INTERHACK

Interhack
5 E Long St 9th Fl
Columbus, OH 43215

VOX +1 614 545 HACK
FAX +1 614 545 0076
WEB http://web.interhack.com/

Using Science to Combat Data Loss: Analyzing Breaches by Type and Industry

C. Matthew Curtin, CISSP and Lee T. Ayres, CISSP

Distribution: Unlimited
Mandatory Access Control Code: PUBLIC/1/3
Date: 2009-04-21 00:01:00

Abstract

Where should defenses be deployed? Security managers can answer the question by knowing what types of breaches there are, and the rates that they occur relative to one another. A number of methods for determining such rates have been proposed with a view to helping with this decision making. Unfortunately, such methods sometimes tend towards anecdote, might be part of a marketing campaign, or lack the context needed to drive informed decisions.

We propose a taxonomy to classify incidents of the loss of control over sensitive information. The taxonomy is hierarchical in nature, allowing classification of incidents to a level of precision appropriate to the amount of information available. Analysis of incidents using the taxonomy may also work with the precision appropriate given the question at hand and data available. We then explore the proportion of breach types in a subset of data losses accumulated by the Identity Theft Resource Center (ITRC). Using the 2002 North American Industry Classification System (NAICS), we classify breach events according to the industry sector in which they occurred.

We conclude that the taxonomy is useful and that analysis of incidents by type and industry yields results that can be instructive to practitioners who need to understand how and where breaches are actually occurring. For example, the *Health Care and Social Assistance* sector reported a larger than average proportion of lost and stolen computing hardware, but reported an unusually low proportion of compromised hosts. *Educational Services* reported a disproportionately large number of compromised hosts, while insider conduct and lost and stolen hardware were well below the proportion common to the set as a whole. *Public Administration*'s proportion of compromised host reports was below average, but their share of processing errors was well above the norm. The *Finance and Insurance* sector experienced the smallest overall proportion of processing errors, but the highest proportion of insider misconduct. Other sectors showed no statistically significant dif-

ference from the average, either due to a true lack of variance, or due to an insignificant number of samples for the statistical tests being used.

Contents

I A Taxonomy of Data Losses

II Analysis of Three Years of Breach Reports by Breach Type and Industry

Part I
A Taxonomy of Data Losses

1 Introduction

The protection of personal information has become serious business. Over the last few years, consumers have become increasingly aware of the risk of fraud from a third party's misuse of personal information ("identity theft"). In response to this risk, state legislatures have enacted "breach notification" laws that require information about the loss of personal information to be reported to the affected parties. Meanwhile, organizations with personal information about their customers, employees, and business partners are trying to understand the most effective ways to protect personal information.

Privacy, risk, and security officers need to understand the means by which sensitive personal information is lost. With breaches now often being reported publicly, many details are available, but no clear picture of what is really happening emerges. Worse, advocates of one defense over another cherry-pick the data to support their causes. Before we can solve the problem of data loss, we need to understand just what the problem is. In particular, we want to know how data loss happens and where data loss happens.

Our objective is to create a simple means of classifying the loss of sensitive information so that such losses can be assessed even when details are sketchy. Knowing the general means by which the loss occurred and the context of the loss is often achievable even with the limited amount of information published in data loss notifications. Ultimately, the methodology developed should help organizations do a better job of maintaining control of the sensitive information in their care.

Information security programs are put in place by organizations of all types with the hope that these programs will properly protect and manage information, thus supporting the trustworthiness of the organizations' brands. At the heart of many programs is a list of controls,[1] These are often specified in standards such as the ISO/IEC 27001:2005[2] or ISO/IEC 27002:2005,[3]

[1] By "controls," we refer to mechanisms to protect against a weakness or vulnerability. Administrative defenses such as policy and education are included, as are technical defenses like cryptography and network firewalls.

[2] ISO. Information technology—security techniques—information security management systems—requirements. International Standard ISO/IEC 27001, 2005a

[3] ISO. Information technology—security techniques—code of practice for information security management. International Standard ISO/IEC 27002, 2005b

Other applicable standards are found in regulations such as the Health Insurance Portability and Accountability Act (HIPAA) Security Rule[4] and the Safeguards Rule[5] for the Gramm-Leach-Bliley Financial Modernization Act of 1999 (GLBA).[6]

Management of information security often becomes a matter of assessing risks against relevant standards and remediating "control gaps."[7] The efficacy of the controls is sometimes taken as a matter of gospel by inexperienced practitioners. Even where practitioners consider the importance of some controls over others, decisions often rely on anecdote and the experience of individual practitioners.

The management and protection of personal information is, in many cases, a matter broader than the visibility and responsibility of an organization's information security department, falling into the domain of a chief privacy officer or similar official. Here, we present a taxonomy of incidents resulting in the potential exposure of personal information, intended to help privacy officials classify failures to protect personal information so they can be studied, compared, and prevented in the future.

1.1 Related Work

Some attempts have been made to focus deployment of controls into areas of greatest value. Since the time of the 1988 Morris Worm, earnest attempts have been made to understand security on the Internet.[8] These attempts have focused generally on incidents reported to the Computer Emergency Response Team Coördination Center (CERT/CC) at the Software Engineering Institute at Carnegie Mellon University.

Other research has been conducted in the area of vulnerability. The most comprehensive collection comes from CERT/CC, which has published statistics, such as vulnerability remediation, incident reports received, and vulnerability advisories published.[9] Some of this work has been taken over by the United States Computer Emergency Readiness Team (US-CERT). Similar work has been undertaken to study Internet privacy, particularly Web privacy, which has helped to show how systems fail and how to avoid repeating the mistakes of the past.[10]

Broader attempts to manage risk through focus of controls in organizations, without regard to Internet connectivity, have also been made. Examples include the National Institute of Standards and Technology's (NIST) Risk Management method, which puts controls deployment in the context of exploit probability and impact around threat-vulnerability pairs,[11] and the Operationally

[4] Department of Health and Human Services. Health insurance reform: Security standards; final rule. In *Federal Register*, volume 68. U.S. National Archives and Records Administration, February 2003. [online] http://www.cms.hhs.gov/SecurityStandard/Downloads/securityfinalrule.pdf

[5] Federal Trade Commission. Standards for safeguarding customer information; final rule 16 cfr part 314. In *Federal Register*, volume 67. U.S. National Archives and Records Administration, May 2002

[6] GLBA. Gramm-Leach-Bliley Act. PUBLIC LAW 106-102, 1999. [online] http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=106_cong_public_laws&docid=f:publ102.106

[7] "Control gaps" refer to the differences between the controls in place within an organization and the framework against which the organization is being assessed.

[8] Thomas A. Longstaff, James T. Ellis, Shawn V. Hernan, Howard F. Lipson, Robert D. Mcmillan, Linda Hutz Pesante, and Derek Simmel. Security of the internet, 1997. [online] http://www.cert.org/encyc_article/tocencyc.html

[9] CERT/CC. Full statistics, January 2008. [online] http://www.cert.org/stats/fullstats.html

[10] Matt Curtin. *Developing Trust: Online Privacy and Security*. Apress, November 2001

[11] Gary Stoneburner, Alice Goguen, and Alexis Feringa. Risk management guide for information technology systems. NIST SP 800-30, July 2002. [online] http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf

Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE^{SM}) Method's assessment of infrastructure vulnerabilities as they relate to critical assets.[12] Even with such focus, organizations often choose which controls to use based, not on objective assessment of actual incidents, but solely on the persuasiveness of those making the arguments.

