-60	61-90	91-120	> 120	Total
Billing #	05/19/2006	Procedure Code	06/06/2006	41.00	0.00	0.00	0.00	0.00	41.00	41.00
Patient Total:				41.00	0.00	0.00	0.00	0.00	41.00	41.00
Insurance Total:				82.00	0.00	0.00	0.00	0.00	82.00	82.00

Insurance Company Name, Address, Phone Number

Patient 2 | Patient 2 SSN | Date of Birth: Pat. 2 DOB | Insured: Self

Insurance: Secondary ID Number

Billing	Date	Code/CPT	Billed	Amount	Current	31-60	61-90	91-120	> 120	Total
Billing #	03/02/2007	Procedure Code	05/04/2007	110.00	0.00	0.00	110.00	0.00	0.00	110.00
	04/06/2007			-18.00	0.00	0.00	-18.00	0.00	0.00	-18.00
	03/02/2007		05/04/2007	110.00	0.00	0.00	110.00	0.00	0.00	110.00
	04/11/2007		05/24/2007	3300.00	0.00	3300.00	0.00	0.00	0.00	3300.00
	05/17/2007			-2138.40	0.00	-2138.40	0.00	0.00	0.00	-2138.40
	05/17/2007			-924.00	0.00	-924.00	0.00	0.00	0.00	-924.00
Patient Total:				439.60	0.00	237.60	262.00	0.00	0.00	439.60

Patient 2 | Patient 2 SSN | Insurance Company Name, Address, Phone Number

Insurance: Primary ID: Number

Billing	Date	Code/CPT	Billed	Amount	Current	31-60	61-90	91-120	> 120	Total
Billing #	01/23/2007	Procedure Code	02/02/2007	25.70	0.00	0.00	0.00	0.00	25.70	25.70
			02/23/2007	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			04/24/2007	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Patient Total:				25.70	0.00	0.00	0.00	0.00	25.70	25.70

Printed Date Time Page 2 of 1718

Fig. 4. Excerpt of an insurance againg report. It contains 1718 pages of patient names, social security numbers, dates of birth, insurers, group numbers, and identification numbers (exposing nearly 9000 patients). Personally Identifiable Information has been redacted to protect the identities of the disclosers and patients.

For a hospital system, we found two spreadsheet databases that contained detailed information on over 20,000 patients including Social Security numbers, contact details, and insurance information. Up to 82 fields of information (see Figure 5) were recorded for each patient—representing the contents of the popular HCFA form. In this case, the hemorrhage came from an outsourced collection agency working for the hospital. However, besides the patients and hospital system, many other

1. FAFA billNumber
2. providerName
3. providerAddressLine1
4. providerCityStateZip
5. providerPhoneNumber
6. providerFederalTaxId
7. patientFirstName
8. patientMiddleInitial
9. patientLastName
10. patientSSN
11. patientPhone
12. patientAddressLine1
13. patientAddressLine2
14. patientCity
15. patientState
16. patientZipCode
17. patientSex
18. patientBirthDate
19. patientEmployerName
20. patientEmployerAddressLine1
21. patientEmployerAddressLine2
22. patientEmployerCity
23. patientEmployerState
24. patientEmployerZipCode
25. patientEmployerPhone
26. caseType
27. admissionDate
28. dischargeDate
29. patientMedRecNo
30. patientMaritalStatus
31. guarantorFirstName
32. guarantorLastName
33. guarantorSSN
34. guarantorPhone
35. guarantorAddressLine1
36. guarantorAddressLine2
37. guarantorCity
38. guarantorState
39. guarantorZipCode
40. guarantorBirthDate
41. guarantorEmployerName
42. guarantorEmployerAddressLine1
43. guarantorEmployerAddressLine2
44. guarantorEmployerCity
45. guarantorEmployerState
46. guarantorEmployerZipCode
47. guarantorEmployerPhone
48. guarantorRelationship
49. totalCharges
50. amountBalance
51. totalPayments
52. totalAdjustments
53. accidentCode
54. accidentDate
55. firstInsuranceName
56. firstInsuranceAddressLine1
57. firstInsuranceCity
58. firstInsuranceState
59. firstInsuranceZipCode
60. firstPolicyNumber
61. firstAuthorizationNumber
62. firstGroupName
63. firstGroupNumber
64. firstInsuredRelationship
65. firstDateEligible
66. firstDateThru
67. secondInsuranceName
68. secondInsuranceAddressLine1
69. secondInsuranceCity
70. secondInsuranceState
71. secondInsuranceZipCode
72. secondPolicyNumber
73. secondGroupName
74. secondGroupNumber
75. secondInsuredRelationship
76. secondDateEligible
77. secondDateThru
78. primaryDiagnosisCode
79. attendingPhysician
80. attendingPhysicianUPIN
81. lastPaymentDate
82. providerShortName

