Copyright © 2011 International Journal of Cyber Criminology (IJCC) ISSN: 0974 – 2891
Jan – July 2011, Vol 5 (1): 794–810

Organizational Data Breaches 2005-2010: Applying SCP to the Healthcare and Education Sectors

Jason D. Collins[1]
Vincenzo A. Sainato[2]
David N. Khey[3]
Loyola University New Orleans, USA

Abstract

An exhaustive literature review was performed to assess the current state of organizational data breaches within the United States. Explicitly, this research reviewed the applicability of Situational Crime Prevention, the influence of current breach notification laws, findings drawn from macro-level studies of data breaches, and reporting issues unique to the entities of health care and education. To assess the results of the literature, a six year sample of reported data breaches was compiled from the Privacy Rights Clearinghouse, consisting of a total 2,219 data breaches disclosed between 2005 and 2010. This analysis specifically addressed four individual variables: type of breach, reporting entity, year the breach was disclosed, and the geographic region in the United States where the breach was reported. Bayesian statistics were further employed to create probabilities that determined the likelihood of these four variables within the healthcare and education entities. Ultimately, we concluded that: the passage of reporting legislation within the healthcare field increased the number of incidents reported; breaches reported by educational institutions appear to be on the decline; the lack of a centralized reporting database for all data breaches prevents a definitive analysis of the field; and that situational crime prevention measures can be proactive in preventing future data breaches within these entities.

Keywords: Situational Crime Prevention; SCP; Data Breach; Cybercrime; Computer Crime; Information systems; IT; Bayes; Privacy Rights Clearinghouse; Healthcare; Education.

[1] Graduate Student, Loyola University New Orleans, Department of Criminal Justice, 6363 St. Charles Avenue, Campus Box 55, New Orleans, LA 70118, United States of America. Email: nightjc@gmail.com

[2] Assistant Professor, Loyola University New Orleans, Department of Criminal Justice, 6363 St. Charles Avenue, Campus Box 55,New Orleans, LA 70118, United States of America. Email: vsainato@loyno.edu

[3] Assistant Professor, Loyola University New Orleans, Department of Criminal Justice, 6363 St. Charles Avenue, Campus Box 55,New Orleans, LA 70118, United States of America. Email: dkhey@loyno.edu

Introduction

Situational crime prevention (SCP) is a practical application of routine activity theory (RAT) that reduces the frequency of likely criminal opportunities (Felson & Clarke, 1998). Cohen and Felson (1979) originally proposed RAT when attempting to analyze the increase in crime rates following World War II. Specifically, these researchers developed a criminal theory that focused on the environmental 'opportunities for crime' to occur characterizing the temporal and spatial attributes of a criminal transgression. Their model stated that a criminal act possessed three fundamental variables: (1) a suitable target, (2) a motivated offender, and (3) the absence of a capable guardian. Essentially, when a potential criminal opportunity arises the act will occur at a juncture in time and space between a motivated offender and a suitable target for victimization. This crime will ultimately take place in a location that lacks a capable guardian to protect the 'suitable target,' which is considered to be either a vulnerable person or one's unguarded property. Thus, the absence of any one of these three situational factors should theoretically make the commission of a crime impossible (Davis, 2002). As a result, routine activity theory is considered to be a macro-level theory applicable to numerous types of crime as it seeks to explain the criminal victimization process and not a criminal's specific motivations (Akers & Sellers, 2009).

Information Security (IS) officers can use SCP to reduce the possible motives for an offender to engage in crime by: increasing the effort and risks of a crime; reducing the potential rewards and provocations; and removing the excuses for committing a crime (Willison & Siponen, 2009). SCP achieves these goals through a preventative technique called "target hardening" that identifies the specific situational exploits that allow criminals to commit an offense in a particular area (Felson & Clarke, 1998, p.27). Applications of target hardening include: installing entry-phones to apartment buildings to regulate access; visible security cameras and guards to deter crime; rapid clean up of graffiti to deny the visual benefit to an offender; and requiring registration at the front desk of a hotel to discourage people from leaving without paying.

Willison and Siponen (2009) applied SCP methodology to the information technology (IT) divisions of corporations in an effort to deter employees from stealing valuable information from the company. An effective way of applying SCP techniques is through the creation of 'crime scripts'[4] that outline the various steps potential offenders would need to execute in order to circumvent security measures and gain access to the restricted areas of an organization. Consequently, these scripts allow security officers to devise countermeasures for each step preventing these would-be offenders from taking advantage of ambiguities in an organization's security procedures. For example, if an offender knew where a coworker wrote down their passwords they could potentially access the employee's account without their knowledge. To deter this scenario from unfolding, a crime script would recommend that the company respond by reducing the number of passwords employees have to remember, incorporate biometric technology into login procedures, or mandate that the staff attend refresher classes on basic security protocols.

[4] 'Crime scripts' in this instance should be thought of as a movie script where all of the character's actions and reactions are meticulously planned out. There are also 'computer scripts' that are automated computer programs that commonly execute multitudes of algorithms that can be employed to determine potential vulnerabilities in one's network, system, or server. Conversely, these computer scripts can be deployed by computer criminals to attempt to exploit vulnerable victims.

Thus, incorporating SCP practices into corporate security procedures can be advantageous in reducing the number of deliberate and accidental security breaches.

Highlighting the current procedures for security threats within corporations, Willison (2008) found that roughly fifty percent of the security breaches reported in the 2004 CSI/FBI Computer Crime Security Survey and the 2006 Global Security Survey occurred within the victimized organization. Furthermore, it was discovered that much of the literature on IS does not employ a concrete analytical theory in their research. Approximately 1,280 IS-related articles written between 1990 and 2004 were analyzed by Siponen and Willison (2007) who ultimately discovered that only 237 of these studies applied any form of an analytical theory. Thus, the available literature on employee computer crime within the IS field was primarily focused on advanced computer-based defenses and largely neglected to research the underlying sociological factors involved in the offense. Therefore, Willison (2008) feels that corporate IS would benefit greatly from SCP by developing a new "socio-technical" perspective that integrates both technical knowledge and insight into the conventional routines of authorized employees (p.170).