Additionally, studies have been executed to analyze security incidents within particular industries. For example, Adam Dodge has produced reports that examine the information security incidents that have occurred at colleges and universities for the years 2006[13] and 2007.[14]

1.2 *Analysis of Actual Data Control Loss Incidents*

We propose that analysis of data control loss incidents should be made more formal. Most significantly, we propose consideration of data control loss incidents over anecdotes and hypothetical weakness. We have a variety of questions we would like to be able to assess scientifically:

1. What types of failures occur?
2. What are the failure rates of various controls?
3. What is the impact of their failure?
4. Are lists of controls produced for one industry relevant to another?
5. Are controls used to protect one type of information useful to protect another?

Our taxonomy seeks to establish a common language needed to discuss these questions.

[12] Christopher J. Alberts, Audrey J. Dorofee, and Julia H. Allen. Octave(sm) catalog of practice, version 2.0, October 2001. [online] http://www.cert.org/archive/pdf/01tr020.pdf

[13] Adam Dodge. Educational Security Incidents: Year In Review—2006, 2007. [online] http://www.adamdodge.com/esi/yir_2006

[14] Adam Dodge. Educational Security Incidents: Year In Review—2007, February 2008. [online] http://www.adamdodge.com/esi/year_review_2007

2 Method

Assessment of control loss failure requires a set of data with sufficient information to allow the assessor to understand which control failed. Failure of a particular control is better understood when put into a specific context, such as the nature of the organization that suffered the failure.

2.1 Classification of Failure

We have focused our present work specifically on failures that result in the loss of control over sensitive information, and on understanding how such failures vary among industries. We begin with a hierarchical view of the types of failure that lead to loss of control over such information. This hierarchy allows for accurate representation of control failures even given information of variable precision, as would likely be the case where various organizations are self-reporting. Although some losses have historically been reported directly and openly, others are described in vague terms.

We break all failures into three categories: physical, logical, and procedural. From there, we specialize further.

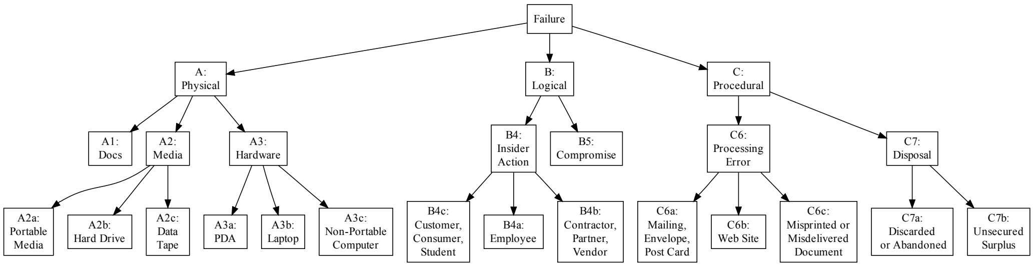

Figure 2.1: Diagram of Data Control Loss Taxonomy

A Physical Failure

Failures in the physical category are those in which control was unintentionally lost over a physical asset containing sensitive information.

A1 Documentation

Loss of control over documentation includes paper or other physical representations of sensitive information. This could come about from a physical break-in and theft of documents.

A2 Media

Loss of control over media is where the data are in electronic form for use by a computing device, but where the computing device is not part of what was lost.

A2a Portable Media

The portable media category addresses readily-accessible media such as CD-ROMs. Someone with basic computer-usage skills would likely be able to read the files from the device.

A2b Hard Drive

In this category, we refer to fixed hard disk drive (HDD) storage. In this category, the HDD is separate from the computer itself.

A2c Data Tape

Tapes are separate from portable media because reading a tape generally requires special equipment and expertise. While portable media is intended for easy interoperability among systems, tapes are optimized for writing and restoration from a known environment; interoperability is not generally a design concern.

A3 Hardware

This category includes all types of computing devices with sensitive data on their connected storage facilities.

A3a Portable Digital Assistant (PDA)

PDAs are the most mobile of computing devices. The category includes smartphones. Generally, these devices have limited storage and computing power, as well as limited security. Information on them tends to be readily accessible.

A3b Laptop

Laptop computers tend to have significant storage potential—dozens of gigabytes at least—and computing power. Security can be better than on PDAs, but in practice, cryptographic controls are rarely used and the information is easily accessible. Passwords typically are in place, making access to the data a matter of reading the disk from another machine.

A3c Non-Portable Computer

Any other computer not designed for mobility falls into this category. This includes desktops and servers.

B Logical Failure

Logical failures are those where access to sensitive information was granted through intentional action, but without giving access to the physical asset housing the data. Exposure of the sensitive information might or might not have been the objective of the action.

B4 Insider Action

This category includes instances where someone with legitimate access intentionally abuses access to sensitive information, thus causing a loss of control.

B4a Employee

Generally, the most damaging type of failure occurs where control over sensitive information is lost by means of a person in a trusted position within the organization—a direct employee.

B4b Contractor, Partner, Vendor

Control failures that come from outsiders who have a business relationship with the organization fall into this category. Thus, a degree of access is likely proper and supported by contractual agreements.

B4c Customer, Consumer, Student

Finally, outsiders with proximity to the organization can sometimes cause control failures.

B5 Compromise

We identify a loss of control of sensitive information that results from the exploitation of a vulnerability in an information system as a compromise.

C Procedural Failure

Procedural failures result from a data custodian mishandling sensitive information, publishing it to an inappropriate audience.

C6 Processing Error

Legitimate and normal business activity can lead to errors that result in a loss of control over sensitive information. These exposures fall into the category of processing errors.

C6a Mailing, Envelope, Post Card
Sensitive information can be exposed by printing it in a visible location, such as on a post card, on the outside of an envelope, or on a part of a document visible through an envelope window.

C6b Web Site
When sensitive information is published in a file on a Web site, leaving it open for download, we classify an exposure to be in this category. In this case, the information is readily available and is subject to indexing, caching, and archival by third parties.

C6c Misprinted or Misdelivered Document
Sending a document by fax to the wrong party, putting a document in the wrong envelope, or misdirecting email sometimes exposes sensitive information.

C7 Disposal

Improper disposal of information or the media that store it lead to exposure of this type.

C7a Discarded or Abandoned
This category includes exposure resulting from instances where sensitive information is carelessly “thrown away.” This includes records thrown into the trash without first being shredded.

C7b Unsecured Surplus
Computer equipment being released though a surplus process that does not include “sanitization” of the media can lead to exposure of sensitive information. Desks and filing cabinets containing paperwork with sensitive information also fall into this category.

