A Statistical Analysis of Disclosed Storage Security Breaches

Ragib Hasan
rhasan@ncsa.uiuc.edu

William Yurcik
byurcik@ncsa.uiuc.edu

National Center for Supercomputing Applications (NCSA)
University of Illinois at Urbana-Champaign (UIUC) Urbana, IL 61801.

ABSTRACT

Many storage security breaches have recently been reported in the mass media as the direct result of new breach disclosure state laws across the United States (unfortunately, not internationally). In this paper, we provide an empirical analysis of disclosed storage security breaches for the period of 2005-2006. By processing raw data from the best available sources, we seek to understand the what, who, how, where, and when questions about storage security breaches so that others can build upon this evidence when developing best practices for preventing and mitigating storage breaches. While some policy formulation has already started in reaction to media reports (many without empirical analysis), this work provides initial empirical analysis upon which future empirical analysis and future policy decisions can be based.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and Protection*; D.4.2 [**Software**]: Operating Systems—*Storage Management*; H.3.4 [**Information Systems**]: Information Storage and Retrieval—*Systems and Software*

General Terms

Security, Economics, Legal Aspects

Keywords

storage security, security breaches, breach disclosure laws

1. INTRODUCTION

There have been a wide range of organizations with disclosed storage security breaches that have subsequently been reported in the mass media between 2005-2006 [2, 3, 11]. PrivacyRightsClearingHouse reports a total of 90 million records containing sensitive personal information have been compromised [1]. Risks from releasing private information in a storage security breach are twofold: (1) privacy risk and (2) identity theft fraud [10] – with damages resulting from these two risks estimated to be on the order of billions of dollars in the United States alone.

The only reason we know about most storage security breaches are new state laws mandating disclosure to affected parties of incidents that release private data due to security compromise. Before the first breach disclosure state law in 2003, organizations did not notify affected parties when their private data was compromised, leaving them at risk for identity theft fraud often only to find out when it was too late. New state disclosure laws allow individuals to take proactive steps to safeguard their identities after a compromise has occurred – thus returning control of private data back to individuals.

Breach disclosure laws have done much more than giving individuals notice breach disclosures have also improved protection by providing metrics upon which to measure security where no metrics existed before. However, since there are typically no public disclosure requirements in state laws and disclosure laws have not been actively and uniformly enforced, reporting in the mass media has been spotty and focused on the sensational rather than insightful analysis.

The goal of this paper is to provide in-depth analysis of storage security breaches (beyond media reports) by processing raw data from a combination of best available sources for emerging patterns. In previous work, we framed a storage security threat model which organized potential attacks into categories along multiple dimensions [7]. In this work, we seek to understand the risks from potential attacks by analyzing the mechanisms, frequency, and victims of storage security breaches from empirical data. While past experience may or may not be indicative of future attacks, understanding vulnerabilities that are being exploited in the current environment is an important starting point for future improvement. Future attacks are unpredictable, but known risks can be measured to serve as a foundation for looking ahead. Due diligence dictates that security investment to mitigate risks should be based on evidence; otherwise it will expose the organization to continuing breaches and liability from shareholder/customer/third-party lawsuits [8].

The remainder of this paper is organized as follows: Section 2 introduces the current breach disclosure state laws in the U.S. (at the time of publication). Section 3 provides details about the best available data sources we use in this investigation. Section 4 presents statistical processing results (along multiple dimensions) describing the source data along

StorageSS'06, October 30, 2006, Alexandria, Virginia, USA.

with analysis and potential explanations. Section 5 provides a brief overview of related work. We end with a summary and future work in Section 6.

2. THE STORAGE BREACH DISCLOSURE LAWS

In the United States, 28 states have enacted storage security breach laws (at time of publication), see Table 2 in the appendix of this paper. These state laws are similar, but may have different requirements for the notice trigger, timing, content, and recipients [9]. While other federal laws[1] also require reporting of storage security status of various forms, these federal laws are focused on compliance with financial requirements for companies and non-profit organizations to federal regulators. In contrast, when private information is compromised, storage breach state laws typically require only direct notification between the third party organization with the compromise and each affected party, without involvement from federal/state regulators or any level of law enforcement. Private information is defined to be any of the following: social security numbers, drivers license numbers, bank account numbers, credit/debit card numbers as well as any other personal identifying information.

