An Analysis of HIPAA Breach Data

Patrick Morrison, Laurie Williams
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
{pjmorris, lwilliams} @ncsu.edu

Abstract—As software developers, we have a responsibility to protect our user's data. When this data is protected health information (PHI), breaches can have serious financial and reputational consequences. The goal of this research is to analyze trends in breaches of PHI that point to software design guidelines that can prevent or lessen the impact of breaches. We examine the US Office of Civil Rights public data on HIPAA breach notifications and examine some of its implications for software system design We observe that a significant number of breaches involve the use of portable stores of unencrypted records and present software design strategies to address these breaches before they happen.

Keywords-component; HIPAA; security; privacy; software design

I. INTRODUCTION

"*However beautiful the strategy, you should occasionally look at the results*." – Winston Churchill

Organizations that use software in regulated environments are responsible for following all applicable regulations. Failure to do so often carries financial, civil and even criminal penalties. For example, in the United States, healthcare organizations must comply with the Health Insurance Portability and Accountability Act (HIPAA) [1] . These regulations stipulate that failure to protect personal health information can lead to fines of up to $50,000 per violation and imprisonment for up to one year.

For example, in a recent incident[2], Blue Cross Blue Shield of Tennessee agreed to pay $1.5 million to settle violations involving the loss by theft of 57 unencrypted hard drives containing over 1 million individual's protected health information (PHI). PHI is "individually identifiable health information", including past, present physical and mental health conditions, records of health care provision and related payment records[3].

As part of the HITECH act of 2009, organizations experiencing exposure of PHI must report the incident, known as a 'breach', to the department of Health and Human Services[4]. When a breach affects 500 or more individuals, the organization must submit a breach report that is posted for public display on an Office of Civil Rights web page.

We analyzed this breach data in order to understand common problems faced in the practice of EHR security and to develop responses to these problems that can be applied in the software design phase. This paper describes the breach data, summarizes our analysis and offers recommendations for software system design based on our findings.

II. HIPAA BREACH DATA AND ANALYSIS

The large breach notification data is available online[5]. There were 392 breach reports dating between September 2009 and November 2011 available when we did the analysis. We focus on three reported fields; 'Type of Breach' (Type), 'Location of Breach' (Location), and 'Summary'. Type characterizes the nature of the breach, e.g. 'Theft', 'Hacking/IT Incident', 'Loss'. Location indicates the device or medium involved in the breach, e.g. 'Laptop', 'Email' or 'Backup tape'. The Summary field was provided for 102 of the breach reports, and provides several sentences in which the breach and breach response were described. In some cases, a single breach report lists multiple Types and/or Locations. We have counted each Type and Location as a distinct entity, so one or more locations and one or more types may describe a single breach.

After identifying each distinct Type and Location we identified 461 Breach/Type/Location instances. Table 1 presents instance counts by Type and Location. Both Type and Location are sorted from highest number of occurrences to lowest. The Summary field reported 77 instances of encryption status; the data was encrypted in four of these cases. From this data we infer that breached data is largely unencrypted.

[1] Pub. L. No. 104-191, 110 Stat. 1936 (1996)
[2] http://www.hhs.gov/news/press/2012pres/03/20120313a.html
[3] 45 C.F.R., Section 160.103
[4] HITECH act, section 13402(e)(4)
[5] http://www.hhs.gov/ocr/privacy/hipaa/administrative/breachnotificationrule/breachtool.html

Table 1

Breach Count Type/Location	Theft	Unauthorized Access/Disclosure	Loss	Hacking/ IT Incident	Improper Disposal	Unknown	Other	
Paper	27	46	13		20	2	1	109
Laptop	87	5	4	1				97
Computer	51	10	4	10		1		76
Other Portable Electronic Device	40	2	25			1		68
Network Server	16	18	1	21	1			57
Other	4	7	11			1	1	24
Email	2	6		2				10
Electronic Medical Record		4		1		1		8
X-ray fim	3				2			5
Backup tape	1		3					4
Compact Disc	1		1					2
Hard drive	1							
Total	235	98	62	35	23	6	2	461

We assumed that the breach data would be a rich source of information on attacks used against digital systems. What we observed were much more basic attack approaches, summarized in the following observations:

- Portability is both weakness and strength. Every Location category was subject to Theft, but the more portable the device (or media), the greater the likelihood of a breach, whether by Theft, Loss, or Improper Disposal. These three categories combined account for almost 70% of breach instances.
- If it can be printed, it can be breached. The most common media breached across Types was Paper, with 24% of breach instances (109/461).
- Encryption is almost non-existent in reported breach instances.

III. RECOMMENDATIONS

Considering the nature of HIPAA breaches experienced in practice suggests several strategic responses:

- Disable the print button by default. Analyzing system use cases and workflow from the perspective of a paperless environment is likely to have benefits beyond data privacy management, as it seems likely to streamline health provider workflow by not requiring a trip to the printer.
- Encrypt PHI at the time of its creation. This requires analysis of key management and distribution needs for each system role during the design phase. Segregation of data fields within storage schemas can provide further 'defense in depth' by restricting the amount and kind of data available at each point during processing.,
- Limit the amount of data stored on portable devices to what will be consumed by their users in a typical workday or week. This will require analysis and design of use cases and workflows that accounts for and limits the size of the 'buffer' of available PHI records.

An extensive set of standards and certifications has begun to develop around the security of PHI. We believe that supplementing these standards with observations of how system security and data privacy are breached in practice offers useful information for the next generation of software design.

ACKNOWLEDGMENTS *(HEADING 5)*

The authors would like to thank the members of the Realsearch group for their reviews and comments on early drafts of this work.