Information and Knowledge Management
ISSN 2224-5758 (Paper) ISSN 2224-896X (Online)
Vol 2, No.8, 2012

Statistical Insight into Breach Data toward Improved Countermeasures

Adewale O Adebayo* Olawale J Omotosho Yinka A Adekunle
School of Science and Technology, Babcock University, P.M.B.21244 Ikeja, Lagos, Nigeria
*E-mail of corresponding author: adebayoa@babcock.edu.ng

Abstract
Information is essential to the continued growth of any society. A secure information infrastructure is required. Despite creditable efforts, there are visible failures of Information Security (IS). Breach data, by its nature, offers factual data about what is happening that should reveal what more is required for improved countermeasures. This work, therefore, provides statistical insight into breach data through analysis of the data set of disclosed breaches for the period January 1st, 2011 to May 18th 2011 kept by a leading and open repository. The analysis informed that mitigation efforts should include improved perimeter security, document archiving and disposal procedure, and physical security for foreseeable future.
Keywords: Security countermeasure, breach data, storage security, information security, security breach

1. Introduction

Information is essential to the continued growth of any society. Information is created, managed, processed, and archived by an information system. The confidentiality, integrity, availability, and accountability issues of information must be ensured for appropriate usage (Somasundaram and Shrivastava, 2009; Stamp, 2006). A secure information infrastructure is required. Information Security (IS) is ensuring that the information systems perform according to stipulation and retain optimum performance in the face of clever wrongdoers or oppositions.

Information is vulnerable to technical, physical and human threats, and risks to IS must be continually assessed and effectively addressed (Richo Security Solutions, 2010). Threats to information security include errors and omissions, fraud and theft, malicious hackers, malicious code, denial-of-service attacks, and social engineering (Shostack and Stewart, 2009). There exist security measures to combat IS modelled threats, but how effective and efficient are they in reality, and what more is needed?

It is known that enterprise-wide security programs, establishing security policies (Peltier, Peltier, and Blackley, 2005), and designing and configuring information systems for reliable and successful operation from the start (Martin and Weadock, 1997), are security measures believed to combat threats to IS. Against these: Neutralisation theory provides a compelling explanation for information system security policy violations and offers new insight into how employees rationalize their behaviour (Siponen and Vance, 2010); employee's intention to comply with information security policy is significantly influenced by attitude, normative beliefs, and self efficacy to comply (Bulgurcu, Cavusoglu and Benbasat, 2010); and designing and configuring information systems for reliable and successful operation from the start introduces complexities that increase the risk of bugs. Use of encryption provides storage confidentiality but introduces system performance issues and denial-of-service vulnerabilities. There are a number of fundamental issues militating against successful IS measures.

Despite credible efforts, some generally visible failures of IS are spam and associated problems (Helbush, 2009), malicious codes (Eichin and Rochlis, 1989), bugs in software including operating systems (Keizer, 2010), and data breaches (Aitoro, 2007). Notable failings of IS (FBI, 2010; Helbush, 2009; SonicWALL, 2008; Klienman, 2007; FBI, 2007; Gartner, 2007), IS industry not currently organized for IS leadership (Gordon, Loeb and Sohail, 2010; Johnson and Warkentin, 2010; Baskerville and Myers, 2009), and inadequacies of existing evidence to support IS decision making (Shostack and Stewart, 2009; Hoffer and Straub, 1989) are cogent reasons for fresh perspective to the subject of IS.

Evidence-based practice implies the use of field or empirical research findings as evidence supporting effective development and use of information systems, thereby linking theory to practice (Oates, 2009), but doing any survey in IS in a scientifically defensible manner (designing a set of questions that do not betray bias and finding a suitable

set of respondents) is very challenging, and there is the impediment of imprecise vocabulary (Shostack and Stewart, 2009; Ryan and Jefferson, 2003). Further militating against collection of factual data is what psychologists call the "valance effect", which is people's tendency to overestimate the likelihood of good things happening rather than bad things (Rosenhan and Messick, 1966).