Fig. 5. File contents for over 20,000 patients in one inadvertent disclosure.

organizations were comprised. The data disclosed in this file well-illustrates the complexity of US health care with many different constituencies represented, including 4 major hospitals, 335 different insurance carriers acting on behalf of 4,029 patient employers, and 266 different treating doctors (Figure 6). Each of these constituents was exposed in this disclosure. Of course, the exposure of sensitive patient health-information may be the most alarming to citizens. Figure 7 shows one very small section of the spreadsheet (just three columns of 82) for a few patients (of the nearly 20,000). Note that the diagnosis code (IDC code) is included for each patient. For example, code 34 is streptococcal sore throat; 42 is AIDS; 151.9 is malignant neoplasm of stomach (cancer); 29 is alcohol-induced mental disorders; and 340 is multiple sclerosis. In total the file contained records on 201 patients with different forms of mental illness, 326 with cancers, 4 with AIDS, and thousands with other serious and less serious diagnoses.

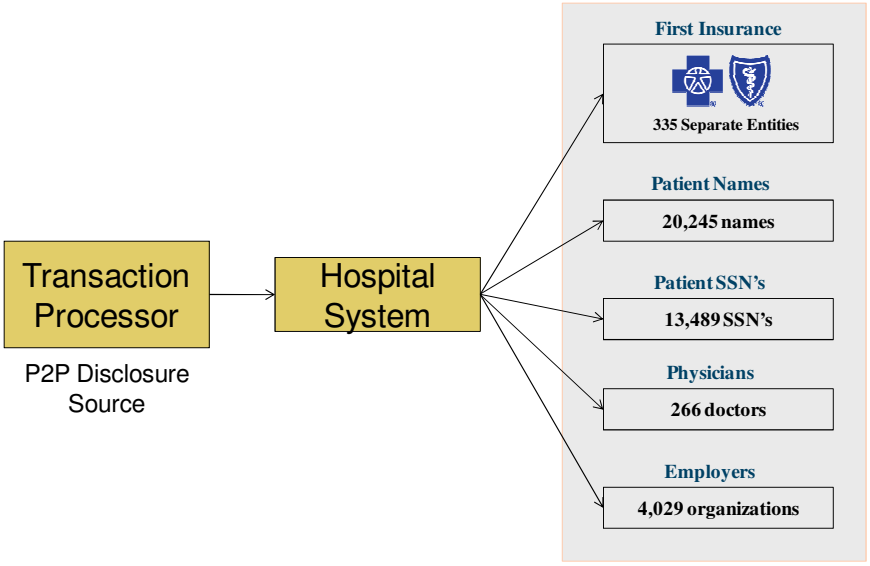
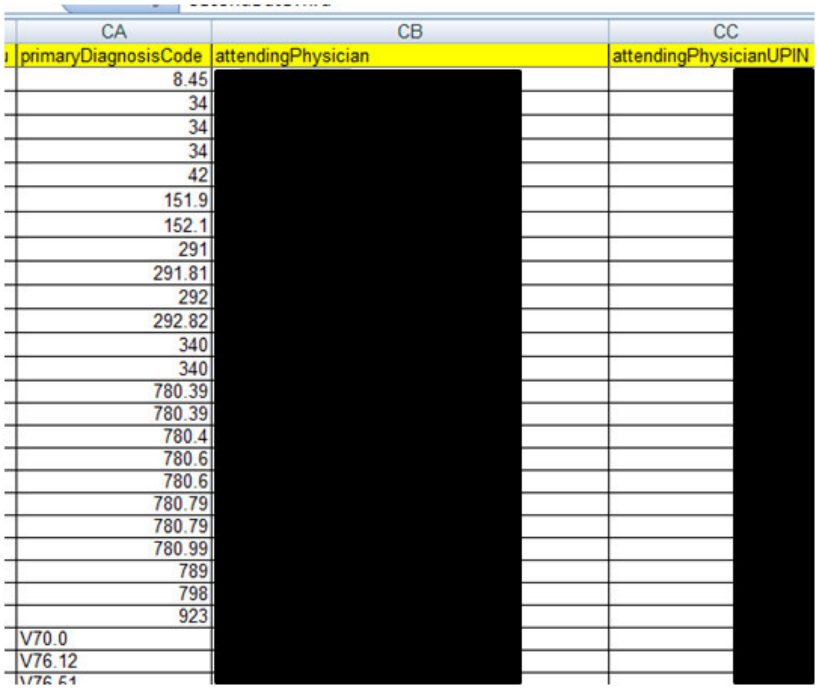