Federal and State Breach Notification Laws

Presently, at the national level there is no all-encompassing law that governs the security of citizens' sensitive information (see Stevens, 2010). What the federal government does have in place is a variety of individual regulations that address particular sectors and types of information. These privacy laws apply to a myriad of industries such as the credit, financial, and healthcare sectors. However, these limited regulations only require certain government entities to employ information security programs and provide a public notification in the event of a breach.[5] Meanwhile, the private sector has various federal laws that can apply, but it depends on the industry and the type of business involved. This "sectoral approach" to personal information security has consequently drawn various criticisms (Stevens, 2010, p.1). Some believe that these laws concentrate too much on how specific information is used (e.g. credit reports, medical data), rather than on defending the individual privacy of citizens. Others are suspicious of how different industries vary in their theory of what consumer information is explicitly considered 'sensitive.' Meanwhile, additional critics feel a national standard of information privacy is necessary in order to clarify the legal accountability of organizations and explain citizens' legal rights in the use of their personal information. While these inconsistencies on information privacy are evident at the federal level, many individual states have chosen to supplement these legal ambiguities by passing their own privacy laws.

As a result, the data breach notification laws enacted by most states have created tangible datasets that enable researchers to acquire a better picture of these incidents within the United States. At their core, these statutes typically require businesses to disclose any security incidents where consumer information may have been compromised. As of July 2010, forty-six states, the District of Columbia, Puerto Rico, and the Virgin Islands have all established breach notification laws (see Intersections Inc., 2010). Surprisingly, these regulations have only been established recently as California was the first state to ratify such legislation in July of 2003. Furthermore, many foreign nations still have not ratified

[5] A breach is defined as an event in which an individual's name plus Social Security Number (SSN), driver's license number, medical record, or a financial record/credit/debit card is potentially put at risk – either in electronic or paper format (Identity Theft Resource Center, 2010a, p.18).

similar laws requiring businesses to divulge this critical information to consumers. Thus, acquiring a sample that is generalizable to the broader, international scale of organizational data breaches is largely unattainable.

Current Issues within the Healthcare Industry

Compromised health records can be especially profitable to criminals who seek to not only exploit social security numbers for financial gain, but use health insurance policies to file fraudulent claims and write counterfeit prescriptions.[6] These types of scams are especially popular with Russian, Armenian, and Nigerian gangs in Los Angeles, California where the stolen credentials of doctors and patients are used to charge millions of dollars to Medicare (see Chernoff & Steffen, 2009). In order to address the legal ramifications that occur when an organization's health records are breached, Congress has only recently passed several laws designed to safeguard patient's private healthcare information. The Health Insurance Portability and Accountability Act (HIPAA) was the first such measure implemented in April of 2003 (see Fairwarning, Inc., 2010). The HIPAA's primary functions are to homogenize the exchange of electronic health information among health care providers by ensuring the continuance of health plan coverage, fraud enforcement, and provide privacy protection. By the end of December 2010, approximately 52,367 HIPAA privacy violations had been investigated and resolved by the Office of Civil Rights within the Department of Health and Human Services (DHHS) (see U.S. Department of Health and Human Services, 2010). However, the strongest privacy legislation was only recently enacted by Congress in 2009 in the form of the Health Information Technology for Economic and Clinical Health (HITECH) Act, which outlined a series of security provisions for healthcare-related breach notifications (see Kaufman, Rossin & Co., 2011). Specifically, the law requires health care entities to notify the Secretary of the DHHS of any security incidents affecting the records of 500 or more people and for a public notification of the breach to be posted on the DHHS website. The HITECH Act took effect on September 23rd 2009 and by September 23rd 2010 a total of 166 data breaches were reported that jeopardized 4.9 million health records.

Current Issues within the Education Sector

Any intentional use of a student's personal information is subject to the guidelines established by the federal government under the Family Educational Rights and Privacy Act (FERPA) ("Family Educational Rights and Privacy Act (FERPA)", 2011). Under this regulation, educational privacy rights are granted to the parents of every child until the age of eighteen when these rights are subsequently transferred to the adolescent. One of the fundamental rights permitted by this legislation allows students to review their own academic records as maintained by the attended school. If the student believes an inaccuracy exists within these documents the student has the right to request a correction be made to the record. In principal, schools must have written permission from a student to disclose any information from their records; however several exceptions exist. Records can be disclosed without consent to: school officials with a justifiable educational interest, schools to which a student is transferring, particular officials conducting audits, parties providing financial aid to a student, organizations conducting research on behalf of a

[6] ID fraud at hospitals goes largely uninvestigated as 34% of hospitals keep inadequate photo ID records and 70% of hospitals investigate less than one case per week (see Keckley, Coughlin, & Gupta, 2011).

school, accrediting organizations, a court of law with a judicial order or subpoena, designated officials during a health or safety related emergency, and to state authorities pursuant to individual state law.

Even with these rights, the education industry is still a vulnerable domain that contains a wealth of valuable information on millions of personal records related to students, faculty, and alumni. A review of the literature has revealed that this field is littered with ambiguities as to why a disproportionate amount of compromised records are reported from academic institutions. Siegel (2008) believes that the main reason why educational institutions appear to have higher totals than other entities is that universities have a greater sense of transparency within their communities. Even before states established mandatory reporting guidelines, colleges tended to err on the side of caution and notify their members of security issues rather than ignore the situation. This attitude is in stark contrast to other sectors, such as financial businesses, whose presumption of security can result in a loss of consumer confidence when a breach is publicly reported. Thus, it is believed that non-academic fields have a greater propensity to underreport breaches as they traditionally face more significant financial and legal ramifications; which would account for the discrepancy in the number of exposed records between educational entities and other fields. As a result, any significantly reduced totals within the academic sector will be difficult to determine in the future because other fields will continue to look more secure by comparison.