The use of breach data and other new sources of data that would eliminate or reduce some of the setbacks of survey in IS, and would provide new perspective to the subject of IS were proposed (Mahmood, et al., 2010; Shostack and Stewart, 2009). Breach data is generated as a result of reports of data breaches. Data breaches may involve personal health information (PHI), personally identifiable information (PII), trade secrets or intellectual property. The catalyst for reporting data breaches to the affected individuals has been the US California law that requires notice of security breaches implemented July 2003 ([Online] Available: http://www.ncsl.org/default.aspx?tabid=13489, 8/4/12). More than forty of US states have since passed laws requiring that individuals be notified of security breaches (Attrition, 2011; PrivacyRights, 2011). The breach data by its nature offers widely spread, unbiased, and easily accessible data for analysis to provide fresh insight into issues surrounding data breaches and therefore IS. Breach data is gathered and shared at PogoWasRight.org, Attrition.org, Privacy Rights Clearing House, and other sites (Shostack and Stewart, 2009, p187).

A number of works have been done analyzing breach data (Gordon, et al, 2010; Culnan and Williams, 2009; Hasan and Yurcik, 2006; Acquisti, Friedman and Telang, 2006; Tehan, 2005), but none yet examined the current collection of breach data towards improved countermeasures.

The goal of this work was, therefore, to gain insight into breach data toward improved countermeasures against storage security breaches through statistical analysis of certain current breaches data set. This should provide enlightening information regarding security incidents that would inform improved security measures.

1.1. Method of the Research

The data requirement for this work includes details of who did the breach, methods used, what assets were involved, what were the asset attributes affected, and so on, sufficient for meaning report. The data generation method was found documents; publicized reports of breach incidents between January 1st, 2011 and May 18th 2011, which avoids the biases indicated against designing questionnaires and finding appropriate respondents (Ryan and Jefferson, 2003) were sought. Sampling frame of all storage security breaches between January 1st, 2011 and May 18th 2011, from which a sample is to be chosen was not possible because there is no way of knowing all storage security breaches during the period of interest; not all the breaches involved PII, not all that involved PII are reported, not all that involved PII that are reported may have been captured, and not all the breaches have been discovered. Sampling technique; this piece of work took the data set of a leading repository of storage security breaches data, Open Security Foundation, Datalossdb.org (Adebayo, 2012), of reported breach incidents for the period between January 1st, 2011 and May 18th 2011 as a cluster representing the population of interest, in order to save the cost and time of searching for the individual breach incident and its details. Datalossdb.org offers certain links to original publications, a random selection of which was verified. The details datalossdb.org provides for each breach include: date of incidence, country of incidence, organization name, type of organization, type of breach, number of records lost, and whether insider or outsider breach. Number of breaches during this period, in datalossdb.org data set, was two hundred and ten, but one incidence involving the loss of seventy-seven million records was ignored as an outlier. This data set though could not be said to be statistically representative of total breaches that occurred during the period, offers useful insight into storage security breaches for improved countermeasures. The data obtained was analyzed, using SPSS 15.0 for Windows, along the following dimensions: number of records per incidence summary statistics, reported and percentage of reported storage breach incidents and storage records lost by organization type, number of breaches and records lost by breach mechanism, number of breaches and records lost by insider or outsider, number of breaches and records lost in time, number of records lost per incidence by organization type, number of records lost per incidence by breach mechanism, number of law suits on storage security breaches, and fraction of perpetrators arrested or prosecuted, in order to communicate what could be learnt about storage breach events toward improved countermeasures.

2. Outcomes

The succeeding sections are 2.1 - Data Presentation, 2.2 – Discussion, and 2.3 – Other Related Works.

2.1 Data Presentation

Table 1 presents summary statistics of number of records per breach incident. Figure 1 depicts number of breaches according to type of organization. Figure 2 depicts reported storage records lost by organization type. Table 2 and Figure 3 present number of breaches by detailed breach mechanism. Figure 4 presents number of records lost by breach mechanism. Figure 5 depicts number of breaches by insider or outsider. Figure 6 shows the number of records lost by insider or outsider. Figures 7 and 8 show scatter diagrams of the number of breaches and number of records lost in time, respectively. Figure 9 depicts fraction of law suits on storage security breaches. Figure 10 depicts fraction of perpetrators arrested or prosecuted.

2.2 Discussion

The mean number of records per breach incident is 170,563 and the highest record loss was 24,600,000 (Table 1). Incidences with no record lost were included for other knowledge that may be gained.