2.2 Classification of Industry

Understanding how failures relate to one another comes from understanding the context of those failures. Context is easily determined by the industry of the affected organization, which allows for comparisons among industries. Efficacy of controls, or groups of controls, can be compared; as different industries manage information differently, lessons in the use of controls learned early in one industry—say, financial services—could well be applied to others, such as health care. Other studies could be possible by making classifications within a particular industry, including the effects of regulation.

The United States used the Standard Industrial Classification (SIC) system since the 1930s to analyze business activity in the U.S. economy and to make comparisons among industries.[1] That system has since been replaced by the North American Industry Classification System (NAICS).[2] The U.S. Economic Classification Policy Committee, Statistics Canada, and Mexico's Instituto Nacional de Estadistica, Geografia e Informatica jointly developed the standard.

Hierarchical classifications of industries, like the NAICS, could be beneficial in allowing for analysis of data at various levels of precision: high-level views could be established based on the top-level (two-digit) NAICS classification, and more detailed industry views (full six-digit) would allow for comparison of different sections of an industry. Variable precision would also be advantageous, as excessive precision would reduce the number of observations per classification down into statistical insignificance.

[1] Esther Pearce. History of the standard industrial classification. Washington, D.C., Executive Office of the President Office of Statistical Standards, U.S. Bureau of the Budget, July 1957. [online] http://www.census.gov/epcd/www/sichist.htm

[2] Office of Management and Budget. North american industry classification system: Revision for 2007; notice. In *Federal Register*, volume 71. U.S. National Archives and Records Administration, March 2006

Discussion 3

Data for this sort of analysis have historically been difficult to find as many organizations will not report when failures take place. Recent changes in the legal environment of the United States, beginning with California's breach notification law (most commonly referred to by its bill number, S.B. 1386), has led to a significant increase in the number of publicly-reported incidents.

These breach notices are helpful in raising consumer awareness with regard to the threat of fraud stemming from identity theft. Additional utility should come from the analysis of such information to establish areas where efforts to protect personal information are failing and areas where controls should be deployed. Such analysis might well show that organizations are structurally ill-equipped to protect the information in their care. Information security departments are often viewed as a group within information technology, even though the department reports to management in finance, risk management, or audit. Realignment of the organization's resources might be required to put effective controls in place.

3.1 Alternative Classification of Failure

One sort of analysis might view data through the prism of an information security controls framework. We believe that, for the purpose of understanding how control loss incidents have taken place, this would present an incomplete view. We will consider an alternative here.

Where the controls framework used for assessment is hierarchical, as is true with ISO/IEC 27002, failures may be categorized according to some defined level of precision. For example, consider the hierarchy of controls around media handling from the ISO/IEC 27002.

In the event of an information security failure caused by the loss of a backup tape, the problem can quickly be identified as a failure at control 10.7. Suppose that more specific information is known: that the problem came from a failure to control access to the tape when moving it from one location to another. The failure then could be correctly identified at 10.7.1—a more precise specification of the type of media-handling failure. Suppose now that less information is known, perhaps where it is unclear whether the loss came from the handling of media (10.7), management of network security (10.6), or an error in the information exchange policy with a business partner (10.8.1). This case would thus best be classified as a failure at the level of communications and operations management (10). With such a system, we can always ensure accuracy given the precision of information available.

Number	*Title*
10	Communications and operations management
	. . .
10.6	Network security management
	. . .
10.7	Media handling
10.7.1	Management of removable media
10.7.2	Disposal of media
10.7.3	Information handling procedures
10.7.4	Security of system documentation
10.8	Exchange of information
10.8.1	Information exchange policies and procedures

Table 3.1: ISO/IEC 27002 Controls Around Media Handling

Herein lies the problem: the mechanism of actual incidents is often a combination of issues at various levels. For example, when the State of Ohio lost a backup tape, the State's Inspector General noted several types of failures. These included inappropriate mechanisms for the handling of sensitive information (ISO/IEC 27002, section 9.2.5), allowance of untrained employees to handle sensitive information (ISO/IEC 27002, sections 8.1.1 and 8.2.2), response to the theft (ISO/IEC 27002, section 13.2.1), and the storage of sensitive information in an unsecured folder (ISO/IEC 27002, section 11.6.1).[1] Classification of this one incident, therefore, would mean generalizing to a level of uselessness to maintain accuracy. The only workable alternative would be to track multiple failures for a particular incident.

With the taxonomy that we have proposed, the entire incident falls into one category, Physical Failure:Media:Data Tape (A2c). Such a classification scheme is more suitable for the sort of analysis of data control loss incidents across organizations and across industries.

3.2 *Alternative Classifications of Context*

Classification by industry is not the only means of determining context for an incident. Two other options that we considered include classification by region and classification by information type. Ultimately we determined that, for our purposes, classification by industry was the most appropriate. However, other studies might benefit from one of these alternatives.

[1] State of Ohio Office of Inspector General. Report of Investigation, July 2007. File ID No 2007190

3.2.1 Classification by Region

Classification of control failure by region could be accomplished in a variety of ways, depending on the needs of the study. For example, a multinational organization might want to understand where to prioritize its remediation activities; analyzing control failure in a particular country could help the organization to prioritize its remediation activities in a manner that best addresses the unique cultural, legal, and technical issues in that country. Similar analysis could be made to compare across trade blocs like the European Union or North America.

With an understanding of where actual control failures most often happen, a multinational organization might decide to reörganize its information processing functions away from those regions, since addressing the specific control failures head-on may ultimately be a less cost-effective option to provide proper protection of information.

This sort of analysis seems available only to multinational organizations that can work with their own control failure data. We are unaware of any significant compilation of control failure data across national boundaries.

3.2.2 Classification by Information Type

Studying control failures by the type of affected information could shed light on ways in which different types of information are exposed when things go wrong. An organization looking to give special attention to particular types of information could then consider whether the organization depends on controls that are known to fail at an unacceptably high rate and whether sufficient secondary controls are in place to protect against such failure.

While some correlation between information type and industry classification might exist, following information type would allow for a different sort of analysis. For example, if a hospital reports the exposure of a medical record, the industry reported would be quite different from a medical record exposed by an information service provider working for the hospital. In both cases, however, it is a medical record that has been exposed.

3.3 Future Work

The immediate next step is to produce an analysis of a set of incident data using the taxonomy in order to determine whether the taxonomy is sufficient for classification of incidents and ultimately to assess the incidents against one another.

We hope that organizations such as the Identity Theft Resource Center, US-CERT, and the International Association of Privacy Professionals (IAPP) that have an interest in the promoting effective methods to keep sensitive information confidential will use the taxonomy as a means of discussing incidents that fall within their various areas of focus.

Privacy and security officers in particular would do well to encourage their organizations to release information regarding data control loss incidents in such a manner that will allow for proper classification of incidents. We believe that this taxonomy will allow for disclosure of such information in a way that is beneficial for classification while protecting organizations from further risk through the publication of specific vulnerabilities to potentially hostile actors. More ready and specific classification of incidents in a standardized system will aid in the understanding of actual events and ultimately promote better risk management.