Annual Reports on Data Breaches

Few researchers have attempted to analyze data on organizational data breaches within the past few years. However, three encompassing studies have set the benchmark by which future analyses will undoubtedly be compared against. Widup (2010) compiled an international sample of organizational security incidents reported between 2005 and 2009 in twenty-eight different countries. The report compiled its sample of data breaches from four separate Internet databases: the Open Security Foundation, the Privacy Rights Clearinghouse, Sound Assurance, and the Identity Theft Resource Center. On an annual basis, Verizon Business's RISK team publishes a report on IT-based data breaches (see Baker et al., 2010). Verizon's current report, entitled the 2010 Data Breach Investigation Report (DBIR), is comprised of evidence from Verizon's own forensic investigations as well as several federal cases prosecuted by the United States Secret Service (USSS). Their sample from 2009 included a total of 141 data breaches, incorporating 57 incidents from Verizon and 84 breaches contributed by the USSS. Lastly, the Identity Theft Resource Center (ITRC) is another organization that publishes an annual report on the number and frequency of data breaches that occur within the United States (See Identity Theft Resource Center, 2010a). Their 2010 Data Breach Stats report is updated daily throughout the year and only includes credible sources into their sample, such as Attorney Generals' websites and reputable media outlets. The ITRC also published several supplementary reports based on the data from the 2010 Data Breach study that addressed the prevalence of hacking, insiders, and the accidental exposure of information (see Identity Theft Resource Center, 2010b; 2010c; 2010d). The statistical results of these studies are summarized below in Table 1:

Table 1. Summary of Findings from Annual Reports[7]

Annual Report:	*Widup (2010)*	*Baker et al. (2010)*	*ITRC (2010)*
Number of Data Breaches	2,807	141	662
Number of Total Records Breached	721.9 million	143 million	16.1 million
Year(s) Included in Sample	2005-2009	2009	2010
Computer Hacking	456 incidents / 327 million records breached	40% of incidents / 94% of the total records breached	113 incidents / 3.6 million records breached
Insiders	29% of incidents / 205.9 million records breached	48% of incidents / 3% of the total records breached	102 incidents / 3.4 million records breached
Unintended Disclosure			71 incidents / 1.1 million records breached
Misuse of Network Privileges		48% of incidents / 3% of the total records breached	
Lost/Stolen Laptop	589 incidents / 42 million records breached		
Business Entities	1,378 incidents / 507.2 million records breached	(% of breaches) Hospitality 23% Retail 15% Manufacturing 6% Tech Services 5% Business Services 4%	279 incidents / 6.6 million records breached
Educational Entities	549 incidents / 10.4 million records breached		65 incidents / 1.6 million records breached
Government Entities	539 incidents / 191.4 million records breached	4% of incidents	(Government/Military) 104 incidents / 1.2 million records breached
Healthcare Entities	341 incidents / 12.8 million records breached	3% of incidents	160 incidents / 1.8 million records breached
Banking/Credit / Financial Entities		(Financial Services) 33% of incidents	54 incidents / 4.8 million records breached

[7] Black areas indicate that the report did not have information on the topic listed in the left column.

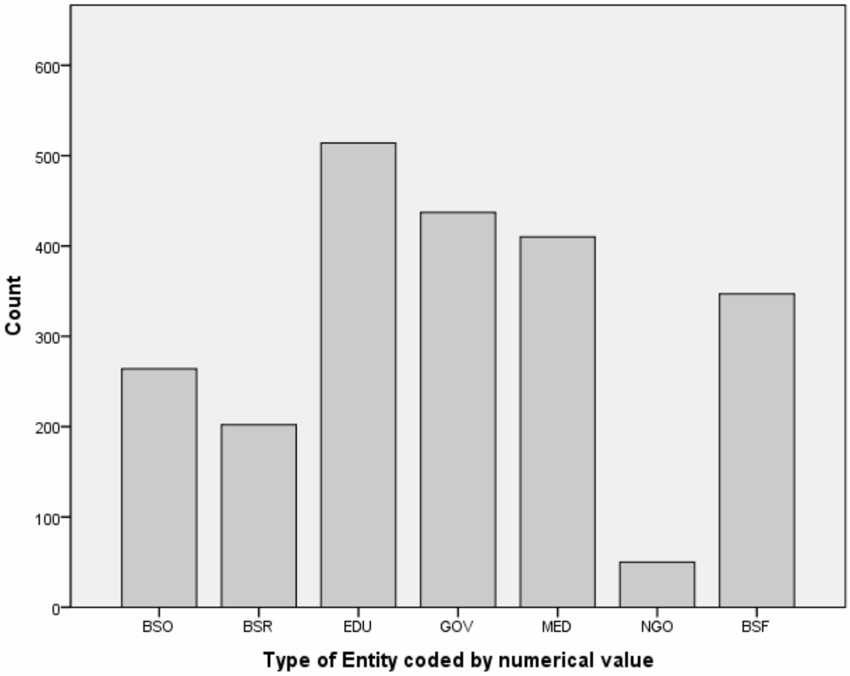
The Sample

Statistical analyses were performed on 2,219 documented security breaches in the United States compiled over a six year period (2005-2010) by the Privacy Rights Clearinghouse (PRC) in their ongoing efforts to maintain the Chronology of Data Breaches section of their website (see Givens, 2011). The PRC obtained its database from media reports, security bulletins, the Open Security Foundation's list-serve (datalossdb.org), databreaches.net, phiprivacy.net, and the National Association for Information Destruction (NAID). Our final sample in SPSS included over twenty unique variables that detailed various characteristics of each organizational data breach. However, for this article we have focused on the four most pertinent variables: the year the data breach was reported, the kind of entity that announced the incident, the type of data breach that occurred, and the region of the United States where the breach took place. The 'year' variable in our sample was assigned a chronological number to each of the six included years, so 2005 was coded as 1 and subsequently 2010 was coded as 6. The more sophisticated classifications for the 'entity,' 'type 'and 'region' variables are outlined below in Table 2:

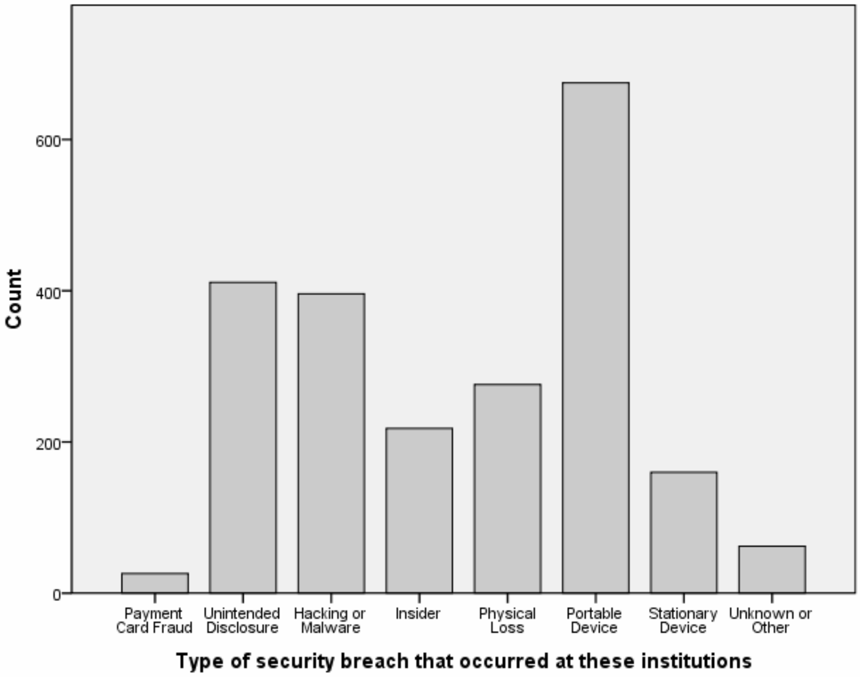
Table 2. *Coded values for experimental variables*

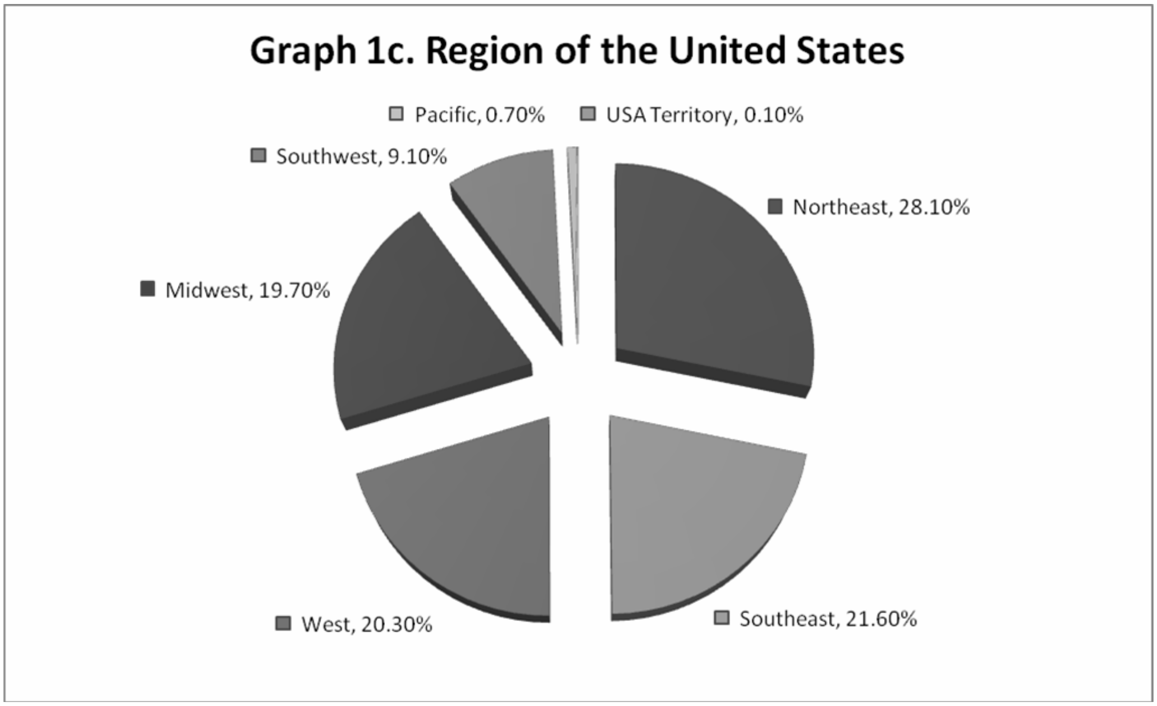
Corporate Entity	CODE: Corporate Entity	Type of Breach	CODE: Type of Breach	Region of the United States	CODE: Region of the United States
Business – Financial and Insurance Services	BSF	**Payment Card Fraud**	CARD	**Mid-West**	MW
Business – Retail/Merchant	BSR	**Unintended Disclosure**	DISC	**North-East**	NE
Business - Other	BSO	**Hacking or Malware**	HACK	**South-East**	SE
Educational Institutions	EDU	**Insider Abuse**	INSD	**South-West**	SW
Government and Military Agencies	GOV	**Physical Loss**	PHYS	**West**	W
Medical Healthcare Providers	MED	**Theft/Loss of a Portable Device**	PORT	**Pacific**	PAC
Non-Governmental Organizations	NGO	**Theft/Loss of a Stationary Device**	STAT	**Federal Territory**	FED
		Unknown/Other	UNKN		