Fig. 6. Hemorrhage exposed a large array of health-care constituents.



	CA	CB	CC
	primaryDiagnosisCode	attendingPhysician	attendingPhysicianUPIN
	8.45	[redacted]	[redacted]
	34		
	34		
	34		
	42		
	151.9		
	152.1		
	291		
	291.81		
	292		
	292.82		
	340		
	340		
	780.39		
	780.39		
	780.4		
	780.6		
	780.6		
	780.79		
	780.79		
	780.99		
	789		
	798		
	923		
	V70.0		
	V76.12		
	V76.51		

Fig. 7. Disclosures expose extreamly personal diagnosis information. A very small section of a spreadsheet for a few (of over 20,000) patients showing IDC diagnosis codes (see http://www.cms.hhs.gov/ICD9ProviderDiagnosticCodes/ or http://www.icd9data.com/). Personally Identifiable Information has not been included in the illustration to protect the identities of the patients and physicians.

For a mental health center, we found patient psychiatric evaluations. All would be considered extremely personal and some were disturbing. We found similar clinical evaluations leaking from Alabama to Nebraska to California.

Of course, these are just few of many files we uncovered. For a group of anesthesiologists, we found over 350MB of data comprising patient billing reports. For a drug and alcohol rehab center, we found similar billing information. From an AIDs clinic we found a spreadsheet with 232 clients including address, Social Security number, and date of birth. And the list goes on. It is important to note that all of these files were found without extraordinary effort and certainly far less effort than criminals might be economically incented to undertake.

With the vulnerability well established, we also investigated the search activity in P2P networks to see if users were looking for health-care data hemorrhages. Again, using our simple digital signature we captured a sample of user-issued searches along with our files. Figure 8 lists a sample of these searches and clearly shows that users are searching for very specific health-care related data in P2P networks.

compudoc medical	computers medical doctors	child medical exam
care office nbc health	billing medical august	child medical release form
hospital records	canada medical test	cigna medical dr
mental hospitals	canadian medical	classified medical records
hospital	caulfield general medical	complete medical exam
hospital letterhead	certficat medical	comprehensive medical
niagara hospital	doctor medical exam	letter for medical bills
american medical	doctors office medical exam	ltr client medical report
connolly medical	doctors orders medical	ltr hjh rosimah medical
dear medical insurance my	doug medical bill	ltr medical body4life
denial of medical insurance	edimis medical software 3.9	ltr medical maternity portland
isilo medical	electronic medical	ltr orange medical head center
medical	electronic medical record	ltr to valley medical
medical claims	electronic medical record.pdf	lytec medical billing
medical exam	electronics & bio medical	medical investigation
medical history	emt medical software	medical journals password
medical passwords	forms medical	medical.txt
medical permission	forms medical liability form	medical abuse records
medical records certification	forms medical office	medical authorization
medical release	ge medical	medical autorization
authorization for medical	ge medical syatems	medical benefits
basic medical forms	medical coding and billing	medical bill
billing medical	medical coding exam	medical billing

Fig. 8. Selection of User-Issued searches related to medical.