Part II

A Comparative Analysis of Three Years of Breach Reports by Breach Type and Industry

Introduction

4

Evidence suggests that budgets for information security have been on the increase.[1] Even so, there are not enough resources to address all plausible threats to the confidentiality, integrity, and availability of information. With limited resources, trade-offs must be made: increased expenditure for controls in one area necessarily means the loss of options in others. In order for organizations to make economically sound security trade-offs, they must first understand the prevalence and severity of the threats and vulnerabilities that face their particular organizations.

Though the need for accurate information is clear, the guidance available to the relevant decision maker is routinely of questionable value. Anecdote and personal opinion are in wide circulation. Many studies lack context, focusing on a single type of failure or a single class of control, without relating them to the bigger picture. Even more suspect is the propaganda disseminated by vendors of control services, applications, and appliances whose principal motivation is sales. Their perspective is designed to focus an audience's attention on the very threat or vulnerability that the vendor's product was designed to address. Such propaganda provides little or no guidance to organizations that are attempting to prioritize efforts to protect the information they possess.

While the severity of threats and vulnerabilities, and the corresponding risks that they may pose, are subject to debate, when these risks become concrete in the form of a breach, we are provided with a much clearer data point. This is not to say that breach reports provide unequivocal information or a total picture of a compromise. Many breach reports fail to provide a complete understanding of the threat agent. The nature of the vulnerability, however, is usually made far clearer. In fact, often a breach report results from merely detecting a vulnerability, even when there is no known case of a threat having exploited it. Some announcements may withhold information on the threat and vulnerability, opting instead simply to detail the compromised data.

The absence of a breach notification is not the same as the ab-

[1] Robert Richardson. 2007 csi/fbi computer crime and security survey, 2007

sence of a breach. An undetected breach cannot be reported. A compromise that is detected internally may not be communicated to the larger public, either because the likelihood of a threat having exploited a vulnerability is deemed too unlikely, or because the organization determines that it would rather accept the consequences of a lack of disclosure than the additional expenditure that might result from publicizing a compromise.

Though breach notices provide imperfect information, it is possible to devise a method that effectively handles information of varying precision, allowing us to glean valuable information contained in imperfect notices. In Part I: A Taxonomy of Data Losses, we attempted to provide a foundation for the discovery of quantitative knowledge about the nature, distribution, and frequency of security breaches—knowledge that could help inform those responsible for applying information security dollars about the relative control gaps they may need to address.

In this part, we apply the taxonomy to a real world data set in order to gain insight into the frequency of various data breaches. Then we apply that insight to determine whether there are significant differences in the occurrence of different types of breaches between industries.

We have undertaken an in-depth analysis of a limited set of data breaches—those that pose a threat to the confidentiality of personally identifying information, or data that might lead to the threat of identity theft. While this type of failure is not the only sort that interests information assurance practitioners, the growing body of state and federal legislation and the motivation to make these breaches public, means a greater number of samples are available for review.

Method

5

To perform this analysis, we used data describing individual security breaches. These breaches were first separated by type; then, each organization or entity with first-order responsibility for the data that was compromised was categorized by industry. Data were subjected to standard analytical methods, including tests for statistical significance, to uncover whether discernible patterns exist within and between industries.

5.1 Data

Data for this analysis were drawn from a collection of breach reports collected by the Identity Theft Resource Center (ITRC). ITRC describes itself as "a nonprofit, nationally respected organization dedicated exclusively to the understanding and prevention of identity theft. The ITRC provides consumer and victim support as well as public education. It also advises governmental agencies, legislators, law enforcement, and businesses about the evolving and growing problem of identity theft."[1]

ITRC has published breach reports for 2005,[2] 2006,[3] 2007,[4] and 2008.[5] We have limited the scope of this project to the years 2005 through 2007. This data set is by no means intended to give complete coverage of all information security breaches during the sample period. In keeping with its mission, ITRC focuses its attention on those breaches occurring in the United States that pose a risk for identity theft, which they define as "a crime in which an impostor obtains key pieces of personal identifying information (PII) such as Social Security numbers and driver's license numbers and uses them for their own personal gain."

Information included in the breach reports was chosen solely at the discretion of ITRC, and, as stated in the description of the 2007 report, the only breaches published in the report are from "real and credible" sources.

Breach report data were entered into a spreadsheet for analysis.

[1] Identity theft resource center. Web Site. [online] http://www.idtheftcenter.org/

[2] Identity Theft Resource Center. 2005 disclosures of U.S. data incidents, 2006. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202005.pdf

[3] Identity Theft Resource Center. 2006 disclosures of U.S. data incidents, January 2007. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202006.pdf

[4] Identity Theft Resource Center. 2007 breach list, January 2008a. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202007.pdf

[5] Identity Theft Resource Center. 2008 breach list, February 2008b. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202008.pdf

In some instances a single breach report cited more than one breach incident. Where possible, we separated the report into multiple observations.

5.2 Taxonomy

We labeled each entry in the data set using the nomenclature set out in Part I: A Taxonomy of Data Losses. The hierarchical structure allows us to classify information breaches to the level of precision available in the published reports.

We used the textual descriptions of the breaches contained in the ITRC sources to categorize the data. When it was not possible to gain enough information from the description to categorize the breach to at least the second level of the hierarchy, we consulted the original sources. When this did not provide enough information, we performed a Web search in an attempt to uncover more information. If these efforts failed, we labeled the breach to the level of precision possible. Further analyses of the breach descriptions may be undertaken as the taxonomy matures. For example, if all that was known about a breach is that information was compromised during a break-in, we labeled that breach to the first level: an instance of "A: Physical Failure." However, if we knew the burglar took several laptop computers containing unencrypted, sensitive data, we labeled that at the third level: "A3b: Lost or Stolen Laptop."

While the entries were scored to the degree of precision allowed by the breach notification, for the purposes of this analysis we have focused on the second level in the taxonomy hierarchy. Those entries that could be distinguished only to the first level were excluded from analysis, and those that were precise to the third level or deeper were grouped together by their second level.

Part I presents a complete discussion of the taxonomy. Briefly, the levels considered here are shown in Table 5.1.

A1	Lost or Stolen Documentation, labeled in figures and tables as *Docs*.
A2	Lost or Stolen Digital Media, labeled in figures and tables as *Media*.
A3	Lost or Stolen Computing Hardware, labeled in figures and tables as *Hardware*.
B4	Insider Misconduct, labeled in figures and tables as *Insider*.
B5	Compromised Host, labeled in figures and tables as *Compromise*.
C6	Insecure Surplussing, labeled in figures and tables as *Processing*.
C7	Discarded Data, labeled in figures and tables as *Disposal*.

Table 5.1: Taxonomy at the Second Level

5.3 Industry

Rather than risk introducing bias by introducing our own scheme, we chose the 2002 North American Industry Classification System (NAICS)[6] as our means for analyzing breach distribution by industry. Though a newer version of NAICS was released in 2007, the greater availability of NAICS 2002 data in free sources led us to use the older classification. A best effort was made to identify the NAICS code of the organization responsible for the

[6] Office of Management and Budget. North american industry classification system: Revision for 2007; notice. In *Federal Register*, volume 71. U.S. National Archives and Records Administration, March 2006

data exposed in each breach in our data set. Each observation was then labeled with the first two digits of the code, representing the broadest distinction between industrial sectors.