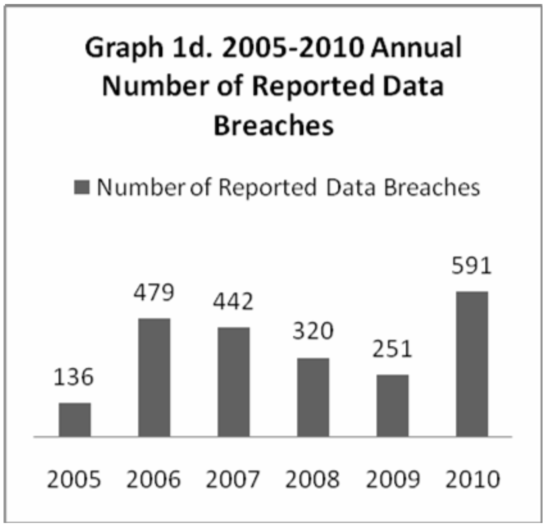
Graph 1a.2005 - 2010 Type of Entity



Graph 1b. 2005 - 2010 Type of Breach







Graphs 1a, 1b, 1c, and 1d partition our total six-year sample into one of the four aforementioned variables. All entities combined disclosed 735.4 million records; of which 510.9 million records contained customer's critical financial information. However, in this article our analysis will only be discussing these four variables as they pertain to the healthcare (410 total data breaches) and education (513 total data breaches) entities that each reported a substantial number of cases in our sample.

Methodology

Bayesian Statistics

In order to draw accurate conclusions from this sample, inferential statistics were necessary to address how the dataset was indicative of the larger population of corporate data breaches within the United States. The ultimate goal of comparing these four particular variables was to estimate the likelihood of certain entities reporting a specific feature of a data breach. For example, given data breaches at educational institutions, what were the odds that personal information was unintentionally disclosed? As noted in Table 2, the number of classifications for each of these four variables ranged from six to eight possible conditions. To produce these probabilities, unique statistical methods were needed that could simultaneously compare conditional research questions and account for the many conditions of each variable. As a result, a more sophisticated form of classical, conditional probabilities was used in our probabilistic model known as 'Naïve Bayes classifiers.'

Naïve Bayes classifiers operate by presuming that the occurrence or omission of a specific attribute within a group is not related to the occurrence or omission of any other attribute in the data (see Shmueli, Patel, & Bruce, 2007). In other words, all the features of a particular condition independently support the likelihood of a certain scenario. For example, the characteristics of a typical car are four wheels, a steering wheel, an engine, and transmission. Regardless if these features depend upon each other to function within a car, Naïve Bayes classifiers treat each of these characteristics as individually contributing to the odds that a particular object with these features is a car. By assuming independence among these characteristics, the probability that these conditions will occur together is equal to the multiplication of all relevant variables. Consequently, when Naïve Bayes classifiers determine the odds of a particular situation occurring between two or more variables importance is placed on the likelihood that a specific order of variables (*A* given *B*) will occur in comparison to the inverse arrangement (*B* given *A*). For example, if we were comparing the 'number of times a horse won a race' to 'whether or not it rained during the race' we would have four possible scenarios: the horse won when it was raining, the horse lost when it was raining, the horse won when it did not rain, and the horse lost when it did not rain. Therefore, in order to address a specific probability like 'what are the odds a horse will win when it is raining,' it is vital to also know how many times the horse lost and which races occurred without rain.

In order to account for such a particular scenario amid several probabilities, Naïve Bayes classifiers derive their equation from Bayes' Theorem expressed as:

$$p(A|B) = \frac{p(B|A)\ p(A)}{p(B)}$$

P(A|B) is the odds of finding observation A given the presence of data from B or using the question from the previous example 'the odds (A) a horse will win a race (B) when it is raining.' P(B|A) is the specific numerical intersection where 'the number of wins' and 'rain during a race' come together among the four previously mentioned scenarios. P(A) is the probability of A occurring without accounting for B, or simply 'the odds of the horse winning the race' regardless of weather conditions. P(B) is the probability of B occurring without accounting for A, or 'the odds it was raining during the race' without considering which horse won. Thus, Bayesian statistics allow us to account for several probabilities when determining the odds of a specific set of conditions.

Bayesian Excel Consoles

A more intricate version of Bayes' formula was implemented into our Bayesian estimator as the variables in this study had up to eight different classifications. Using Microsoft Excel, the Bayesian estimator we created allowed us to cut and paste the raw data from our experiment into a worksheet that automatically calculated the probability of a particular set of circumstances that could be adjusted on the fly. Other statistical software, such as SPSS, could have been used for this process; however the design of SPSS's interface lacked the flexibility to enter and calculate a singular set of conditional settings in a timely manner. Performing these calculations in SPSS would have required navigating through a series of cumbersome menus and a timely delay as the software slowly processed the request. Thus, one of the benefits of having used Microsoft Excel allowed for calculations to occur in an instant as soon as a single cell value was modified to match the specific condition the user wanted to explore.

Results

In a series of tables, this section lists the probabilities generated from our distinctive use of Bayesian statistics. Again, the probabilities chosen for inclusion in this paper focused on the influence of four variables as they pertain to the literature on the healthcare and education entities. As such, the (A) or given value was the 'entity' variable, while (B) was either 'year,' 'type of breach,' or 'region.' Translated into a form of Bayes' expression 'p(A|B)' the probabilities reported in these results were from one of the three following comparisons: p(entity|year), p(entity|type of breach), and p(entity|region).

