Some Observations on the Occurrences of Data Breaches

Mario Guerra
mariog14@gmail.com

Murphy Joseph
murphyj169@gmail.com

Ernest Ramirez
ernestramirez79@hotmail.com

Kai S. Koong
koongk@utpa.edu

Lai C. Liu
liul@utpa.edu

Computer Information Systems and Quantitative Methods Department
The University of Texas Pan American
Edinburg, Texas 78539, USA

Abstract

This study examines the occurrences of data breaches that were reported to have occurred in the United States. Specifically, this research deals with all those data breaches that were reported to Privacy Rights Clearinghouse (PRC) from 2005 to the first quarter of 2009. First, based on the number of breaches reported and the number of records compromised, businesses represented the leading category affected. The rate of breaches being reported is slowly decreasing. Like in the case of previous studies, laptops are the most commonly stolen hardware reported.

Keywords: data breaches, computer security, data breaches trends, and data breaches characteristics

1. INTRODUCTION

Information security has become a very important subject. It was once stated, "In this information-saturated age, the use of personal data has significant consequences for every American" (Carlson, 2005, p.26). It is now common for websites to require registration of some sort, especially for the purpose of performing a transaction. Others require some type of a registration to use their services like searching for a job, online

gaming and even online dating. Providing all that personal information to someone can be a very dangerous thing to do because talks of Internet fraud, identity theft, and credit card fraud have been circulating the nation for quite a while. Many Web users will remember that "2007 was an unprecedented year for data security breaches when the subject made the headlines regularly throughout the year" (Meadowcroft, 2008, p.10). Actually, these occurrences are constantly on the rise and it has become a real threat to users as well as institutions of all types.

2. STATEMENT OF THE PROBLEM

Data breaches have become a major issue this day and age. They can generally be referred to as an "organization's unauthorized or unintentional exposure, disclosure, or loss of sensitive personal information, which can include personally identifiable information such as Social Security numbers (SSN) or financial information such as credit card numbers" (The Government Accountability Office, 2007, p. 2). It is no surprise that research indicates that data breaches have serious financial consequences on an organization (Ponemon, 2009). For example, ChoicePoint, an identification and credential verification service company, was breached and that resulted in the loss of $27.3 million in 2005 alone to cover legal fees, notify victims, and seek audits. Poor policy was found to be part of the problem (Otto, Anton & Baumer, 2007). With the loss of customers and monetary value among others, Since then, ChoicePoint has made numerous changes in its policies and procedures.

Similarly, the retailer parent company of T.J. Maxx and several other chains stores took "a $12 million charge in its first fiscal quarter of 2008 related to the loss of more than 45 million credit and debit card numbers stolen from its IT systems over an 18-month period" (Gaudin, 2007, p.19). More recently, Heartland Payment Systems also became "the victim of a security breach within its processing system", possibly part of a "global cyber fraud operation" in January 2009 (Wikipedia, 2009). It was the largest criminal breach of card data ever, compromising up to 100 million cards from more than 650 financial services companies. These are just three relatively well known examples that show how expensive breaches can be to a company.

A data breach can cause detrimental effects to an organization, and in some cases can cause them to declare bankruptcy. Not only does the organization have losses occurred because of the information being compromised, they can also suffer from clients losing trust and causing them to cease doing business with them. Even though data may get recovered, the fact that it happened is a detail that will always be remembered. When a company announces publicly that they have encountered a data breach, one of several actions must take place. "Companies often must hire security consultants, engage legal counsel and offer credit monitoring services to affected customers" (Knowledge@Wharton, 2009). A study conducted by the Ponemon Institute in 2008 reported that the cost of a data breach averaged $202 per customer record. The average total cost per reporting company in 2007 was $6.3 million and $4.7 million in 2006 (Ponemon, 2009). According to Ponemon, 65% of these costs are incurred because of lost business. With these outrageous numbers associated with the cost of a data breach, it is apparent that this is a major issue.