6 Results

Over the three years of ITRC data, we cataloged 925 observations. Some of these were presented in the breach lists as a single notice that described multiple distinct incidents. These were separated into different observations. Only 905 observations could be classified to the second level. Five entries were unclassifiable; they noted only that a breach had occurred that required some form of credit monitoring or other corrective action. Fifteen entries could be classified only to the first level, one fell into category A (Physical), thirteen in category B (Logical), and one in category C (Procedural). Of the 905 classifiable observations, six did not have enough information about the responsible organization to be assigned a NAICS code. The 899 remaining observations are detailed in Figure 6.1. While most NAICS codes were represented in the data, there were no observations for codes 11 (*Agriculture, Forestry, Fishing, and Hunting*), 42 (*Wholesale Trade*), and 55 (*Management of Companies and Enterprises*), so these are excluded from all tables, figures, and analysis.

11	Agriculture, Forestry, Fishing and Hunting
21	Mining
22	Utilities
23	Construction
31–33	Manufacturing
42	Wholesale Trade
44–45	Retail Trade
48–49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services
92	Public Administration

Table 6.1: NAICS 2007 Codes and Descriptions

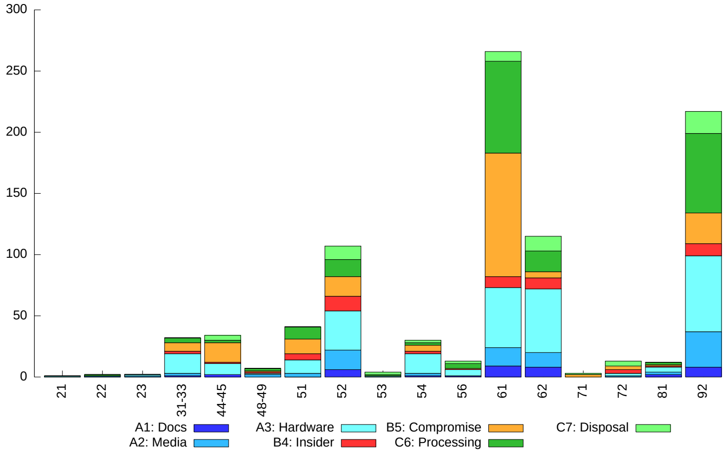

Figure 6.1: Breach Type Observations by Industry

6.1 *Distribution of Observations*

To test the hypothesis that a significant difference exists in the distribution of breach types between industries, we assumed the opposite, generating a table of expected observations for each breach type by industry (Table 6.3). We calculated each cell in this "expected" table by taking the product of the total number of observations in that industry and the percentage of that specific breach type across all industries. Taking each industry in turn we then performed a χ^2 test[1] for statistical significance, comparing the observed distribution of breach types against the expected values. P-values resulting from those tests are detailed in Table 6.2.

Agriculture	.8760759196
Mining	.4832063360
Utilities	.5512226506
Manufacturing	.1817398353
Retail	.0034487983
Transportation	.5437545766
Information	.2034512769
Finance	.0157029420
Real Estate	.0612124637
Prof. Services	.1271919999
Admin. Services	.3659340107
Educational	.0000000001
Health Care	.0000096276
Arts	.2484658204
Accommodation	.0022826685
Other Services	.3049347667
Public Admin.	.0018734368

Table 6.2: χ^2 P-Values by Industry

A decision crucial to our analysis was the selection of a p-value for our tests of statistical significance. The p-value is the probability, assuming the truth of a hypothesis (typically a null hypothesis[2]), of generating a distribution at least as extreme as the observed distribution. This is a standard method for determining the "statistical significance" of the observed set of data. We are not aware of a body of work in Information Assurance that sets a standard for statistical significance. Since the threshold of $p = 0.05$ is common across many disciplines, we use that here.

	Docs		Media		Hardware		Insider		Compromise		Processing		Disposal		Total
Agriculture	0.0	0	0.1	0	0.3	1	0.1	0	0.2	0	0.2	0	0.1	0	1
Mining	0.1	0	0.2	1	0.6	0	0.1	0	0.4	0	0.4	1	0.1	0	2
Utilities	0.1	0	0.2	1	0.6	1	0.1	0	0.4	0	0.4	0	0.1	0	2
Mfg.	1.4	1	3.1	2	9.3	16	2.0	2	6.9	7	7.1	4	2.3	0	32
Retail	1.4	2	3.3	0	9.9	9	2.1	1	7.3	16	7.5	2	2.4	4	34
Transport	0.3	0	0.7	2	2.0	1	0.4	1	1.5	1	1.5	2	0.5	0	7
Information	1.7	0	3.9	3	11.9	11	2.6	5	8.8	12	9.1	10	2.9	0	41
Finance	4.5	6	10.2	16	31.2	32	6.7	12	23.1	16	23.7	14	7.6	11	107
Real Estate	0.2	0	0.4	0	1.2	1	0.2	0	0.9	0	0.9	1	0.3	2	4
Prof. Svc.	1.3	1	2.9	2	8.7	16	1.9	2	6.5	5	6.6	2	2.1	2	30
Admin. Svc.	0.5	1	1.2	0	3.8	5	0.8	1	2.8	0	2.9	4	0.9	2	13
Education	11.2	9	25.4	15	77.5	49	16.6	9	57.4	101	58.9	75	18.9	8	266
Health Care	4.9	8	11.0	12	33.5	52	7.2	9	24.8	5	25.5	17	8.2	12	115
Arts	0.1	0	0.3	0	0.9	0	0.2	0	0.6	2	0.7	0	0.2	1	3
Accomm.	0.5	0	1.2	1	3.8	2	0.8	3	2.8	3	2.9	0	0.9	4	13
Other Svc.	0.5	2	1.1	2	3.5	4	0.7	1	2.6	1	2.7	2	0.9	0	12
Pub. Admin.	9.2	8	20.8	29	63.2	62	13.5	10	46.8	25	48.0	65	15.4	18	217
Total	37.9	38	86.0	86	261.9	262	56.0	56	193.8	194	199.0	199	63.8	64	899

Table 6.3: Breaches Expected and Observed by Type and Industry

[1] A χ^2 test is a standard test to approximate the degree to which observations in a contingency table are independent of one another. Put another, less precise way, the test gives us an indication of how likely it is that each observation in a table was placed into a category at random.

[2] A "null hypothesis" is a hypothesis intended to be refuted. Our null hypothesis is: "There is no difference in the distribution of breach types across industries." By finding little evidence of the null hypothesis we become correspondingly more confident in the "alternate" hypothesis.

Strictly speaking, six of the seventeen observed industries satisfied our p-value requirement for significance. To address risk of Type II errors[3] in the χ^2 analysis, we exclude industries that have fewer than five breaches of a particular type. While the p-values for codes 44–45 (*Retail Trade*) and 72 (*Accommodation and Food Services*) pass our minimum standard, they both show a majority of cells with values below 5. This leaves us with codes 52 (*Finance and Insurance*), 61 (*Educational Services*), 62 (*Health Care and Social Assistance*), and 92 (*Public Administration*) as subjects for our analysis.