Table 3. Probability of a data breach within a specific year, given a health care entity.

Year	2005	2006	2007	2008	2009	2010
Probability	0.081481481	0.121593291	0.136054422	0.163009404	0.18875502	0.308474576

Table 4. Probability of a data breach within a specific region, given a health care entity.

Region	Mid-West	North-East	South-East	South-West	West	Pacific	Federal
Probability	0.180778032	0.173354735	0.1875	0.201970443	0.188470067	0.266666667	1.0

Table 5. Probability of a specific type of a data breach, given a health care entity.

Type	Unintended Disclosure	Hacking/Malware	Payment Card Fraud	Insider Abuse	Physical Loss	Theft/Loss of a Portable Device	Theft/Loss of a Stationary Device	Unknown
Probability	0.117073171	0.04859335	0	0.266055046	0.254545455	0.25297619	0.251572327	0.083333333

Table 6. Probability of a data breach within a specific year, given an educational entity.

Year	2005	2006	2007	2008	2009	2010
Probability	0.474074074	0.213836478	0.238095238	0.304075235	0.285140562	0.125423729

Table 7. Probability of a data breach within a specific region, given an educational entity.

Region	Mid-West	North-East	South-East	South-West	West	Pacific	Federal
Probability	0.290617849	0.179775281	0.25	0.221674877	0.228381375	0.40	0

Table 8. Probability of a specific type of a data breach, given an educational entity.

Type	Unintended Disclosure	Hacking/Malware	Payment Card Fraud	Insider Abuse	Physical Loss	Theft/Loss of a Portable Device	Theft/Loss of a Stationary Device	Unknown
Probability	0.356097561	0.398976982	0	0.050458716	0.130909091	0.157738095	0.301886792	0.166666667

Observations from the Healthcare and Education Sectors

Within the healthcare sector, the passage of the HITECH Act improved the frequency of cases that were reported by the medical industry. When the HITECH Act took effect in September of 2009 the probability of a data breach being reported by a healthcare entity grew to 18.9%. By the end of 2010, this same probability grew to a considerable 30.8%. The health care probabilities by region illustrated a fairly even distribution across the continental United States. Overall, the regional odds varied by less than 3%; with a range between 17.3% and 20.2%. The four types of breaches that were overwhelmingly responsible for security incidents at medical entities were: insider abuse, physical loss of records, compromised portable devices, and compromised stationary devices. The probabilities for each of these four methods deviated by less than 1.5%; with a range between 25.2% and 26.6%. Based on this observation, it would appear that the healthcare industry was highly vulnerable to these breach methods and that it is essential for this sector to implement appropriate security precautions. Possible suggestions to combat these issues include improving employee education programs that deal with information security, employing SCP to proactively deter insider abuse, and enhancing the technological security of sensitive equipment via encryption or the implementation of virtual private networks (VPNs).[8]

As previously mentioned in the literature review, educational entities have had a long history of open disclosure regarding data breaches. This attitude towards transparency appears to be one of the major reasons why educational entities had the highest probabilities of annual reporting. For example, in 2005 these education-related breaches had a probability of 47.4%; or in other words that roughly 1 in 2 security incidents in 2005 were reported by an educational entity. This colossal prospect declined between 2006 and 2009 as the odds of a data breach ranged from 21.3% to 30.4%. Unexpectedly, in 2010 this annual probability hit an all time low of 12.5%. This sudden decline could be

[8] Virtual Private Networks are used as a secure means of allowing authorized users to access the private network of a company from an off-site location via the Internet.

a result of the increased number of health care breaches that were now being disclosed due to the HITECH Act. Consequently, the escalation in health care reporting may have become responsible for a larger portion of data breaches in 2009 and 2010. However, it was also possible that this low probability was the result of improved security measures implemented by universities over the past six years. The dissection of these probabilities by region shows a wider range of values between 17.9% and 40%. Educational institutions in the North-East region appear to have had the best information security policies as this area had the lowest odds of announcing a data breach. Meanwhile, academic institutions in the Mid-West were the most vulnerable to potential data breaches within the continental United States. The most probable types of breaches that occurred at educational entities were hacking or malware attacks, the unintended disclosure of information, and the theft or loss of a stationary device. As a result, it would appear that generous resources should be made available to IT departments at educational entities as the internal computer network of these institutions is highly susceptible to attack. The 'human factor' involved in unintended disclosures and compromised stationary devices can be effectively controlled with SCP measures such as recurring information security classes for employees and issuing physical locks for all computer equipment to deter theft.

Conclusions

Policy Implications

Widup (2010) believes that the best way to deal with organizational data breaches would be with the creation of a federal agency that is responsible for notifying, documenting, and evaluating corporate data breaches. Such a bureau would simplify the notification process for American businesses by creating better metrics for more thorough investigations and offer data security professionals a central reference for the latest network exploits. Moreover, the standardization of reporting guidelines for data breaches would allow researchers to have the same details in each report. Similarly, the researchers at ITRC found that there was no consistent transparency in breach reporting apart from a few exceptional media outlets and progressive state websites (see Identity Theft Resource Center, 2011). The current approach of "risk of harm" in breach notification legislation is not an acceptable method for determining public notification because it still allows the company to make the decision on whether or not to report a breach (Identity Theft Resource Center, 2011, p.1). This argument was quantified in the statistics of the 2010 ITRC Breach List as 49% of these incidents failed to include the number of potential records placed at risk and 38.5% of these breaches did not identify the manner in which the information was compromised. For the sake of public transparency, the author felt that the power to make reporting decisions should be made by a federal IT forensic specialist who is called in to investigate when a data breach occurs.