There are many reasons as to how the data breaches occurred. They range from theft, malicious attacks, dishonest employees, and negligence (A Chronology of Data Breaches, 2009). Securing an organization is a difficult task. Protection methodologies are constantly changing because people are always looking for loopholes to find their way in. What is effective six months ago could be obsolete today. In order to ensure a secure environment, an organization must build strong policies and continue to revise them as time passes. Obviously, many of the breaches reported may have incurred less damage than they actually were affected had stricter and more dynamic policies were in place.

3. STATEMENT OF OBJECTIVE

This study examines the occurrences of data breaches that occurred in the United States. Specifically, this research deals with all those data breaches that were reported to

the Privacy Rights Clearinghouse (PRC), a non-profit consumer information and advocacy organization. The website of this organization can be located at http://www.privacyrights.org. Data reported to PRC during the period of 2005 to the first quarter of 2009 was analyzed to show the types and magnitude of data breaches. This research would be beneficial to information systems educators, information security experts, law enforcement persons, and consultants.

Based in San Diego, California, PRC was established in 1992 and its goals are as follows (Privacy Rights Clearinghouse, 2009):

- Make consumers aware of how technology affects personal privacy.
- Assist consumers to take action to control their own personal information by providing practical tips on privacy protection.
- Respond to specific privacy-related complaints from consumers, intercede on their behalf, and, when appropriate, refer them to the proper organizations for further assistance.
- Document the nature of consumers' complaints and questions about privacy in reports, testimony, and speeches and make them available to policy makers, industry representatives, consumer advocates, and the media.
- Advocate for consumers' privacy rights in local, state, and federal public policy proceedings, including legislative testimony, regulatory agency hearings, task forces, and study commissions as well as conferences and workshops.

According to disclosures on its organizational website, the major services provided by PRC include (Privacy Rights Clearinghouse, 2009):

- A hotline for consumers to report privacy abuses and request information on ways to protect their privacy.
- An extensive series of fact sheets on privacy issues, available in English and some in Spanish.
- A web site (www.privacyrights.org) that provides texts of all fact sheets, transcripts of PRC speeches and testimony, FAQ and index by topic, stories of consumers' experiences, and more.
- Assistance and interviews for journalists, providing background and comments for stories.
- A referral service for journalists and policymakers who are seeking victims of privacy abuses who have indicated a willingness to talk with the media and/or testify in legislative and regulatory agency hearings.
- A speakers service, in which PRC staff make presentations at conferences, employee training sessions, and civic and community group meetings.

4. METHODOLOGY

The institutions that reported a data breach in the United States during the period of 2005 through the first quarter of 2009 are selected as the targeted population of this research. The data were collected using a chronology of data breaches that was retrieved by the Privacy Rights Clearinghouse (A Chronology of Data Breaches, 2009). In order to study the magnitudes and types of data breaches, the following data were extracted from the report for a total of seventeen quarters:

- type of industry that was affected by years and quarters
- type of breach that occurred by years and quarters
- number of reported incidents by years and quarters
- number of records reported being lost by years and quarters
- locations where the data breaches occurred.

The chronology did not list the type of industry that was affected directly; therefore each record was analyzed and classified according to the primary deliverable of the institution itself. The organizations were classified into seven different areas; business, federal government, education, state government, county government, city government, and medical. Institutions classified as federal government included all government offices that are headquartered in Washington. Education institutions included any type of schools, colleges and universities, and school district. State government, County government, and City government are agencies that are run at their respective levels. Medi-

cal institutions included hospitals, clinics, or cases where medical records were lost. All others are labeled as businesses and they can range from a mom-and pop operation to a retail giant. On a few occasions, a breach that occurred could have fallen into more than one category and was classified as such.