Figure 6.2 shows the percentage of each breach type within our four subject industries, as well as for all observations collectively.

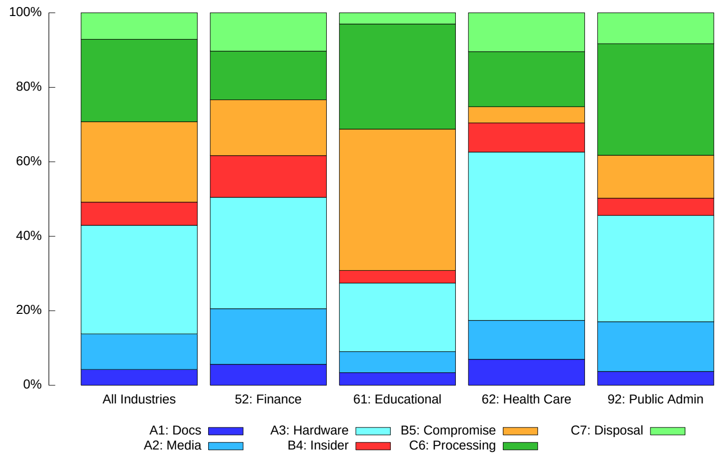

Figure 6.2: Breach Type Percentages by Industry

Table 6.4 shows the distribution of all observations after consolidating all low p-value industries into a catch-all category. We ran a χ^2 test on each column, comparing each breach type in turn to Table 6.5, which contains our expected values. Table 6.6 contains the p-value results of those tests. The results show a high degree of significance for lost and stolen media and hardware, insider misconduct, compromised hosts, processing errors, and insecure disposal. Only the loss or theft of documentation does not appear to differ significantly between industries.

	Finance	**Educational**	**Health Care**	**Public Admin.**	**All others**	**Total**
Docs	6	9	8	8	7	38
Media	16	15	12	29	14	86
Hardware	32	49	52	62	67	262
Insider	12	9	9	10	16	56
Compromise	16	101	5	25	47	194
Processing	14	75	17	65	28	199
Disposal	11	8	12	18	15	64
Total	107	266	115	217	194	899

Table 6.4: Observed Values with Low P-Value Industries Consolidated

	Finance	**Educational**	**Health Care**	**Public Admin.**	**All others**
Docs	4.5	11.2	4.9	9.2	8.2
Media	10.2	25.4	11.0	20.8	18.6
Hardware	31.2	77.5	33.5	63.2	56.5
Insider	6.7	16.6	7.2	13.5	12.1
Compromise	23.1	57.4	24.8	46.8	41.9
Processing	23.7	58.9	25.5	48.0	42.9
Disposal	7.6	18.9	8.2	15.4	13.8

Table 6.5: Expected Observations with Low P-Value Industries Consolidated

Docs	.5116643991
Media	.0172283798
Hardware	.0001473399
Insider	.0344524032
Compromise	.0000000001
Processing	.0001688209
Disposal	.0384812865

Table 6.6: χ^2 P-Values by Breach Type

[3] A "Type II error," also known as a "false negative," is a rejection of a correct hypothesis.

6.2 Trends

We separated observations by industry and by year in an effort to compare the trends in breaches from year to year. Figure 6.7 shows the results for all observations. Statistical analysis of the trend values is a work in progress. It is likely that there are not enough observations per year in many instances to make concrete statements about the significance of our results.

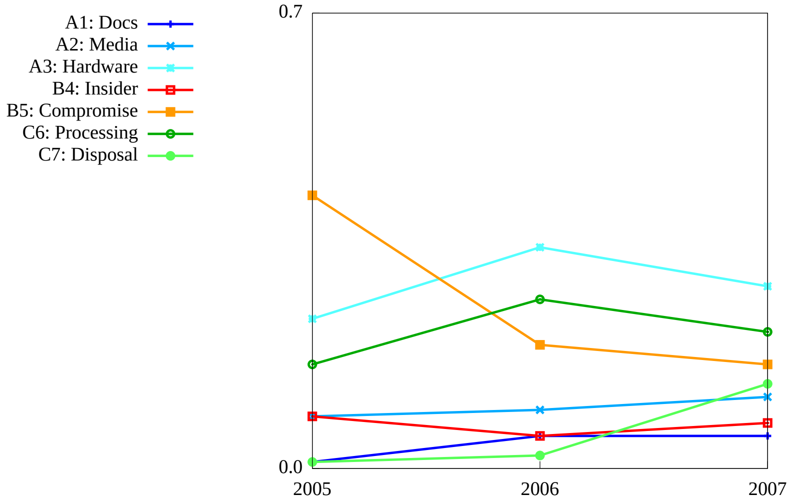

Figure 6.7: Proportion of Breach Types By Year, All Observations

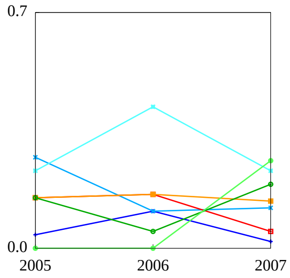

Figure 6.3: Proportion of Breach Types By Year, *Finance*

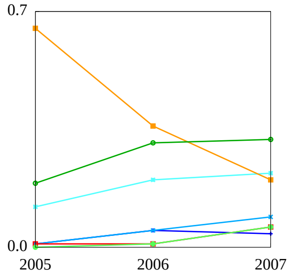

Figure 6.4: Proportion of Breach Types By Year, *Education*

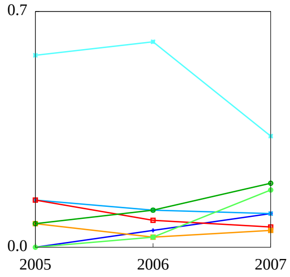

Figure 6.5: Proportion of Breach Types By Year, *Health Care*

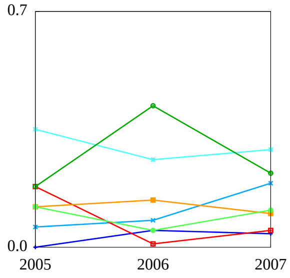

Figure 6.6: Proportion of Breach Types By Year, *Public Administration*

7 Discussion

While the bulk of media attention on threats to private information is given to the activity of outside attackers, these breaches account for only approximately 22% of the instances in our data set. More significant is the number and type of breaches caused by people within an organization. Poor procedures, human errors by staff (*Processing* and *Disposal*), and the malicious activities of people on the inside of an organization account for greater than 35% of our observations.

Noteworthy is the finding that the single largest contributor to our data set comes from the loss and theft of computing hardware. In many cases there was no way to distinguish lost from stolen, though in practice the distinction is unimportant. Once a device has left an organization's control, the organization can no longer rely on the security of the information that the device contains. It is reasonable to assume that the majority of these thefts were attempts to acquire the hardware—thus the data were not the primary target. As the perceived profitability of information theft increases and the retail value of laptops declines, it is likely that the proportion of theft specifically targeting data will increase.