Currently, a rough prototype of a centralized reporting bureau exists within the federal system that allows individual consumers to report incidents of cybercrime. The FBI established the Internet Crime Complaint Center (IC3) in 2000, allowing citizens to notify the agency of computer-based crimes such as online fraud, hacking, identity theft, money laundering, child pornography, and economic espionage (see U.S. Department of Justice, 2010). Annually, this office publishes a report that details such information as the number, type, and cost of documented cybercrimes; allowing the government to focus their resources on preventing the most problematic offenses. Designing a similar system for organizational data breaches would not only reduce the financial loss incurred by these companies, but also aid in national security by allocating government resources to only the most vulnerable industries. As the federal government has been willing to use their assets to aid victims of data breaches in the healthcare industry with the passage of the HITECH Act; the creation of a dedicated federal agency to deal with national data breaches across all industries would clearly be the next logical step.

Future Research

This study is open to numerous applications for future research as only a specific portion of the sample was analyzed in this paper. In this study, attention was drawn to only the four most pertinent variables associated with organizational security incidents. Altogether, twenty-four variables were determined for each of the 2,219 data breaches included in the dataset that presents future researchers with an abundance of characteristics to rearrange and compare to one another. While not directly addressed in this paper, the sample compiled in this study provides potential researchers with detailed information on the monthly totals of data breaches from January 2005 to December 2010. Other aspects of the data can be explored by studying new combinations of variables, such as a comparison of state figures by season, month, or year. Additionally, the Bayesian statistical methods applied in this paper were independently concerned with comparing only two variables at a time. This means that future researchers can expand upon our Bayesian model by using three or more variables in their own analyses. Also, interviewing the IT specialists that handled the computer-based data breaches in our sample would further add to this field of research. In conclusion, this dynamic research can be used by future scientists who will either analyze this subject from the perspective of a specific industry or attempt an all-encompassing viewpoint of national data breaches. Ultimately, it is our hope that one day a centralized database will exist to simplify the data acquisition process for forthcoming researchers interested in organizational data breaches. Ideally this event will coincide with an end to the industry-specific reporting discrepancies that continue to mask the underlying prevalence of corporate data breaches.

Limitations

Several observations need to be considered when reviewing the probabilities in the results section. Firstly, the Pacific region and the federal territories had very few data breaches reported within the six year time period covered in our sample. These two regions were only responsible for a total of seventeen data breaches. Fifteen of these cases originated in the Pacific region, while the remaining two incidents were reported by healthcare entities within the federal territories. Consequently, the small sample size available for these regions generated larger probabilities when compared against other variables. Secondly, payment card fraud was only reported in twenty-six data breaches and all of these incidents were exclusive to the three business entities categorized in the sample.[9] As such, our probabilities for payment card fraud were always zero considering that the results focused on healthcare and educational institutions. Furthermore, the entirety of cases obtained from 2005 only accounted for 136 data breaches; compared to subsequent years that reported between 251 and 591 incidents. This moderately small sample size for 2005 was a consequence of the few resources that tracked data breaches at that point in time. Comparatively, the 591 data breaches reported in 2010 were partially

[9] These business entities include Business – Financial and Insurance, Business – Retail, and Business – Other.

due to the inclusion of three additional reporting databases in the Privacy Rights Clearinghouse sample. Thus, as the years have passed it seems that national interest in organizational data breaches has grown, suggesting that more security incidents are now being reported than ever before.

The most notable limitation of this study was the absence of a centralized reporting database for organizational data breaches. This detail prevented the author from conducting a definitive multivariate analysis in order to determine the significance of our results. Instead, we conducted a non-parametric assessment of our data by simulating specific conditions among our variables in order to reach our conclusions.[10] Consequently, the methods used in this study can be regarded as a form of exploratory research. This impediment is regrettably familiar in criminal justice research because most crime data are characteristically deficient as it only accounts for victims who make the effort to report a particular crime (see Coleman & Moynihan, 1996). Statistics associated with underreported crimes such as data breaches, robberies, and rapes are only considered to be a fraction of their actual crime rates. Conversely, homicides tend to have more accurate statistics because there is usually a corpse for law enforcement agencies to process. Other obstacles faced by criminal researchers are voluntary crime reports that tend to leave out specific information. For example, the extensively referenced Uniform Crime Report (UCR) follows a "hierarchy rule" only counting the most serious offense committed during a criminal incident (U.S. Department of Justice, 2004, p.10). Thus, if a murder took place during the course of an armed robbery and carjacking, the report would only classify the episode as a murder.

References

Akers, R. L., & Sellers, C. S. (2009). *Criminological Theories: Introduction, Evaluation, and Application*. New York: Oxford University Press.

Baker, W., Goudie, M., Hutton, A., Hylender, C.D., & Niemantsverdriet, J. (2010). *2010 data breach investigations report* Verizon Business, RISK Team. Retrieved from http://www.verizonbusiness.com/resources/reports/rp_2010-data-breach-report_en_xg.pdf

Berger, J. O. (1985). *Statistical decision theory and bayesian analysis*. (Second ed.). Springer-Verlag.

Bernardo, M., & Smith, A. F. M. (1994). *Bayesian theory*. John Wiley & Sons Inc.

Bishop, C. M. (2007). *Pattern recognition and machine learning*. Springer-Verlag.

Chernoff, A., & Steffen, S. (2009, October 24). *Organized crime's new target: medicare*. Retrieved from http://www.cnn.com/2009/CRIME/10/22/medicare.organized.crime/

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: a routine activity approach. *American Sociological Association*, *44*(4), 588-608.

Coleman, C., & Moynihan, J. (1996). *Understanding crime data: haunted by the dark figure*. United Kingdom: Open University Press Crime and Justice Series.

Davis, M. S. (2002). *The Concise Dictionary of Crime and Justice*. Thousand Oaks, CA: Sage Publications.