While the compromise of any individual identity has the potential for fraud, it should be noted that experience indicates only a percentage of compromised private data will be involved in identity theft fraud. For example, criminal investigators found only 800 cases of fraud among the 163,000 identities exposed by the ChoicePoint storage security breach in 2004 (less than 0.5%) [6]. Nearly all state laws provide an exemption for breach disclosure if the personal data was encrypted at the time of the compromise [9].

3. DATA SOURCES

Storage breach disclosure laws are currently established only in the United States and are not mandatory in every state. However, even though a majority of states now have breach disclosure laws, disclosure reporting is only required between the organization and the affected parties (employees, customers, etc.) and there is no requirement for public reporting. As a result, there is no comprehensive data source on storage security breaches although there are several lengthy lists of breach incidents maintained on a growing number of websites [1, 3].

Potential costs to an organization for a storage breach reported in the mass media includes damage to reputation, loss of current/future customers, liability from other state's laws, and possible lawsuits from shareholders/customers. In the storage security breaches that have been disclosed, many were reported in the mass media first; thus leading one to infer that many storage breaches, required to be disclosed by law, are not being disclosed unless forced to do so.[2]

No organization has been sued for not disclosing a storage breach they were required by law to disclose. However, several organizations (particularly ChoicePoint) have been sued for negligence by parties affected by storage breaches after disclosure. This provides a strong additional economic incentive not to disclose storage breaches – hopefully this may change with future litigation.

Since there is not a standard format for disclosures, information that would be valuable for analysis is reported inconsistently and often not reported at all. In this paper, we have attempted to provide the best available view of disclosed storage breaches by merging data from the two leading sources of storage breaches: PrivacyRights.org [1] and Attrition.org [2]. The time period of analysis is between January 1, 2005 and June 5, 2006. PrivacyRights.org has 182 storage breach incident reports for this period. For each report, this data source provides date of the incident, organization name, type of breach, and number of records lost. Attrition.org has information on 183 storage breach incident reports for this period. For each entry, it lists the following information: date, organization name, type of business, specific information about the business, type of data, specific nature of data, whether a third party was involved in data handling and loss, total records lost, and a reference to the notification or news item related to the breach. We merged the two databases into a single one, which ultimately contained 219 breach reports for the time period.

The database of storage breaches 2005-2006 upon which our analysis is based is available for query via the Internet at the following URL: *http://dais.cs.uiuc.edu/~rhasan/breachdb*.

4. ANALYSIS

We analyze the data set obtained from the combined two data source and represent the data in various graphical formats in order to communicate the essence of storage breach events along multiple dimensions. Unless otherwise noted, all values are rounded to the nearest integer.

Table 1 shows a statistical overview of the data set during the time interval under study – the mean, median, standard deviation, 95% confidence interval around the mean, and high/low values for the frequency distribution of storage breach incidents and total records lost distribution (per month and per incident). The large standard deviations are due to two large breach incidents which skew the variation statistics.

4.1 Type of Organizations

Fig. 1 and 2 show the number of reported storage breach incidents for different types of organization. The frequency of such incidents is the highest in case of educational institutions (35%), which may be due to a combination of lax security and more openness in reporting. Businesses have incentives not to report breach incidents, so the number of events reported by them is likely low; but it is currently impossible to determine how low. By volume, the second tier of organizations reporting incidents are medical institutions, state government agencies, and banks. The third tier of organizations reporting incidents are the Federal government, data brokers, and organizations (profit/non-profit). This third tier of organizations have large constituencies retaining large volumes of private information, but also more restricted scopes for transactions. Grouping organizations by these three tiers is consistent for other statistics we report.

Fig. 3 provides the insight that, even though educational

[1]Federal laws relevant to reporting storage security status include: Sarbanes-Oxley, Gramm-Leach-Bliley, and HIPAA.
[2]As one example, ChoicePoint first disclosed its 2005 breach only to California residents which had the first disclosure law in the nation and only later disclosed to residents in other states as new state laws were enacted.