In breach notifications where a computer was stolen, the reporter often hastens to note that the device was password protected. While this may serve to ease the fears of the general public, a password does little to protect the information if the information itself was the thief's target. Simply attaching the hard drive to another device or booting the stolen device from alternate media will give the attacker access to the data. However, proper use of good encryption is an effective control for data on stolen hardware or electronic media. While cryptography would not stop the thefts in a case where the hardware is the target, it could change the incident into one that is resolved by a simple police report on the value of the hardware or media rather than a public breach announcement.

An early impetus for this research was a heated debate with a colleague who argued that the impact of cryptographic controls is wildly overstated. If one accepts the position that the loss of properly encrypted data does not constitute a breach, then 38% of all losses that led to our data, those observations stemming from both lost and stolen media and hardware, would never have led to disclosures if the data they contained had been encrypted.

7.1 *Health Care and Social Assistance*

Upon looking at the distribution of breaches across industries (Figure 6.2), we find the most striking characteristic is the proportion of lost and stolen hardware events in the *Health Care and Social Assistance* sector. At over 45% of reported breaches, physical control of hardware clearly needs attention. Other forms of loss and theft are slightly higher in *Health Care* as well, though we did not get favorable significance numbers overall for lost and stolen documents, so we warn against any conclusions drawn from that category from this analysis.

We had some concern that lost hardware, being such a large proportion of the data loss in the *Health Care* industry, might skew differences in other categories. We removed lost hardware from the data and regenerated proportions (Table 7.1) to see what effect this had. We did not perform any statistical tests on the results, so their significance is suspect; however, the results give additional support to a greater proportion of loss and theft of media in *Health Care*.

	All Industries	**Health Care**
Docs	5.97	12.70
Media	13.50	19.05
Insider	8.79	14.29
Compromise	30.46	7.94
Processing	31.24	26.98
Disposal	10.05	19.05

Table 7.1: Breach Type Percentages Compared Between All Observations and Health Care and Social Assistance with Lost/Stolen Hardware Removed

The relatively small rate of processing errors in the *Health Care* sector relative to all industries (15% compared to 22%) may imply an above average level of attention to the quality of mandated procedures regarding the handling of information, either through the inherent culture of the industry, or the impact of HIPAA[1] on the procedures in use. If such procedures exist, they do not adequately address the improper disposal of documentation, as we observe that *Health Care* has the highest proportion of improperly disposed documentation, at nearly 10.5%.

While compromised hosts account for about 22% of all of our data points, they account for less than 4.5% of the reported breaches in *Health Care*. We are led to believe that the industry either has done an above average job in securing logical access to its systems, or the industry faces a disproportionate shortfall in detective controls.

[1] HIPAA. Health Insurance Portability and Accountability Act of 1996. PUBLIC LAW 104-191, 1996. [online] http://aspe.hhs.gov/admnsimp/pl104191.htm

7.2 Educational Services

The proportion of compromised hosts in the *Educational Services* sector is the polar opposite of that in the *Health Care* sector. Nearly 38% of all breaches in *Education* are of this type. Given the large number of reports in this sector overall, compromised hosts in *Education* account for more than 11% of all observations in the full data set. This may indicate that *Education* faces a greater proportion of threats from outside parties seeking to target their computing infrastructure, a greater proportion of vulnerabilities due to outdated or unpatched services and missing controls, or some combination of these issues. Alternatively, it might signal a better than average ability to detect compromised hosts. Either way, it is likely that the proportion is significant and worthy of future study.

Though in absolute magnitude *Education's* processing errors are the second largest contributor to our data set (8.3% of all observations), the percentage of breaches of this type does not appear to deviate significantly from the overall percentage. However, if we factor compromised hosts out of the category (Table 7.2), the numbers lead us to reëxamine that conclusion. We have yet to determine the appropriate test for significance, so these results are as yet undetermined.

Lost and stolen hardware and media appear to be lower in *Education*, though in Table 7.2 the gap closes for media. Insider misconduct is correspondingly lower than the general case. This could be attributed to the type of personally identifying information that an educational institution is likely to maintain. If a criminal's goal is to steal identities for financial gain, there are probably more potentially lucrative targets than students. Improper disposal of sensitive information is also lower in *Education*. Note, however, that the differences for insider misconduct and improper disposal may also be within an as-yet uncalculated margin of error.

	All Industries	**Educational**
Docs	5.39	5.45
Media	12.20	9.09
Hardware	37.16	29.70
Insider	7.94	5.45
Processing	28.23	45.45
Disposal	9.08	4.85

Table 7.2: Breach Type Percentages Compared Between All Observations and Educational Services with Compromised Hosts Removed

7.3 Public Administration

Public Administration has a disproportionately lower rate of compromised hosts relative to the overall data set: 11.5% versus 21.5%. As with *Health Care*, there are two competing hypotheses. The numbers imply that either fewer compromises occurred due to a higher degree of preventative control, or fewer compromises were reported, possibly because of lower coverage by detective controls. A specific review of practices and behaviors in *Public Administration* relative to other industries would likely clear up the ambiguity.

The proportion of processing errors relative to all industries (30% as opposed to 22%) in *Public Administration* is also worth investigation. Obviously, the sorts of organizations that make up this sector—the military, law enforcement, public services—will have greatly different needs and procedures. Further study into the specifics of these incidents should point to the root cause for these differences.

7.4 Finance and Insurance

Of our four "statistically significant" data sets, *Finance and Insurance* has the least significant p-value (see Table 6.2), but there remain a few points of interest worth noting. The *Finance and Insurance* industry has the smallest proportion of processing errors, at 13%. This is not surprising given the high degree of formalization and regulation of the procedures in the industry as a whole. Also not surprising, given the type of data the industry uses and its value in fraud, is that *Finance* manifests proportionally the highest rate of insider misconduct. At over 11%, insider misconduct occurs at more than double the rate in the overall data set.

7.5 Trends

We have not performed any statistical tests on the trend lines, and believe that it is likely that the sample size per year may be too small to draw any meaningful conclusions. Furthermore, having only three years of data makes it difficult to put much faith in any trend we spot. We will nonetheless present some preliminary impressions, with the warning that the information contained in this section may be subject to personal bias. *Caveat lector.*

The level of lost and stolen documentation appears to have been slightly on the rise across years and industries, with the exception of *Finance*, the only industry in which the rate of lost and stolen documentation was lower in 2007 than in 2005. Loss or theft of media has been increasing slightly in the full data set, though different industries have been moving in different directions. Most noteworthy is the precipitous drop in the *Finance* sector, from more than a quarter of all reports in 2005, to about 12% in 2007. No industry presents a clear trend regarding the loss and theft of hardware, though the steep drop in reports in the *Health Care* industry may signal a turnaround in what has been its largest breach type.

There is little clear about trends of insider misconduct, but the *Education* sector may be worth watching if its proportion continues to climb in 2008. The proportion of compromised hosts dropped significantly from 2005 to 2007, but with the exception of *Education*, our statistically significant data sets changed little.