[10] Bayesian probability, is a non-parametric technique initially discovered by Thomas Bayes and pioneered by Pierre-Simon Laplace in the 1800's and is a fundamental model use in Knowledge Discovery (aka: Machine Learning or Data Mining) for assessing the relative probabilities of an event (see Berger, 1985; Bernardo & Smith, 1994; Bishop, 2007; Shmueli, Patel & Bruce, 2007).

Fairwarning Inc. (2010). *Privacy breach benchmarks compel care providers to deploy breach monitoring and commit to a culture of privacy and compliance*. Retrieved from http://www.fairwarningaudit.com/documents/2010-FAIRWARNING-FINDINGS-REPORT.pdf

Family educational rights and privacy act (ferpa). (2011, April 8). Retrieved from http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

Felson, M., & Clarke, R. (1998). Opportunity makes the thief: Practical theory for crime prevention: Great Britain.

Givens, B. (2011, January 5). *Chronology of data breaches - privacy rights clearinghouse*. Retrieved from http://www.privacyrights.org/data-breach

Identity Theft Resource Center. U.S. Department of Justice, Office for Victims of Crime, Office of Justice Programs. (2010a). *Identity theft resource center: 2010 data breach stats* (Grant No. 2007-VF-GX-K038). Retrieved from http://www.idtheftcenter.org/artman2/uploads/1/ITRC_Breach_Stats_Report_2010 1229.pdf

Identity Theft Resource Center. U.S. Department of Justice, Office for Victims of Crime, Office of Justice Programs. (2010b). *Identity theft resource center: 2010 data breach hacking category summary* (Grant No. 2007-VF-GX-K038). Retrieved from http://www.idtheftcenter.org/artman2/uploads/1/ITRC_Breach_Stats_-_Hacking_Summary_20101229.pdf

Identity Theft Resource Center. U.S. Department of Justice, Office for Victims of Crime, Office of Justice Programs. (2010c). *Identity theft resource center: 2010 data breach insider theft category summary* (Grant No. 2007-VF-GX-K038). Retrieved from http://www.idtheftcenter.org/artman2/uploads/1/ITRC_Breach_Stats_-_Insider_Theft_Summary_20101229.pdf

Identity Theft Resource Center. U.S. Department of Justice, Office for Victims of Crime, Office of Justice Programs. (2010d). *Identity theft resource center: 2010 data breach accidental exposure category summary* (Grant No. 2007-VF-GX-K038). Retrieved from http://www.idtheftcenter.org/artman2/uploads/1/ITRC_Breach_Stats_-_Accidental_Exposure_Summary_20101229.pdf

Identity Theft Resource Center. (2011). *Data Breaches in 2010: Indicates Mandatory Reporting Needed*. Retrieved from http://www.idtheftcenter.org/artman2/publish/lib_survey/Breaches_2010.shtml

Intersections Inc., BreachReadiness. (2010). *Intersections data breach consumer notification guide* Intersections.com. Retrieved from http://www.intersections.com/library/Intersections%20Data%20Breach%20Services%20Consumer%20Notification%20Guide%20July%202010.pdf

Kaufman, Rossin & Co. (2011). Preventing a data breach and protecting health records one year later: are you vulnerable to a breach? *Kaufman, Rossin & Co. White Paper Series*, Retrieved from http://www.kaufmanrossin.com/whitepapers/

Keckley, P.H., Coughlin, S., & Gupta, S. (2011). *Issue brief: privacy and security in health care: a fresh look*. Deloitte Center for Health Solutions. Retrieved from http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/Health%20Reform%20Issues%20Briefs/US_CHS_PrivacyandSecurityinHealthCare_022111.pdf

Shmueli, G., Patel, N. R., & Bruce, P. C. (2007). *Data mining for business intelligence: concepts, techniques, and applications in microsoft office excel with xlminer.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Siegel, P. M. (2008). Data breaches in higher education: from concern to action. *EDUCAUSE Review, 43*(1), 72-73. Retrieved from http://net.educause.edu/ir/library/pdf/ERM08111.pdf

Siponen, M. & Willison, R. (2007). A critical assessment of is security research between 1990-2004. *Proceedings of the 15th European conference of information systems* (pp. 1-17). St. Gallen, Switzerland: Copenhagen Business School.

Stevens, G. Congressional Research Service, (2010). *Federal information security and data breach notification laws* (RL34120). Washington, DC: Congressional Research Service. Retrieved from http://www.fas.org/sgp/crs/secrecy/RL34120.pdf

U.S. Department of Health and Human Services, Office of Civil Rights. (2010). *Health Information Privacy: Enforcement Results by Year.* Rockville, MD. Retrieved from http://www.hhs.gov/ocr/privacy/hipaa/enforcement/data/historicalnumbers.html

U.S. Department of Justice, Federal Bureau of Investigation. (2004). *Ucr: uniform crime reporting handbook* [Rev. 2004]. (Adobe PDF file), Retrieved from http://www.fbi.gov/about-us/cjis/ucr/additional-ucr-publications/ucr_handbook.pdf

U.S. Department of Justice, Federal Bureau of Investigation. (2010). *2009 Internet crime report* (Grant No. 2009-BE-BX-K042). Washington, DC: National White Collar Crime Center. Retrieved from http://www.ic3.gov/media/annualreport/2009_IC3Report.pdf

Widup, S. (2010). *The leaking vault: five years of data breaches* Digital Forensics Association. Retrieved from http://www.digitalforensicsassociation.org/storage/The_Leaking_Vault-Five_Years_of_Data_Breaches.pdf

Willison, R. (2008). Applying situational crime prevention to the information systems security context. *Crime Prevention Studies*, *23*, 169-192.

Willison, R., & Siponen, M. (2009). Overcoming the insider: reducing employee computer crime through Situational Crime Prevention. *Communications of the ACM*, *52*(9).