The type of breach that occurred was the next category that was analyzed for every breach. According to the PRC, the listing stated what occurred to cause the breach, and categorized it as loss of hardware, external breach, or internal breach. Hardware was categorized as a breach that occurred because of a laptop, hard drive, flash drive, computer, back up data tape, or storage device being reported as lost, stolen, or missing by the site. External breaches included cases of a third party retrieving the data without authorization such as hacking, or losing physical files that caused the records to be compromised. Lastly, internal breaches included errors of the parties within the company. These include exposal of records online, mailing errors through e-mail or postal system, improper disposal, security lapse caused by an employee, or dishonest insiders.

Data breaches by location were another area of interest in which this study examined. Regional classification was used to classify the data by location because there was insufficient information to classify by state. The following regions were used and this classification system was modeled after the Population Division of the US Census Bureau: West, Midwest, Northeast, and South.
The Northeast region:

- Connecticut, District of Columbia Massachusetts, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island and Vermont.

The Midwest region:

- Illinois, Indiana, Iowa, Kansas, Ohio, Michigan, Minnesota, Missouri, Nebraska, North Dakota, South Dakota and Wisconsin.

The South region:

- Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, Louisiana, Oklahoma, Texas, Virginia and West Virginia.

The West region:

- Arizona, Alaska, California, Colorado, Idaho, Montana, Nevada, New Mexico, Hawaii, Oregon, Utah, Washington and Wyoming.

Each data breach was examined and categorized by the items mentioned. First, the number of each incident was recorded according to the type of breach and what occurred for that specific quarter. Second, the number of records that were compromised of each quarter was documented. Finally, the location where the incident occurred was also recorded.

5. FINDINGS

According to the breaches that were recorded by the Privacy Rights Clearinghouse, there were some 1,120 reported incidents and over 300 million known records compromised. This led to an average of about 307 thousand records lost per breach. This number should be interpreted with caution due to the fact that the total number was brought about by several outlying cases. There was also a considerable amount of incidents where the amount of records was unknown. Roughly 26% of these records were categorized as such.

When analyzing the data, there was substantial increase in the amount of incidents from the base year of 2005 to the first quarter of 2006. Since the escalation in 2006, the number of incidents is slightly decreasing and could possibly level out. This may indicate that the public have become more aware of the importance and value of private information. However, despite the awareness, the number of cases is showing a cyclical pattern, a spike in the number of cases to about 7 percent every 3 or 4 quarters.

Next, the data set was tallied to show the total number of incidents per region. The South region has 34.2 percent of the total number of breaches which was the highest amount during the 17 quarters. The West region had the second highest number with 24.5 percent. The Midwest was third with 22.8 percent, and the Northeast reported the smallest number of incidents with 18.6 percent. This number appeared to correlate with the population density of the regions. Details about the number of data breaches in the respective regions are shown in Table 1 in the Appendix.

Table 2 (see Appendix) showed the number of data breaches incidents reported per industry. As mentioned before they were classified into seven different categories. Business was the category that was most affected with 409 incidents reported or 35.6%. Education was the next highest industry affected with 329 incidents reported or 28.7%. Medical was third on the list with state run agencies slightly behind. There was a limited occurrence of government, city, and county breaches reported. Based on the results reported during the period, business and education institutions appear to be the higher targeted industries. Business showed a very small growth over the periods in comparison to education which displayed a significant increase. Medical facilities actually had a slight decrease showing that they are becoming less of a target in comparison to their more public counterparts.

When the types of breaches were analyzed, it was found that 41.2 percent of the cases were hardware related. This meant that a physical asset was either reported lost, stolen, or missing. This may contribute to the association between the population and reported incidents within the regions; the higher the population, more theft, loss, or misplacement was reported. Hacking was the second highest of reported incidents. Although hacking was second, it did have the highest number of records reported compromised. From the number of known records lost, there were over 123 million compared to the 77 million for hardware. A surprising discovery was the number of reported incidents about online exposure. Fifteen percent of the breaches were caused by personal information being exposed on public websites with over 80 million records being promised placing it second in the list of known records lost. Other details about the types of breaches reported are presented in Table 3 (see Appendix A).