5 Conclusion

Data hemorrhages from the health-care sector are clearly a significant threat to providers, payers, and patients. The inadvertent disclosers we found and documented in this report point to the larger problem facing the industry. Clearly, such hemorrhages may fuel many types of crime. While medical fraud has long been a significant problem, the crime of medical identity theft is still in its infancy. Today, many of the well-documented crimes appear to be committed out of medical need. However, with the growing opportunity to commit more significant crimes involving large financial rewards, more and more advanced schemes and methods, such as P2P-fueled identity theft, will likely develop. For criminals to profit, they don't need to "steal" an identity, but only to borrow it for a few days, while they bill the insurer carrier thousands of dollars for fabricated medical bills. This combination of medical fraud along with identity theft adds a valuable page to the playbook of thieves looking for easy targets. Stopping the supply of digital identities is one key to halting this type of illegal activity.

The Health Insurance Portability and Accountability Act (HIPAA) was created to protect workers' health insurance coverage when they change or lose employment. It also includes standards for the transfer of healthcare information that are designed to protect the privacy of sensitive patient medical information. The Privacy and Security Rules of HIPAA require covered entities to ensure implementation of administrative safeguards in the form of policies, personnel and physical safeguards to their information infrastructure, and technical safeguards to monitor and control intra and inter-organizational information access (Choi, et al. 2006). Those rules were phased in over time with compliance maturing nearly five years ago (Privacy Rules in April 2003 and Security Rules in April 2005). Unfortunately, recent industry reports suggest low level of HIPAA compliance related to data security and privacy (AHIMA 2006). Variations in provider implementation may also make medical identity theft more difficult to track, identify, and correct. When a patient's medical record has been altered by someone else using their ID, the process used at different providers to correct the record can be confusing for the patient. The erroneous information in the medical file may remain for years. Also people who have been victims of medical identity theft may find it difficult to even know what has been changed or added to their record. Since the thief's medical information is contained within the victim's file, it is given the same privacy protections as anyone under the act. Without the ability to easily remove erroneous information, or figure out the changes contained in a medical record, repairing the damages of medical identity theft can be a very taxing process.

In theory, HHS enforcement of HIPAA is a positive force in the fight against identity theft. It is true that institutions have been fined and required to implement detailed corrective action plans to address inadvertent disclosures of identifiable electronic patient information (HHS 2008). However, many observers note that very few cases have actually resulted in a fine. And while HIPAA could be used to prosecute offending medical professionals, which are historically the largest group of health-care fraud perpetrators, few are ever prosecuted. So it is not clear that this protection of patient identities really discourages inappropriate use of medical information or reduces the chance of hemorrhages. Better compliance with both the security and privacy rules is certainly needed. Of course, HIPAA can do little to stop patients from disclosing their medical identities voluntarily to individuals posing as health care providers, or poorly managing their own computerized documents.

Tighter controls on patient information are a good start, but consumers still need to be educated of the dangers of lost health-care information and how to secure their information on personal computers. Hospitals and others concerned with medical identity theft have begun to undertake measures in order to curb medical identity theft. One of the simplest and most effective measures put in place by hospitals is to request photo identification for admittance to the hospital. In many cases, when a request for photo identification is made, the individual will give up on obtaining care and simply leave the hospital, never to return again. Of course, this measure will likely lose its efficacy in time as criminals become aware of the change in policy. Once a few personal identifiers have been acquired, such as date of birth and Social Security number, a criminal can obtain seemingly valid photo-ID. In the future, insurance companies may need to begin issuing their own tamper-proof photo identification to help stop medical identity theft.