Businesses (Biz) suffer more breaches, followed by medical (Med), education (Edu), and government (Gov) institutions, respectively (Figure 1). Malefactors do breaches for nefarious gains about which businesses and medical institutions seem greener. It was revealed that businesses lost more records than all other types of organizations put together (Figure 2). Number of records lost is one of the main indicators of breach incident severity.

Hacking topped as a breach mechanism (Table 2 and Figure 3) though the sum of physical losses is worse. Hacking topped by far as a breach mechanism in terms of number of records lost. The 24,600,000 records were lost through hacking (Table 1 and Figure 4). This calls for improved perimeter security, and Intrusion Detection and Prevention Systems. More attention should also be given document archiving and disposal, and physical security.

Outsiders perpetrated more breaches than insider malicious (Figure 5). Outsider breaches result in the largest number of lost records (Figure 6). More care should definitely be paid securing network perimeter, and attention should be given stemming insider accidental losses that ranked second, probably through training and policies.

There is no time target for breaches (Figure 7). The number of records loss in time is low (<4,500) (Figure 8). Malefactors seem to avoid the usually successful hunt that comes with hitting big volume records probably because of the policing heat that they generate.

The number of law suits on storage security breaches (<2%) is minimal (Figure 9). Organizations need not fear reporting storage security breaches.

The fraction of perpetrators arrested or prosecuted stood at 18.6 percent (Figure 10). This will likely discourage malefactors.

It is also appropriate to mention the need for improved framework for breach data capture. More specificity in capturing the details about the breach incidents should produce more revelations towards improved IS decisions.

2.3 Other Related Works

A summary of selected storage security incidents reported in the press between 2000 and 2005 was conducted (Tehan, 2005). In this a small data set of incidents was used and biased sampling was noted.

A report of the United States' (US's) state of California Department of Consumer affairs/Office of Privacy Protection

in April, 2006, claims that physical attack is the most prevalent form of storage breach in California. It is observed from this data that physical is still the breach mechanism that is most prevalent, though hacking is more severe.

A study on the impart of security breaches on stock market valuations claims that publicly reporting security breaches does not negatively affect, in the mid-term, the valuation of stock of the reporting organization (Acquisti, et al; 2006). It is noted, though, that the study presented a skewed and partial view because it was limited to incidents affecting only publicly traded firms, and it includes different types of security breaches not limited to storage breaches. It was however shown that voluntary disclosure of information security breaches is associated positively with the market value of a firm (Gordon, et al, 2010). This piece of work shows law suits on storage security breaches is negligible, and therefore offers no contradiction.

A study that claimed to be the first valid statistical analysis of disclosed storage security breaches used combined data set spanning January 1, 2005 to June 5, 2006, from two leading sources. It shows that 35% of breaches occur in educational institutions, followed by Business with 25%; 36% of total number of records lost was from Business against 3% Educational institutions; 41% of breaches occur via external intrusion or hacking with 36% through physical attack (Hasan and Yurcik, 2006). The current analysis shows businesses taking the lead in terms of number of storage security breach incidents and number of records lost, but shows external intrusion as still the leading breach mechanism in severity.

2011 Data Breach Investigation Report by Verizon Risk Team, US secret service and Dutch High Tech Crime Unit (April, 2010) shows that 92% of data breaches stemmed from external agents, 50% of breaches are through some form of hacking, 83% of victims were targets of opportunity, and 96% of breaches were avoidable through simple or intermediate controls (Verizon, 2011). It is noted that the data set used include only Verizon confirmed incidents of data compromise involving deliberate breach and compromise situations, and the data set is not also open to external scrutiny. Besides, it is unlikely that Verizon will be called to investigate the loss of portable devices. The current analysis shows a 50:50 chance of data breaches stemming from external or internal agents, and shows that about 46% of breaches are still through some form of hacking.

The use of proactive system engineering in designing protection for storage systems, based on classical security principles or data lifecycle model, by organising system threats and vulnerabilities into general classes to be addressed with known storage protection techniques (Hasan, et al, 2005) would be enlightened by breach data analysis.

3. Conclusion

This work presents further empirical evidence towards assessing risk of storage security breaches. It informs and it is clearer, for foreseeable future, that mitigation efforts should include improved perimeter security, document archiving and disposal procedure, and physical security.

It also compels, being enlightening itself, that periodic work of this type should be done on improved data set to inform improved IS decision making over time. The applications of threat modelling should also be complimented with what breach data analysis informs.