Disclosure Statistics	Frequency of Disclosures per Month	Record Size per Month	Record Size per Incident
Mean	12.1	5.74M	589K
Standard Deviation	5.68	14.9M	3.8M
95% Confidence Interval around Mean	9.48 – 14.7	0 – 12.6M	26.6K – 1.15M
Median	12	913K	20K
High	21	57.8M	40M
Low	2	42K	13

Table 1: Overview of Statistical Information.

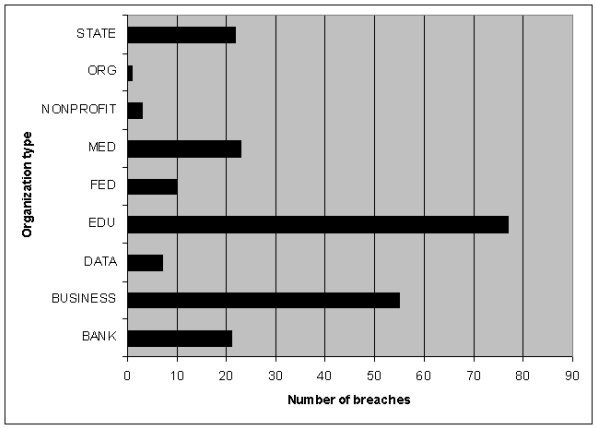

Figure 1: Reported storage breach incidents by organization.

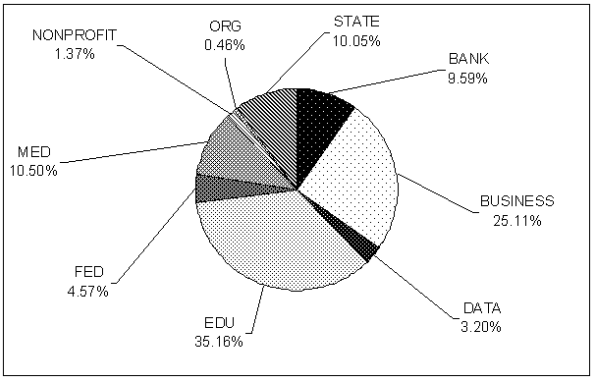

Figure 2: Breakdown of storage breach incidents by organization.

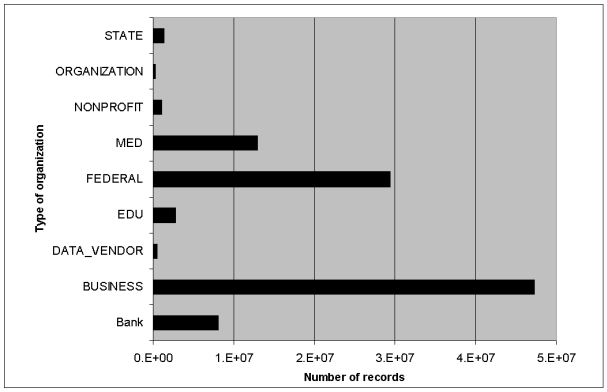

Figure 3: Reported records lost count by organization type.

institutions report more breach incidents, the total number of records lost by educational institutions is roughly an order of magnitude less than the total number of records lost from businesses. Fig. 4 shows that, when considering a percentage breakdown of all records lost categorized by organization type, it is 36% business vs. 3% educational institution. While the Federal government is a third tier organization by breach incident volume, it is a first tier organization by breached record volume – indicative of fewer, but larger breach events in record volume. The Federal government, with only 10 reported incidents in the time period, contributed to almost 30% of total records lost. Medical institutions and banks remain second tier organizations by both breach incident volume and breached record volume.

4.2 Type of Data

We categorized the type of records into the following data type categories: social security numbers (SSN), names and addresses (NAA), credit card numbers (CCN), medical records (MED), account information (ACC), tax information (TAX), passwords (PASS), miscellaneous data (MISC), and unknown records (UNK). From Fig. 5 we see that social security numbers were by far the most common data type stolen or lost (by volume, 62% of records lost). Note that in about 50% of the reported incidents, more than one type of data were among the lost/stolen records (Fig. 5 percentages add to more than 100%).