Finally, health-care providers and insurers must enact better monitoring and information controls to detect and stop leaks (Appari and Johnson 2009). Information access within many health-care systems is lax. Coupled with the portability of data, inadvertent disclosures are inevitable. Better control over information access governance (Zhao and Johnson 2008) is an important step in reducing the hemorrhages documented in this report.

References

1. AHIMA – The American Health Information Management Association. (2006). "The State of HIPAA Privacy and Security Compliance," last accessed on Nov. 2008, http://www.ahima.org/emerging_issues/2006StateofHIPAACompliance.pdf
2. Appari, A and M.E. Johnson (2009), "Information Security and Privacy in Healthcare: Current State of Research," forthcoming in *International Journal of Internet and Enterprise Management.*
3. Ball, E., Chadwick, D.W., Mundy, D (2003), "Patient Privacy in Electronic Prescription Transfer," IEEE Security & Privacy, March/ April, 77 – 80.
4. Bolin, J.N., Clark, L.S. (2004), "Avoiding Charges of Fraud and Abuse: Developing and Implementing an Effective Compliance Program," JONA (34:12), 546-550.
5. Bosworth, M.H. (2006), "Kaiser Permanente Laptop Stolen: Personal Data on 38,000 Members Missing," Consumer Affairs, Nov 29, http://www.consumeraffairs.com/news04/2006/11/kaiser_laptop.html
6. BW (2007), "Diagnosis: Identity Theft," Business Week, January 8, 2007.
7. Choi, Y.B., Capitan, K.E., Krause, J.S. and Streeper, M.M. (2006) "Challenges associated with privacy in healthcare industry: Implementation of HIPAA and security rules," *Journal of Medical Systems*, Vol. 30, No. 1, pp57–64.
8. Claburn, T. (2007), "Minor Google Security Lapse Obscures Ongoing Online Data Risk," Information Week, January 22.
9. De Avila, J. (2007), "The Hidden Risk of File-Sharing," Wall Street Journal, Nov. 7, D1.
10. Dixon, P. (2006), "Medical Identity Theft: The Information Crime that Can Kill You," The World Privacy Forum.
11. FBI (2007), "2006 Financial Crime Report" Federal Bureau of Investigation. [Online] 02 28, 2007. [Cited: 02 04, 2008.] http://www.fbi.gov/publications/financial/fcs_report2006/financial_crime_2006.htm.
12. FTC (2007), "2006 Identity Theft Report," Federal Trade Commission, November, 2007, last accessed on June 18, 2008, http://www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf
13. Good N.S., and A. Krekelberg (2003) "Usability and privacy: a study of Kazaa P2P file-sharing," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, April 05-10.
14. Hanson, G (1994), "Illegal Aliens Bilk Sick U.S. system," Insight on the News. April 18, 1994.
15. Hendrick, B. (2008), "Insurance records of 71,000 Ga. families made public," Atlanta Journal-Constitution, April 08. http://www.ajc.com/metro/content/metro/stories/2008/04/08/breach_0409.html
16. HHS (2008), "HHS, Providence Health & Services Agree on Corrective Action Plan to Protect Health Information," U.S. Department of Health & Human Services, News Release, July 17, http://www.hhs.gov/news/press/2008pres/07/20080717a.html