References

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

Adebayo, A. O. (2012). A Foundation for Breach Data Analysis. Journal of Information Engineering and Applications, Vol.2 No.4, pp 17-23.

Aitoro, J. (2007). Reports of federal security breaches double in four months. Government Executive.com, October 23, 2007, www.govexec.com/dailyfed/1007/102307;1.htm. Retrieved November 11, 2010

Attrition. (2011). Entities that suffer large personal data incidents (list). http://attrition.org/errata/dataloss

Baskerville, R L and Myers, M D. (2009). Fashion Waves in Information Systems Research and Practice. MIS Quarterly December 2009, Vol. 33, Issue 4 (pp. 647-662)

Bulgurcu, B, Cavusoglu, H, and Benbasat, I (2010) "Information Security Policy Compliance: An Empirical Study of Rationality-based Beliefs and Information Security Awareness." MIS quarterly Vol. 34, No. 3

Culnan, M J, and Williams, C C. (2008). How Ethics Can Enhance Organizational Privacy: Lessons from the ChoicePoint and TJX Data Breaches. MIS Quarterly December 2009, Vol. 33, Issue 4 (pp. 673-687)

Eichin, M and Rochlis, J. (1989). With Microscope and Tweezers: An analysis of the Internet virus of November 1998. 1989 IEEE Symposium on Research in Security and Privacy, www.mit.edu/people/eichin/virus/main.html. Visited November 15, 2010

FBI. (2007). United State of America, Federal Bureau of Investigation press release, "Over 1 million potential victims of botnet cyber crime," June 13, 2007, www.fbi.gov/press-rel/pressrel107/botnet061307.htm. Retrieved November 11, 2010

FBI. (2010). United State of America, Federal Bureau of Investigation, Department of Justice press release, "Bellevue Man Sentenced on Computer Hacking Charges," November 9, 2010, http://cleveland.fbi.gov/dojpressrel/pressrel10/cl110910b.htm. Retrieved November 11, 2010

Gartner. (2007). Gartner Research press releases. www.gartner.com/it/page.jsp?id=565125 Retrieved November 12, 2010

Gordon, L. A., Loeb, M. P., and Sohail, T. (2010). "Market Value of Voluntary Disclosures Concerning Information Security." MIS quarterly Vol. 34, No. 3

Hasan R, Myagmar S, Lee A.J, and Yurcik W. (2005). Toward a Threat Model for Storage Systems. *Storage SS'05*, November 11, 2005, Fairfax, Virginia, USA

Hasan, R., and Yurcik, W. (2006). A Statistical Analysis of Disclosed Storage Security Breaches. International Workshop on Storage Security and Survivability: in conjuction with 12th ACM Conference on Computer and Communications Security, October, 2006.

Helbush, A. (2009). Phishing Attacks Still on the Rise. Where to Start Technology Solutions Blog, http://www.wtsci.com/2009/11/Phishing-attacks-still-on the rise/ Visited November 11, 2010

Hoffer, J. A., and Straub, D. W. (1989). "The 9 to 5 Underground: Are You Policing Computer Crimes?," *Sloan Management Review* (30:4), pp. 35-43

Johnson, A C, and Warkentin, M. (2010). "Fear Appeals and Information Security Behaviours: An Empirical Study." MIS quarterly Vol. 34, No. 3

Keizer, G. (2010). Apple Smashes Patch Record with gigantic Update. Computer World.com/s/article/9196118/Apple_smashes_patch_record_with_gigantic_update. Visited November 5, 2010

Klienman, M. (2007). Microsoft helps FBI bust Chinese gang. Daily Telegraph online, July 25, 2007, www.telegragh.co.uk/finance/markets/2812822/Microsoft-helps-FBI-bust-Chinese-gang.html. Retrieved November 11, 2010

Mahmood, M.A, Siponen, M, Straub, D, Rao, H.R, and Raghu, T.S. (2010). "Moving Toward Black Hat research in Information Systems Security: An Editorial Introduction to the Special Issue." MIS Quarterly Vol. 34 No 3. Pp 431-433, September 2010.

Martin, R J, and Weadock, G E. (1997). Bulletproofing Client/Server Systems. New York –McGraw-Hill Co., Inc.