4.3 Type of Breach

Fig. 6 presents a breakdown of the different types of breach mechanisms showing that 41% of breaches occur via external intrusion, a system breach or other type of malicious attack by external entities. The next most common type of breach is physical attack (covering 36% of total breaches), the loss or theft of media (tapes, hard drives, portable drives) or hardware (laptops, computers). Data breach due to misconfiguration occurs in 12% of total breaches, where data records were inadvertently exposed (e.g. on the web, via email, or database query). Insider attacks, frequently cited as the primary computer network security risk, is found in only 9% of all breaches. Accidental data loss via offline methods (e.g. SSNs printed on driver licenses or mailing labels) occur in only 3% of all breaches although they are typically large incidents affecting many people and sensational media stories when they do occur.

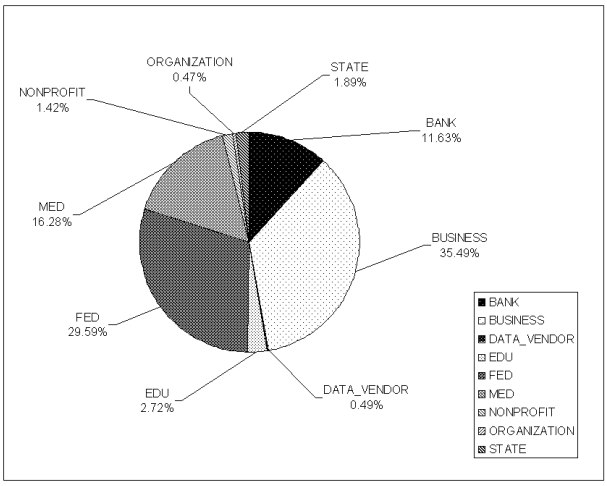

Figure 4: Percentage of reported storage records lost by organization type.

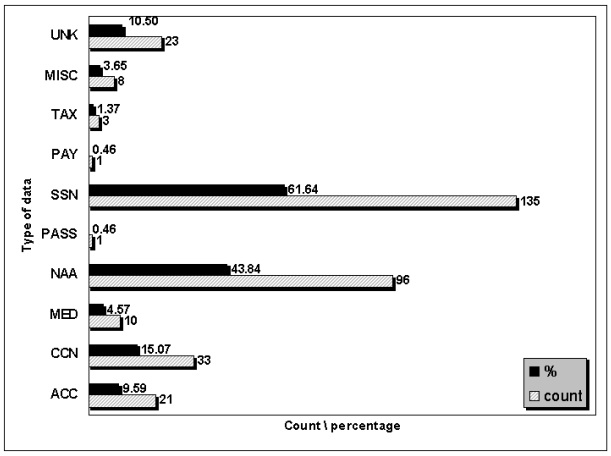

Figure 5: Reported breaches by data type (by volume).

4.4 Times of Breach

Fig. 7 presents a breakdown in the time dimension of the number of reported breach incidents per month. Interestingly, the number of breaches in time shows a periodic pattern – with a peak attained in June 2005 followed by a trough in October 2005, before peaking again in February 2006.

We posit two possible explanations that may work in combination to explain this pattern in time: (1) since educational institutions report the most incidents, there may be a link between breach incidents and the academic calendar and (2) after a particularly large storage breach event is reported (especially in the mass media) then organizational security processes are temporarily tightened, breach incidents temporarily decline, and then over time the number of breach incidents gradually increases as security processes gradually loosen until the next large storage breach (and the cycle continues).