No clear trend is evident in the proportion of processing errors. There is a pronounced jump in the number of instances of improper disposal, although we believe this jump is simply due to an increase in the reporting of discarded documents. Since sensitive data has been improperly disposed of for as long as there has been paper and dumpsters, we do not think this is a real trend in incidents, but a trend in reporting. A large number of the reports in the ITRC collections detail phone calls to media outlets alerting them to the presence of documents in public. Numerous laws at the state[2] and federal[3] levels require the proper disposal of sensitive documentation, and it is likely that this topic simply became a popular news item in 2007, rather than there being any significant increase in processing errors.

[2] Vermont Act No. 162. Vermont act no. 162, 2006. [online] http://www.leg.state.vt.us/docs/legdoc.cfm?URL=/docs/2006/acts/ACT162.HTM

[3] GLBA. Gramm-Leach-Bliley Act. PUBLIC LAW 106-102, 1999. [online] http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=106_cong_public_laws&docid=f:publ102.106; FCRA. Fair Credit Reporting Act. 15 U.S.C. §1681 et seq, 2001. [online] http://www.ftc.gov/os/statutes/031224fcra.pdf; SOX. Sarbanes-Oxley Act. PUBLIC LAW 107-204, 2002. [online] http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:publ204.107; and FACTA. Fair and Accurate Credit Transactions Act. PUBLIC LAW 108-159, 2003. [online] http://www.treasury.gov/offices/domestic-finance/financial-institution/cip/pdf/fact-act.pdf

8 Future Work

Future work includes the calculation of statistical significance of the trend lines within industries. Given our impression that the sample count may be too low to give an accurate accounting, our discussion of trends will be limited.

Undoubtedly, data breaches will be with us for some time, providing more data for analysis in the future. Inclusion of those data will help boost our confidence in our statistical analysis.

9 Credits

C. Matthew Curtin is the founder of Interhack Corporation, a computer expert firm with practice areas in Information Assurance and Forensic Computing. As a forensic computer expert, Mr. Curtin analyzes information technology and electronically stored information to answer questions that arise in adjudication. He has appeared as an expert witness in both civil and criminal cases, dealing with everything from electronic discovery to assessment of information technology in practice. Since 1998, Mr. Curtin has maintained a regular academic appointment as a lecturer at The Ohio State University's Department of Computer Science and Engineering, teaching courses in the Common Lisp programming language and operating systems implementation. He frequently lectures on the topic of forensic computing to audiences of judges and attorneys.

Lee Ayres joined Interhack as a Senior Analyst in January 2007, supporting both the Forensic Computing and Information Assurance practices. Using his experience as an application developer and systems analyst, he helps attorneys in litigation understand how to make use of the systems and data available to them as evidence. For three years prior to joining Interhack, Lee was the lead developer at Mercury Markets in Chicago, building automated trading systems for the global financial markets using proprietary algorithms. His previous experience includes work as a developer and system engineer at I-DEP, building systems for taking depositions over the Internet and onShore Development, building Web applications in Common Lisp. Lee holds a Bachelors Degree in Computer Science and Engineering from The Ohio State University.

We are grateful for commentary we received on early revisions of Part II specifically, and on the work generally. In particular, we wish to thank Steve Romig for helpful advice on interpreting observed differences among industries. We also want to thank the Identity Theft Resource Center for collecting reported data loss incidents and publishing its observations.

Bibliography

Identity theft resource center. Web Site. [online] http://www.idtheftcenter.org/.

Christopher J. Alberts, Audrey J. Dorofee, and Julia H. Allen. Octave(sm) catalog of practice, version 2.0, October 2001. [online] http://www.cert.org/archive/pdf/01tr020.pdf.

Identity Theft Resource Center. 2005 disclosures of U.S. data incidents, 2006. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202005.pdf.

Identity Theft Resource Center. 2006 disclosures of U.S. data incidents, January 2007. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202006.pdf.

Identity Theft Resource Center. 2007 breach list, January 2008a. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202007.pdf.

Identity Theft Resource Center. 2008 breach list, February 2008b. [online] http://idtheftmostwanted.org/ITRC%20Breach%20Report%202008.pdf.

CERT/CC. Full statistics, January 2008. [online] http://www.cert.org/stats/fullstats.html.

Matt Curtin. *Developing Trust: Online Privacy and Security*. Apress, November 2001.

Department of Health and Human Services. Health insurance reform: Security standards; final rule. In *Federal Register*, volume 68. U.S. National Archives and Records Administration, February 2003. [online] http://www.cms.hhs.gov/SecurityStandard/Downloads/securityfinalrule.pdf.

Adam Dodge. Educational Security Incidents: Year In Review—2006, 2007. [online] http://www.adamdodge.com/esi/yir_2006.

Adam Dodge. Educational Security Incidents: Year In Review—2007, February 2008. [online] http://www.adamdodge.com/esi/year_review_2007.

FACTA. Fair and Accurate Credit Transactions Act. PUBLIC LAW 108-159, 2003. [online] http://www.treasury.gov/offices/domestic-finance/financial-institution/cip/pdf/fact-act.pdf.

FCRA. Fair Credit Reporting Act. 15 U.S.C. §1681 et seq, 2001. [online] http://www.ftc.gov/os/statutes/031224fcra.pdf.

Federal Trade Commission. Standards for safeguarding customer information; final rule 16 cfr part 314. In *Federal Register*, volume 67. U.S. National Archives and Records Administration, May 2002.

GLBA. Gramm-Leach-Bliley Act. PUBLIC LAW 106-102, 1999. [online] http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=106_cong_public_laws&docid=f:publ102.106.

HIPAA. Health Insurance Portability and Accountability Act of 1996. PUBLIC LAW 104-191, 1996. [online] http://aspe.hhs.gov/admnsimp/pl104191.htm.

ISO. Information technology—security techniques—information security management systems—requirements. International Standard ISO/IEC 27001, 2005a.

ISO. Information technology—security techniques—code of practice for information security management. International Standard ISO/IEC 27002, 2005b.

Thomas A. Longstaff, James T. Ellis, Shawn V. Hernan, Howard F. Lipson, Robert D. Mcmillan, Linda Hutz Pesante, and Derek Simmel. Security of the internet, 1997. [online] http://www.cert.org/encyc_article/tocencyc.html.

State of Ohio Office of Inspector General. Report of Investigation, July 2007. File ID No 2007190.

Office of Management and Budget. North american industry classification system: Revision for 2007; notice. In *Federal Register*, volume 71. U.S. National Archives and Records Administration, March 2006.

Esther Pearce. History of the standard industrial classification. Washington, D.C., Executive Office of the President Office of Statistical Standards, U.S. Bureau of the Budget, July 1957. [online] http://www.census.gov/epcd/www/sichist.htm.

Robert Richardson. 2007 csi/fbi computer crime and security survey, 2007.

SOX. Sarbanes-Oxley Act. PUBLIC LAW 107-204, 2002. [online] http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:publ204.107.

Gary Stoneburner, Alice Goguen, and Alexis Feringa. Risk management guide for information technology systems. NIST SP 800-30, July 2002. [online] http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf.

Vermont Act No. 162. Vermont act no. 162, 2006. [online] http://www.leg.state.vt.us/docs/legdoc.cfm?URL=/docs/2006/acts/ACT162.HTM.