Oates, B J. (2009). Researching Information Systems and Computing. London - SAGE Publications Ltd

Peltier, T R, Peltier, J, and Blackley, J. (2005). Information Security Fundamentals. Boca Raten, Florida – CRC Press Company

Privacyrights.(2011). A chronology of data breaches reported since the Choicepoint incidence (list). Privacy Rights Clearing House. http://www.privacyrights.org/ar/ChronDataBreaches.htm. Retrieved 3/6/2011

Richo Security Solutions. (2010). www.ricoh.com. Visited November 17, 2010

Rosenhan, D, and Messick, S. (1966). Affect and Expectation. Journal of Personality and Social Psychology, vol. 3, pp. 38-44, as cited in Wikipedia, "Valence effect", http://en.wikipedia.org/wiki/valence_effect. Visited November 17,

2010

Ryan, J C H, and Jefferson, T I. (2003). The Use, Misuse and Abuse of Statistics in Information Security Research. Proceedings of the 2003 ASEM National Conference, St. Louis, Missouri

Shostack, A and Stewart, A. (2009). The new approach to Information Security. Harlow, Essex – Pearson Education Ltd.

Siponen, M and Vance, A. (2010). "Neutralisation: New Insight into the Problem of Employee Information Systems Security Policy Violations." MIS quarterly Vol. 34, No. 3

Somasundaram, G and Shrivastava, A. (2009). Information Storage and Management: Storing, Managing, and Protecting Digital Information. Indianapolis, Indiana – John Wiley and Sons. Chapter 15

SonicWall. (2008). Phishing Facts. www.sonicwall.com/phishing. Retrieved November 12, 2010

Stamp, M. (2006). Information Security: Principles and Practice. Hoboken, New Jersey – Wiley and Sons, Inc.

Tehan, R. (2005). Personal Data Security Breaches: content and incident summaries. In Congressional research Service Report for Congress, December 16, 2005.

Verizon. (2011). Data Breach Investigation Report. ([Online] Available: http://www.verizonbusiness.com/resources/reports/rp_data-breach-investigations-report-2011_en_xg.pdf, 15/3/12)