Fig. 8 shows the percentage of number of records affected

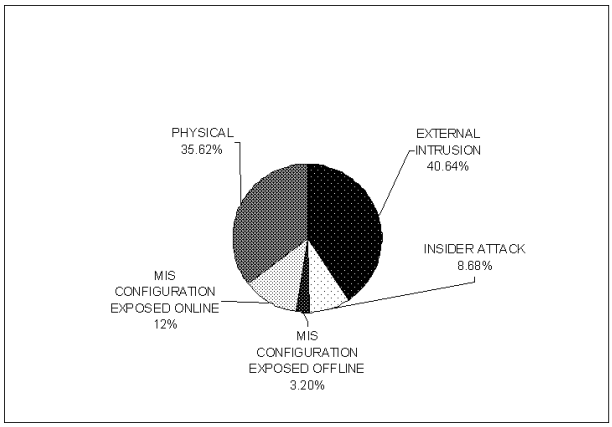

Figure 6: Type of breaches (breach mechanism).

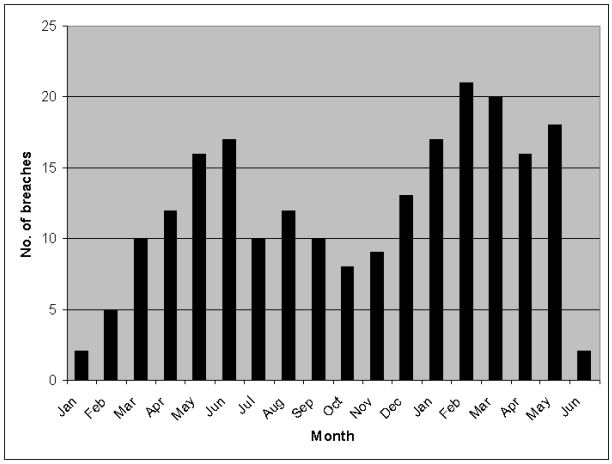

Figure 7: Breach Incidents per Month, January 2005 - June 2006

per month. The figure shows two spikes - one in June 2005 and the other in May 2006. The former refers to a breach of CardSystems, resulting in the loss of 40 million credit card records. The latter is the breach of social security numbers and other private information by the U.S. Department of Veterans Affairs. Fig. 9 presents the record loss per month over time on a log scale to better visualize non-peak months. The average loss in records/month is on the order of 10^6 (a mean of exactly $5.74M$ from Table 1).

4.5 Breach Sizes

By projecting the data set into scatter diagrams, we attempt to provide a better understanding of the relationships between the loss size (in record volume) of individual breach incidents and other dimensions. Fig. 10 presents record size lost over time showing peak events in early summer and a continuous clustering at mid-levels throughout the year. Fig. 11 presents record size loss by breach type showing physical breaches have a tendency toward larger loss incidents and both physical/external breaches clearly are more frequent across the spectrum of loss sizes. Inside attacks occur at the lowest loss sizes and have the widest range. Of-

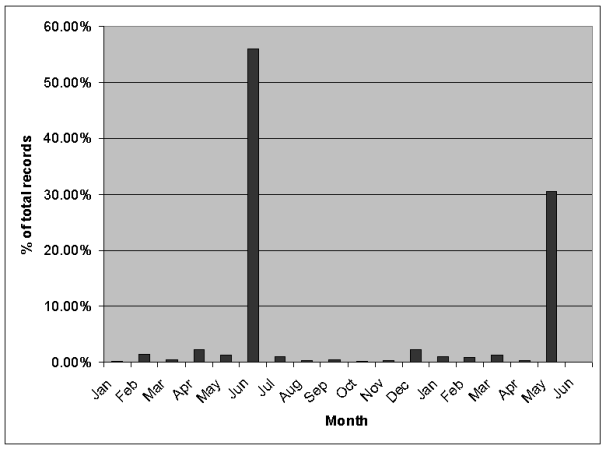

Figure 8: Records Lost per Month, January 2005 - June 2006.

