Healthcare Data Hemorrhages:

Inadvertent Disclosure and HITECH

Presented at *IEEE Symposium on Security and Privacy*, May 16-19, 2010, Oakland CA.

M. Eric Johnson
Dartmouth College
Tuck School of Business
Center for Digital Strategies
Hanover, NH USA
m.eric.johnson@dartmouth.edu

Nicholas Willey
Dartmouth College
Tuck School of Business
Center for Digital Strategies
Hanover, NH USA
nicholas.willey@dartmouth.edu

***Abstract*—** Hemorrhages of confidential patient health data create privacy and security concerns. While the US HIPAA legislation on privacy and security went into effect over five years ago, healthcare information security remains a significant concern as organizations migrate to electronic health records. The recent HITECH legislation aimed at accelerating this migration contained mandates for greater security, including the addition of new requirements on breach reporting. We examine a recently collected sample of inadvertently disclosed files found on internet-based file-sharing networks. We characterize the security risk of these files and also present evidence of the threat by analyzing user-issued searches. Our analysis indicates that the threat and vulnerability for the healthcare sector continued, even as HITECH became effective.

Keywords-Healtcare information, data leaks, security

EXTENDED ABSTRACT

Inadvertent disclosures of private customer information have occurred in nearly every industry from banking to healthcare. Such leaks directly impact customers through embarrassment, fraud, and identify theft. In the healthcare sector, data hemorrhages have multiple consequences [1]. In some cases, the losses translate to privacy violations and social stigma. In other cases, criminals exploit the information to commit fraud or medical identity theft. The fragmented nature of the US healthcare system results in data hemorrhages from many different sources including acute-care hospitals, physician groups, ambulatory healthcare providers, 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 examine the recent Health Information Technology for Economic and Clinical Health (HITECH) legislation and its potential impact on the security of protected health information (PHI). HITECH was enacted as part of the 2009 American Recovery and Reinvestment Act (ARRA) to spurn the adoption of Electronic Health Records (EHR). The act earmarked $20 Billion dollars to be used as incentives and investments in the creation of a digital health information infrastructure. The act also followed up on earlier HIPAA legislation to enhance privacy and security rules, which became effective in 2003 and 2005 respectively. It expanded the breach notification process by extending the parties covered under HIPAA, i.e. care providers and insurers, to include their business associates. It also defined new conditions and penalties for noncompliance. According to US Health and Human Services (HHS) guidance, affected individuals must be notified that a breach has occurred within 60 days after the discovery of the breach. If the HIPAA party does not have contact information for an individual, then they must post the breach on their website or make media notifications (local newspaper, television station, etc.). If a breach effects more than 500 people, state media and government notifications are required.

We examine the impact of HITECH on availability of hemorrhaged PHI on internet-based file-sharing networks. In earlier work [1], we showed that inadvertent disclosures of medical information collected in 2008 made PHI readily available on P2P file-sharing networks. We found leaks throughout the health chain including care providers, laboratories, and financial partners. In one case involving an AIDS clinic in Chicago that was leaking patient data, impacted individuals who experienced fraud and social stigma have recently filed a class action lawsuit against the clinic [2]. In another study, El Emam et al.[3] estimated that 0.4% of Canadian and 0.5% of US IP addresses exposing documents and spreadsheets on P2P networks, leaked ones containing PHI. Their results show that, with tens of millions of simultaneous P2P users in North America, the exposed PHI at any given point of time is substantial.

We begin by surveying recently reported healthcare data losses and the fraud that data fuels. Next we briefly summarize HITECH and its implications for data security. Then we turn to an analysis of inadvertent data hemorrhages. Given the September 23, 2009 effective date of the new HITECH breach reporting requirements, we sampled healthcare related files both before (August) and immediately after (October) the effective date to ascertain the availability the leaked PHI. We present an analysis of thousands of files we found. These files were published in peer-to-peer file-sharing networks like Limewire, eDonkey, and Bearshare and could be downloaded by anyone searching for them. The files found included sensitive patient

correspondence, business documents, and PHI-laden spreadsheets (Figure 1).

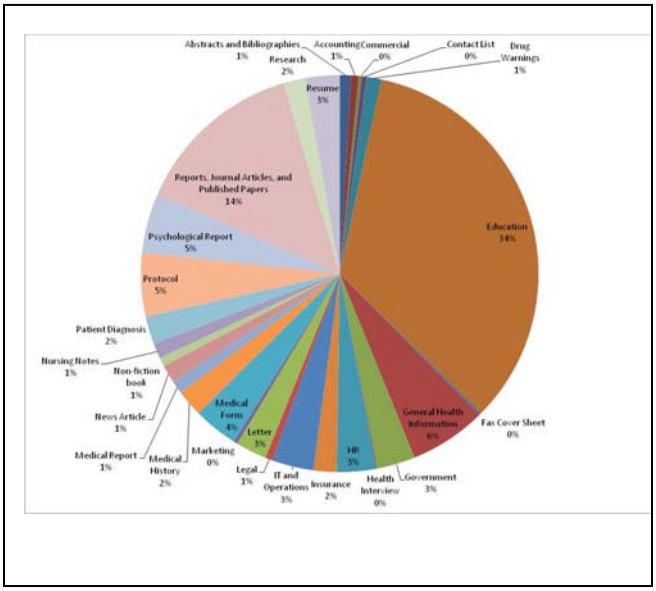

Figure 1. Categorization of files found..