17. Johnson, M. E. and S. Dynes (2007), "Inadvertent Disclosure: Information Leaks in the Extended Enterprise," Proceedings of the Sixth Workshop on the Economics of Information Security, Carnegie Mellon University, June 7-8.
18. Johnson, M. E. (2008), "Information Risk of Inadvertent Disclosure: An Analysis of File-Sharing Risk in the Financial Supply Chain," Journal of Management Information Systems, Vol. 25, No. 2, 97–123.
19. Johnson, M. E., D. McGuire, and N. D. Willey (2008), "The Evolution of the Peer-to-Peer File Sharing Industry and the Security Risks for Users," Proceedings of HICSS-41, International Conference on System Sciences, IEEE Computer Society, Jan 7-10, Hawaii.
20. Johnson, M. E., McGuire, D., and N. D. Willey (2009), "Why File Sharing Networks Are Dangerous," Communications of the ACM, 52, 2, 134-138.
21. Lafferty, L (2007), "Medical Identity Theft: The Future Threat of Health Care Fraud Is Now," Journal of Health Care Compliance; Jan/Feb, 9, 1, 11-20.
22. Levitz, J. and J. Hechinger (2006), "Laptops Prove Weakest Link in Data Security," Wall Street Journal, March 26.
23. Mennecke, T. (2006), "Slyck News – P2P Population Continues Climb," June 14, http://www.slyck.com/news.php?story=1220 .
24. Messmer, E. (2008), "Health Care Organizations See Cyberattacks as Growing Threat," Network World, February 28.
25. Musco, T. D. and K. H. Fyffe (1999), "Health Insurers' Anti-fraud Programs," Washington D.C. Health Insurance Association of America.
26. Nakashima, E. and R. Weiss (2008), "Patients' Data on Stolen Laptop," Washington Post, March 24, A1.
27. Olson, P. (2006), "AOL Shoots Itself in the Foot," Forbes, August 8.
28. PA (2006), "Pennsylvania Attorney General. Attorney General's Insurance Fraud Section charges former SEPTA employee with using co-worker's ID to obtain Viagra." Harrisburg: s.n., July 6, 2006.
29. Peterson, M. (2000), "When Good Drugs Go Gray; Booming Underground Market Raises Safety Concerns," The New York Times, 12 14, 2000, p. 1.
30. Reavy, P. (2006), "What Baby? ID victim gets a jolt," Deseret News (Salt Lake City). May 2, 2006.
31. Robenstein, S. (2008), "Are Your Medical Records at Risk?" Wall Street Journal,
32. Russell, J. (2005), "Harvard fixing data security breaches: Loophole allowed viewing student prescription orders" Boston Globe, January 22.
33. Tokars, L. (2008), "Memorial Hospital loses laptop containing sensitive employee data," WSBT, Feb 7, http://www.wsbt.com/news/local/15408791.html
34. Totty, M. (2007), "Security: How to Protect Your Private Information," Wall Street Journal, January 29. R1.
35. Twedt, S. (2007), "UPMC patients' personal data left on Web," Pittsburgh Post-Gazette, April 12.
36. USDC (2006), "United States of America vs. Fernando Ferrer, Jr. and Isis Machado," 06-60261, s.l., United States District Court Southern District of Florida, September 7, 2006.
37. USDJ (2007), "US Department of Justice. Six Indicted for Health Care Fraud Scheme in Southeast Texas," Houston, TX : s.n., 2007. Press Release.
38. USA (2007), "United States Attorney, District of Nevada. "Las Vegas Pharmacist Charged with Health Care Fraud and Unlawful Distribution of Controlled Substances," Las Vegas, United States Department of Justice, 2 23, 2007.
39. Useem, J. (2007), "Fortune 500: The Big Get Bigger," Fortune Magazine, 155, 8, April 30, 81.Wall Street Journal, March 26.
40. Vijayan, J. (2007), "Personal data on 17,000 Pfizer employees exposed; P2P app blamed," Computer World.

http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9024491
41. Wereschagin, Mike (2006), "Medical ID Theft Leads to Lengthy Recovery." Pittsburgh Tribune-Review, 10 24, 2006.
42. WFTV (2008), "Medical Center Patient Records Posted On Internet," August 14, http://www.wftv.com/news/17188045/detail.html?taf=orlc
43. Zhao, X. and M. E. Johnson (2008), "Information Governance: Flexibility and Control through Escalation and Incentives," Proceedings of the Seventh Workshop on the Economics of Information Security, Dartmouth College, June 26-27.