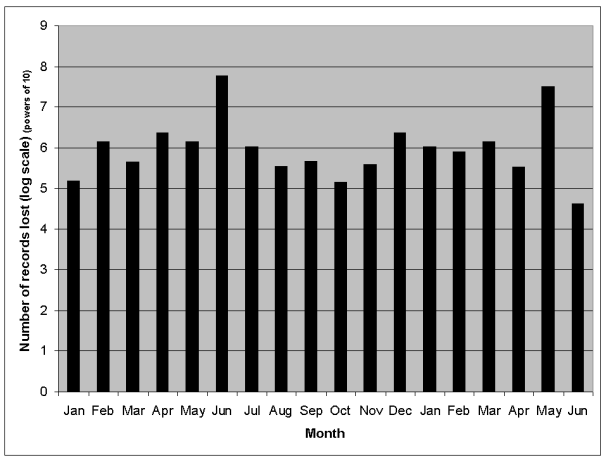

Figure 9: Number of Records Lost per Month (Log Scale), January 2005 - June 2006.

fline/online exposure are sparsely distributed at mid-levels. Fig. 12 presents record size loss per organization type showing education and businesses similarly clustered with more events than other organizations although businesses have several higher volume events without counterparts in educational organizations. Fig. 12 presentation in a scatter diagram clearly highlights that incidents for the following types of organizations are sparsely distributed as exhibited by the small number of incidents and lack of overlapping incidents: profit/non-profit organizations, Federal government, and data broker.

5. RELATED WORK

We are aware of only three related efforts to analyze storage security breaches. First, in [11] the authors summarize selected storage security incidents reported in the press since 2000 and make some claims that we will examine in this section. At present [11] is limited for analysis due its small data set of incidents and biased sampling but the authors themselves state the report is only a start and will be regularly

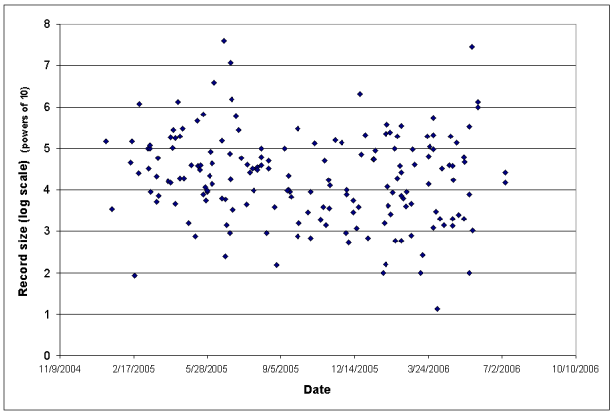

Figure 10: Scatter diagram for Number of Records Lost over Time.

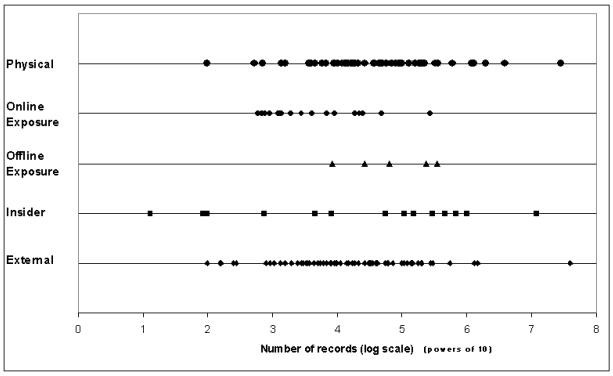

Figure 11: Scatter Diagram for Number of Records Lost by Breach Type.

updated in the future – time will determine the ultimate value of this work.

Tehan *et al.*[11] claims that *almost half of the security breaches occurred at institutions of higher education.* Fig. 2 shows that, considering the total number of breach incidents, educational institutions indeed are the largest (with 35% of total breach incidents). While our percentages are different, we validate this claim that educational institutions are the source of most storage breach incidents. [11] also claims that, *In 2005, a stolen computer (desktop, laptop, or hard drive) was the cause of the security breach 20% of the time.* Our analysis in Fig. 6 shows that 36% of breaches were due to physical attacks including laptop theft (among other types of theft) so our results are again consistent with this claim.