After identifying stolen, lost, or missing hardware as having the highest occurrence, the type of hardware reported was examined. As anticipated, laptops were the most reported type of hardware that was responsible for a breach. Computers as were a distant second and data tapes were the third most stolen. Table 4 (see Appendix A) shows the categories of lost or stolen hardware and their reported frequency. Another statistic that was recorded for the hardware was if it was stolen by a physical break-in. Twenty-four percent of these incidents were recorded as such. Another interesting statistic was found, out of the 445 pieces of hardware that were reported as lost, missing, or stolen, only 2 percent of them were reported to have their information encrypted.

6. CONCLUSIONS, LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

Personal information is important and must be guarded with care. Organizations need to maintain strict policies to ensure the protection of this data. Based on the outcomes in the four tables and the data analyzed, there are several conclusions that can be made about the reported data breaches. First, whether it is measured by the number of breaches or number of records compromised, businesses are the number one industry that has been affected. Second, while the rate of breaches being reported is slowly decreasing it does not mean they can let their guard down because the data breaches incident is drastically increasing in education. In other words, this is still a growing problem because the preferred target has changed or a new group of perpetrators have surfaced. Third, only 2 percent of hardware that was reported lost, stolen, or missing was noted to be encrypted. This illustrates that although policies may exist to protect information leaving a facility; policies are not proactively being implemented to protect those information assets.

Another alarming discovery was the lack of internal regulation of publicly exposed information. The second highest total number of records reported being lost was in the area of online exposure. This goes to show that many online vendors are not implementing proper policies to safeguard what is being displayed to the world. It is the responsibilities of end-users of online systems to exercise extreme caution when providing critical personal information to web-based vendors.

Several limitations arose while conducting this study. First, the data collected is re-

stricted to occurrences reported to the Privacy Clearinghouse for the period 2005 to the first quarter of 2009 and the way it is collected. While this was a sufficient amount of data, it is not a comprehensive list of all data breaches. Since only publicly traded companies are required by law to report such breaches, many cases of data loss go unreported. Secondly, the Privacy Rights Clearinghouse only published data breaches that actually had personal information stolen or compromised. Only a minor amount of breaches where no personal information was compromised were included in this study for the sole purpose of establishing the frequency of data breaches. Finally, this study is targeted at occurrences that were reported in the US and the way it is recorded. The actual effects of data breaches around the world may be different. Other limitations include the reliability of those occurrences. Privacy Rights Clearinghouse does not state whether each data breach reported is a single entity or if more than one occurrence was reported multiple times. Also there were a few records that were listed as unknown regarding the type of breach and what was stolen. The actual number of records and type of breaches may be higher than what is reported in this study.

Despite all limitations listed this study is useful in indicating the types of data breaches that have occurred, the types of industries that have suffered data losses, and the size of data loss. Future research could be geared towards developing a better way of collecting the data. In particular, laws could be established for all companies to report data breaches.

7. REFERENCES

"A Chronology of Data Breaches." (2009, July 28) Available at Privacy Rights Clearing House Website at http://www.privacyrights.org/ar/ChronDataBreaches.htm

Carlson, C. (2005) "Teed up for '06: Data Breaches Spyware." eWeek, Vol. 22, No. 47, p. 26.

Gaudin, S. (2007) "Breach Costs Soar At TJX." Information Week, p. 19.

Knowledge@Wharton (2009, March 18) "A Time for a Data Diet? Deciding What Customer Information to Keep – and What to Toss." Retrieved on June 29, 2009 from http://knowledge.wharton.upenn.edu/artic-lepdf/2186.pdf?CFID=8448159&CFTOKEN=88031812&jsessionid=a830aa4cfbe6bad49ca0cd595e2043106410

Meadowcroft, P. (2008) "Card Fraud – Will PCI-DSS Have." Card Technology Today, pp. 10-11.