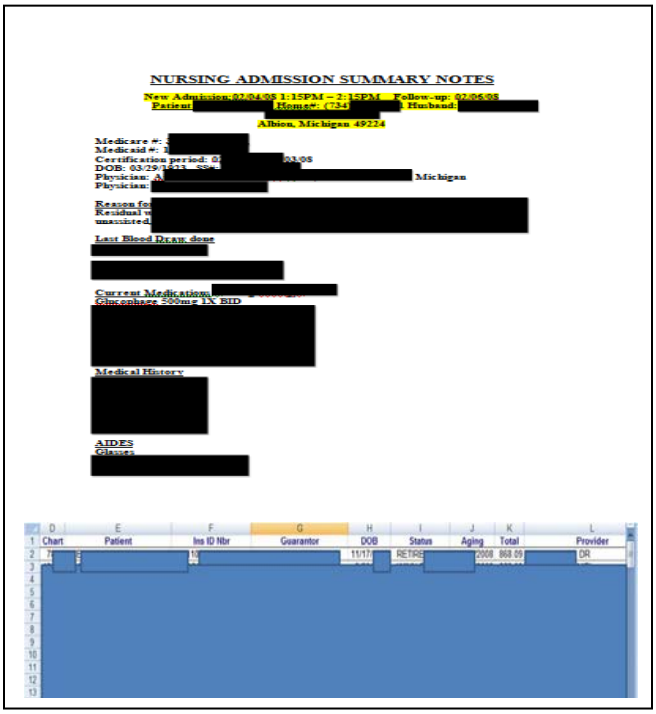

Figure 2. Redacted examples of document and spreadsheet with PHI.

We found files from healthcare firms that contained private employee and patient information for thousands of individuals, including addresses, Social Security Numbers, birth dates, and treatment information (Figure 2). We also found private patient information including medical diagnoses and psychiatric evaluations. Besides our analysis of files, we also present evidence from user-issued searches on these networks that demonstrate the threat to leaked medical data (Figure 3). We conclude with a brief discussion on reducing these inadvertent data hemorrhages.

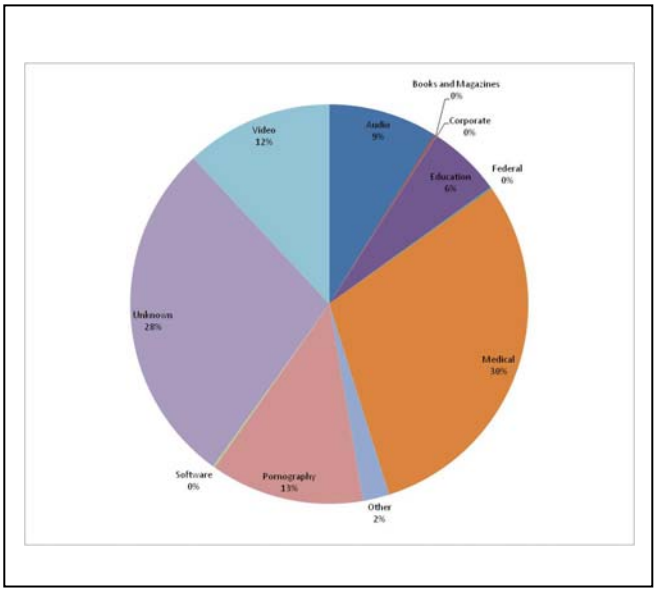

Figure 3. Categorization of user-issued searches..

ACKNOWLEDGMENT

This research was partially supported by the National Science Foundation, Grant Award Number CNS-0910842, under the auspices of the Institute for Security, Technology, and Society (ISTS). 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..

REFERENCES

[1] Johnson, M. Eric (2009) "Data Hemorrhages in the Health-Care Sector," *Lecture Notes in Computer Science*, R. Dingledine and P. Golle (Eds.): FC 2009, *LNCS 5628*, 71–89, ICFA/Springer-Verlag Berlin Heidelberg.

[2] "John Doe et al. vs. Open Door Clinic of Greater Elgin, an Illinois Corporation" 2010.

[3] El Emam, Khaled, Emilio Neri, Elizabeth Jonker, Marina Sokolova, Liam Peyton, Angelica Neisa, Teresa Scasa (2010), "The Inadvertent Disclosure of Personal Health Information through Peer-to-peer File Sharing Programs." *Journal of the American Medical Informatics Association*, 17: 148-15.