Second, a report from the State Government of California [4] recommends best practices for organizations responsible for protecting personal information including making breach notifications to individuals. In addition to recommendations, the report also includes lessons learned from studying breach notifications in California. It makes several claims based on the experience of being the state with the oldest breach disclosure law. The report suggests more precautions should be taken to prevent physical attacks, the most prevalent form of

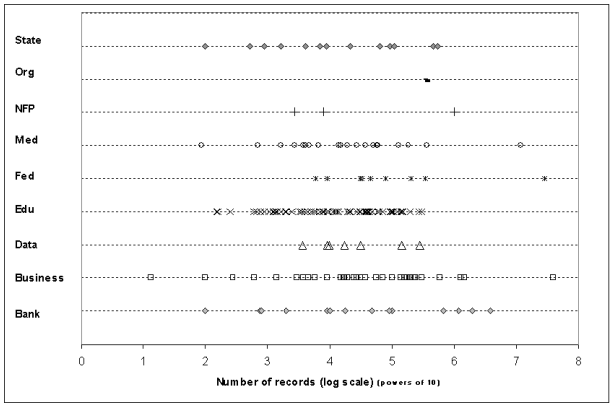

Figure 12: Scatter Diagram for Number of Records Lost by Organization Type.

storage breach in California at 53%. As shown in Fig. 6, our results find that external intrusions are the most prevalent type of storage attack nationwide (41%) followed in second place by physical attacks at 36%. Next, the report claims that, in California, social security numbers are the most common type of data lost in breaches (at 85%). Fig. 5 shows our results are consistent with this claim in that social security numbers are the most common type of data lost nationwide at 62% (in terms of record volume).

Third, [5] studies the impact of security breaches on stock market valuations. The events used in [5] were thus limited to those affecting only publicly traded firms and includes different types of security breaches not limited to storage breaches which disclosure private information.[3] While businesses listed on stock exchanges are an important, they are still only part of the complete storage breach picture. By considering other types of security events and without considering private businesses, non-profit organizations (e.g. universities, hospitals, etc.) and government agencies, the data analysis in [5] presents a partial/skewed view of storage breach events. As our results in Section 4 show, educational institutions report the largest number of storage breach incidents and governments have reported some of the largest individual breach events in terms of records lost so not including these two types of organizations would significantly bias any claims about storage breaches.

6. SUMMARY

Private data on networked devices will always be subject to some risk, but this level of risk can be understood and controlled at a cost. This paper presents empirical evidence of disclosed storage breaches (January 2005 to June 2006) that can be used to assess risk of storage security breaches which release private information. Decisions on type and level of storage protection should be based on such evidence along with the trade-offs and risk posture unique to different environments. To our knowledge, this is the first comprehensive analysis of disclosed storage breaches and it is our hope it will be the first of many more studies. Continuous work is needed to better understand protecting private information in networked environments, especially given the dynamic nature of storage systems and attacks on these same systems.

[3]the [5] data sources include websites, mailing lists, news feeds, and blogs and was not made publicly available.

Acknowledgments

First, we acknowledge special insights on storage security and data breaches from (in alphabetical order): Arshad Noor (StrongAuth, Inc.), Umash Prasad (State of California Criminal Justice Statistics Center – Special Requests Unit) and Professor Marianne Winslett (UIUC). We also thank the constructive criticism of anonymous StorageSS reviewers which we have incorporated to improve this paper.

This paper is, in part, based upon work supported by the Technology Research, Education, and Commercialization Center (TRECC), a program of the University of Illinois at Urbana-Champaign, funded by the Office of Naval Research and administered by the National Center for Supercomputing Applications (NCSA). Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Office of Naval Research.

7. REFERENCES

[1] A chronology of data breaches reported since the choicepoint incident (list). *Privacy Rights Clearinghouse* http://www.privacyrights.org/ar/ChronDataBreaches.htm.

[2] Dataloss mailing list. *Attrition.org* http://attrition.org/security/dataloss.html.

[3] Entities that suffered large personal data incidents (list). *Attrition.org* http://attrition.org/errata/dataloss.

[4] Recommended practices on notice of security breach involving personal information. *State of California Department of Consumer Affairs/Office of Privacy Protection*, April 2006.

[5] A. Acquisti, A. Friedman, and R. Telang. Is there a cost to privacy breaches? an event study. In *Workshop on the Economics of Information Security (WEIS)*, 2006.

[6] C. Conkey. Identity theft: Shielding yourself. July 14, 2006.