Otto, Paul, Annie I. Anton and David L. Baumer (2007) "The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information." IEEE Computer Society, Vol. 5, No. 5 (September/October), pp. 15-23. Available at http://doi.ieeecomputersociety.org/10.1109/MSP.2007.126

Ponemon, Larry (January, 2009) "Fourth Annual US Cost of Data Breach Study." Available at http://www.ponemon.org/local/upload/fckjail/generalcontent/18/file/2008-2009%20US%20Cost%20of%20Data%20Breach%20Report%20Final.pdf

Privacy Rights Clearinghouse (2009) "PRC Mission & Goals." Available at http://www.privacyrights.org/about_us.htm.

The Government Accountability Office. (2007) "Data Breaches and Identity Theft." The Government Accountability Office.

Wikipedia (2009) "Data Breaches." Available at http://en.wikipedia.org/wiki/Data_breach

Appendix A.

Table 1 Data Breaches Incidents by Region

	Midwest	Northeast	South	West	Total	%	Ranking
2005 Q1	5	1	3	9	18	1.6%	16
2005 Q2	12	6	9	9	36	3.2%	13
2005 Q3	3	2	4	14	23	2.1%	15
2005 Q4	3	4	3	7	17	1.5%	17
2006 Q1	8	9	7	11	35	3.1%	14
2006 Q2	19	15	31	13	78	7.0%	8
2006 Q3	22	13	43	26	104	9.3%	1
2006 Q4	24	14	31	26	95	8.5%	4
2007 Q1	22	11	23	25	81	7.2%	7
2007 Q2	24	15	33	24	96	8.6%	3
2007 Q3	20	17	22	18	77	6.9%	9
2007 Q4	16	19	20	9	64	5.7%	11
2008 Q1	15	18	32	19	84	7.5%	5
2008 Q2	16	27	37	17	97	8.7%	2
2008 Q3	19	10	32	15	76	6.8%	10
2008 Q4	9	7	25	14	55	4.9%	12
2009 Q1	18	20	28	18	84	7.5%	5
Total	255	208	383	274	1,120	100.0%	
%	22.8%	18.6%	34.2%	24.5%			

Table 2 Data Breaches Incidents Reported by Industry

	Business	Federal Government	Education	State	County	City	Medical	Total
2005 Q1	6	0	10	1	0	0	1	18
Q2	8	0	6	2	0	0	2	18
Q3	1	1	15	0	1	0	0	18
Q4	8	0	7	0	0	0	3	18
Total	23	1	38	3	1	0	6	72
%	31.9%	1.4%	52.8%	4.2%	1.4%	0.0%	8.3%	
2006 Q1	21	2	9	7	1	2	3	45
Q2	27	9	21	11	3	0	11	82
Q3	42	12	18	12	5	4	18	111
Q4	34	7	21	7	7	11	14	101
Total	124	30	69	37	16	17	46	339
%	36.6%	8.8%	20.4%	10.9%	4.7%	5.0%	13.6%	
2007 Q1	25	8	24	12	1	3	10	83
Q2	32	3	34	10	4	5	9	97
Q3	31	4	19	10	3	4	8	79
Q4	21	4	20	9	4	0	10	68
Total	109	19	97	41	12	12	37	327
%	33.3%	5.8%	29.7%	12.5%	3.7%	3.7%	11.3%	
2008 Q1	32	2	25	13	2	1	14	89
Q2	41	3	32	7	1	4	10	98
Q3	25	2	28	8	2	5	9	79
Q4	16	3	16	9	0	5	8	57
Total	114	10	101	37	5	15	41	323
%	35.3%	3.1%	31.3%	11.5%	1.5%	4.6%	12.7%	
2009 Q1	39	5	24	5	0	5	9	87
Total	39	5	24	5	0	5	9	87
%	44.8%	5.7%	27.6%	5.7%	0.0%	5.7%	10.3%	
Grand Total	409	65	329	123	34	49	139	1148
Overall %	35.6%	5.7%	28.7%	10.7%	3.0%	4.3%	12.1%	