[7] R. Hasan, S. Myagmar, A. J. Lee, and W. Yurcik. Toward a threat model for storage systems. In *ACM International Workshop on Storage Security and Survivability (StorageSS)*, pages 94–102, 2005.

[8] M. Hines. Data losses may spark lawsuits. In *eWeek*, June 12, 2006.

[9] P. Mueller. How to survive data breach laws. *Network Computing*, June 8, 2006.

[10] B. Schneier. Risks of third-party data. *Communications of the ACM*, May 2005.

[11] R. Tehan. Personal Data Security Breaches: Context and Incident Summaries. In *Congressional Research Service Report for Congress*, December 16, 2005.

States	Start Date	State Law	Responsible Party	Likelihood of Harm Threshold	Best Practices Required
(1) California	07/01/03	SB 1386	entities conducting business, separate section for state agencies	no	yes
(2) Arkansas	03/31/05	SB 1167	entities conducting business	yes	yes
(3) Georgia	05/06/05	SB 230	data brokers only, excludes state agencies	no	no
(4) North Dakota	06/01/05	SB 2251	entities conducting business	no	no
(5) Delaware	06/28/05	HB 116	entities conducting business	no	no
(6) Florida	07/01/05	HB 481	entities conducting business	yes	no
(7) Tennessee	07/01/05	HB 2170	"information holder" including people, business, or state agency	yes	no
(8) Washington	07/24/05	SB 6043	any person or business, plus state agencies	yes	no
(9) Texas	09/01/05	SB 122	a person that conducts business	no	yes
(10) Nevada	12/01/05	SB 347	data collectors, including all entities and state agencies	yes	yes
(11) North Carolina	12/01/05	SB 1048	any person or state agency	no	no
(12) New York	12/08/05	SB 5827	any person or business	no	no
(13) Connecticut	01/01/06	SB 650	any person that conducts business	yes	no
(14) Illinois	01/01/06	HB 1633	data collectors, including all entities and state agencies	no	no
(15) Louisiana	01/01/06	SB 205	any person or agency	yes	no
(16) Minnesota	01/01/06	HF 2121	entities conducting business, section for state agencies	no	no
(17) New Jersey	01/01/06	A4001	a business or public entity	yes	yes
(18) Maine	01/31/06	LD 1671	data brokers only, excludes state agencies	no	no
(19) Ohio	02/15/06	HB 104	any person or state agency	yes	no
(20) Montana	03/01/06	HB 732	entities conducting business, plus special requirements for insurers	yes	yes
(21) Rhode Island	03/01/06	HB 6191	any state agency or person, including all businesses]	yes	yes
(22) Wisconsin	03/31/06	SB 164	entities conducting business	no	no
(23) Oklahoma	06/08/06	HB 2357	only state entities	no	no
(24) Indiana	06/30/06	503	person or government agency	no	no
(25) Pennsylvania	06/30/06	SB 712	any entity	yes	no
(26) Idaho	07/01/06	28-51-104	entities conducting business	yes	no
(27) Nebraska	07/13/06	LB 876	entities conducting business	yes	no
(28) Colorado	09/01/06	6-1-7161a	entities conducting business	yes	no
(29) Arizona	12/31/06	SB 1338	entities conducting business	yes	yes
(30) Hawaii	01/01/07	SB 2290	entities conducting business	no	no
(31) Kansas	01/01/07	SB 196	entities conducting business	yes	no
(32) New Hampshire	01/01/07	HB 1660	entities conducting business	yes	no
(33) Utah	01/01/07	SB 69	entities conducting business	yes	no
(34) Vermont	01/01/07	SB 284	entities conducting business	no	no

Table 2: Summary of State Laws for Privacy Breach Disclosures adapted from: (1) "State Laws Governing Security Breach Notification", Crowell Moring LLP, 01/25/06. http://www.crowell.com/; (2) "Security Breach Notice Legislation: Effective Dates, and Security Breach Notification Chart," Perkins Cole Attorneys Al Gidari, Barry Reingold, and Matt Staples; and (3) "Notice of Security Breach State Laws," Consumer Union, June 27, 2006.