Table 3 Incidents Reported by Data Breach Type

	Dishonest Insider	Exposed	Hacking	Hardware	Improper Disposal	Mail Error	Security Lapse	Total	%
2005									
Q1	2	1	10	5	0	0	0	18	23.7%
Q2	3	0	8	8	0	0	0	19	25.0%
Q3	0	0	15	5	0	0	0	20	26.3%
Q4	0	2	5	10	0	1	1	19	25.0%
Total	5	3	38	28	0	1	1	76	
%	6.6%	3.9%	50.0%	36.8%	0%	1.3%	1.3%		
2006									
Q1	4	5	9	16	2	6	0	42	13.1%
Q2	6	9	22	34	3	1	3	78	24.3%
Q3	4	15	14	57	6	6	4	106	33.0%
Q4	7	12	15	40	10	7	4	95	29.6%
Total	21	41	60	147	21	20	11	321	100.0%
%	6.5%	12.8%	18.7%	45.8%	6.5%	6.2%	3.4%		
2007									
Q1	2	16	12	37	8	4	0	79	25.2%
Q2	5	24	21	32	11	4	0	97	31.0%
Q3	6	13	9	31	5	2	4	70	22.4%
Q4	2	10	12	34	4	3	2	67	21.4%
Total	15	63	54	134	28	13	6	313	
%	4.8%	20.1%	17.3%	42.8%	8.9%	4.2%	1.9%		
2008									
Q1	5	12	19	39	5	5	0	85	27.6%
Q2	9	17	19	30	13	3	2	93	30.2%
Q3	7	15	16	22	4	7	3	74	24.0%
Q4	2	10	12	28	1	2	1	56	18.2%
Total	23	54	66	119	23	17	6	308	
%	7.5%	17.5%	21.4%	38.6%	7.5%	5.5%	1.9%		
2009									
Q1	9	6	19	25	9	7	6	81	
%	11.1%	7.4%	23.5%	30.9%	11.1%	8.6%	7.4%		
Grand Total	73	167	237	453	81	58	30	1099	
Overall %	6.6%	15.2%	21.6%	41.2%	7.4%	5.3%	2.7%		

Note: Stolen/Lost paper documents not included

Table 4 Lost or Stolen Hardware Categories

	Laptops	Computers	Flash Drive	Tape	Disk	Hard Drive	Total
2005 Q1	2	1	0	1	0	1	5
Q2	3	2	0	3	0	0	8
Q3	1	0	0	3	1	0	5
Q4	5	3	0	2	0	0	10
Total	11	6	0	9	1	1	28
Percent	39.3%	21.4%	0.0%	32.1%	3.6%	3.6%	
2006 Q1	11	0	1	1	2	1	16
Q2	17	8	1	5	0	3	34
Q3	36	11	3	5	0	2	57
Q4	22	11	1	0	4	2	40
Total	86	30	6	11	6	8	147
Percent	58.5%	20.4%	4.1%	7.5%	4.1%	5.4%	
2007 Q1	15	12	0	3	2	5	37
Q2	14	5	3	4	5	1	32
Q3	14	9	1	3	4	0	31
Q4	23	5	3	3	0	0	34
Total	66	31	7	13	11	6	134
Percent	49.2%	23.1%	5.2%	9.7%	8.2%	4.5%	
2008 Q1	17	8	5	2	1	6	39
Q2	14	10	2	2	0	2	30
Q3	9	4	4	4	0	1	22
Q4	12	5	1	4	3	3	28
Total	52	27	12	12	4	12	119
Percent	43.7%	22.7%	10.1%	10.1%	3.4%	10.1%	
2009 Q1	14	5	1	2	2	1	25
Total	14	5	1	2	2	1	25
Percent	56.0%	20.0%	4.0%	8.0%	8.0%	4.0%	
Grand Total	229	99	26	47	24	28	453
Percent	50.6%	21.9%	5.7%	10.4%	5.3%	6.2%	