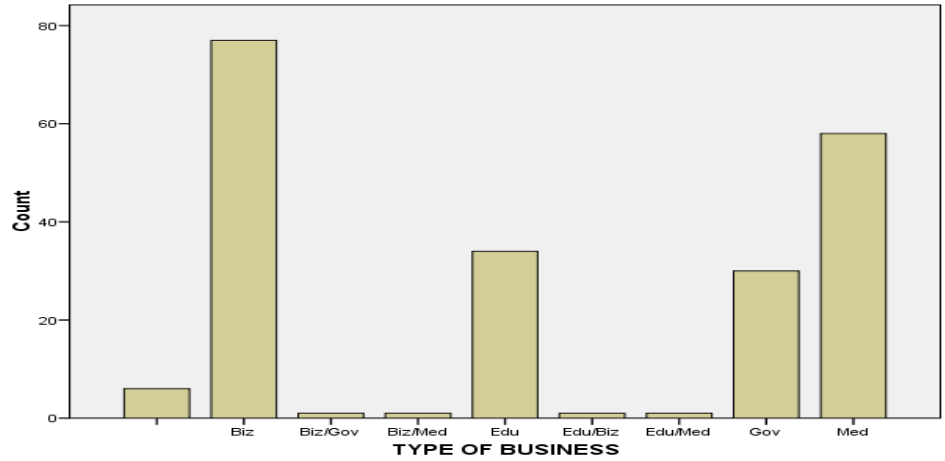

Figure 1 - Number of reported storage breach incidents by organization type

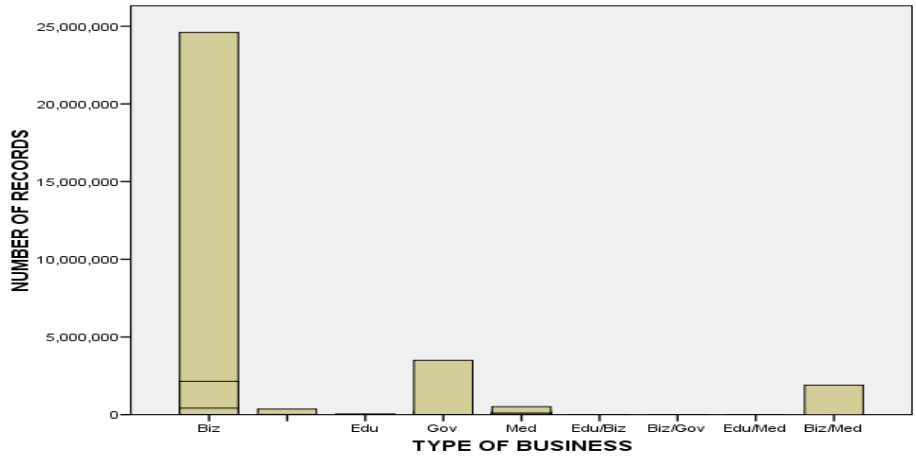

Figure 2 - Reported storage records lost by organization type

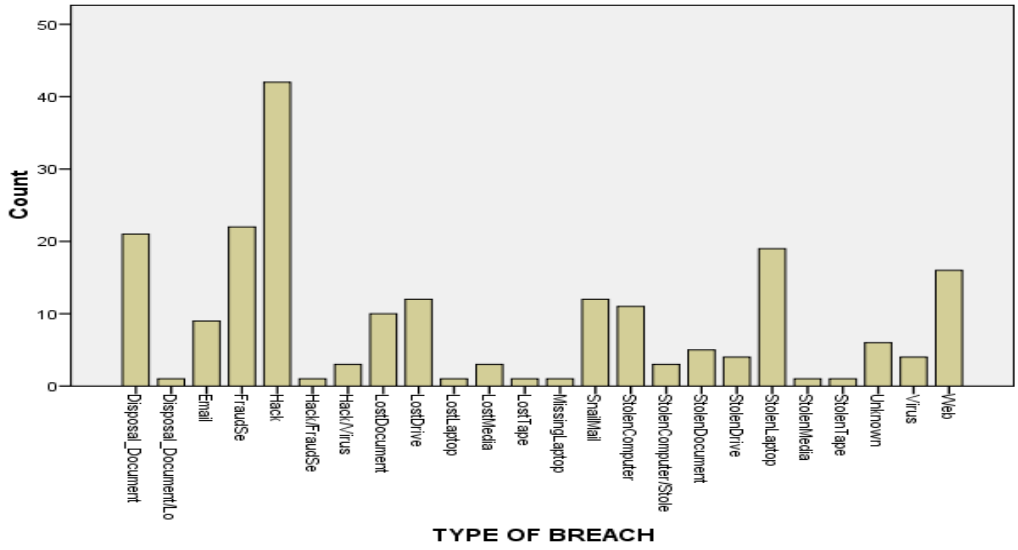

Figure 3 – Number of breaches by Breach Mechanism

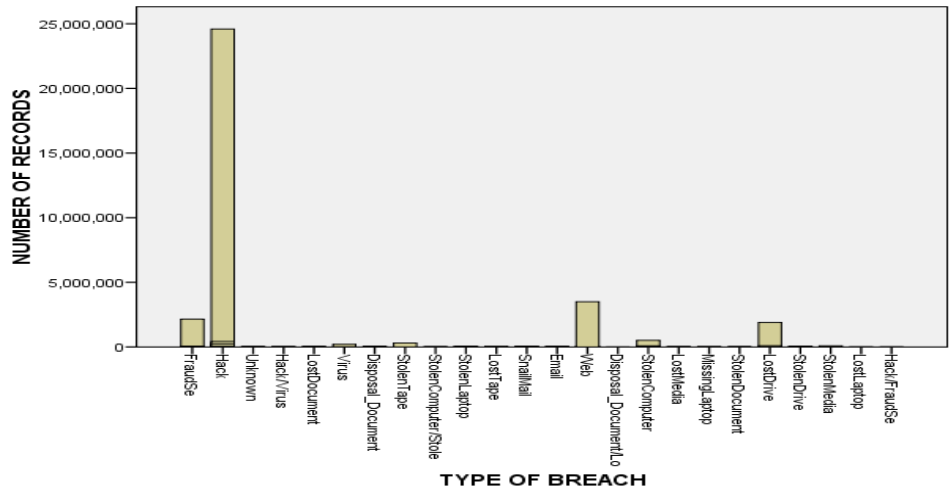

Figure 4 – Number of Records Lost by Breach Mechanism

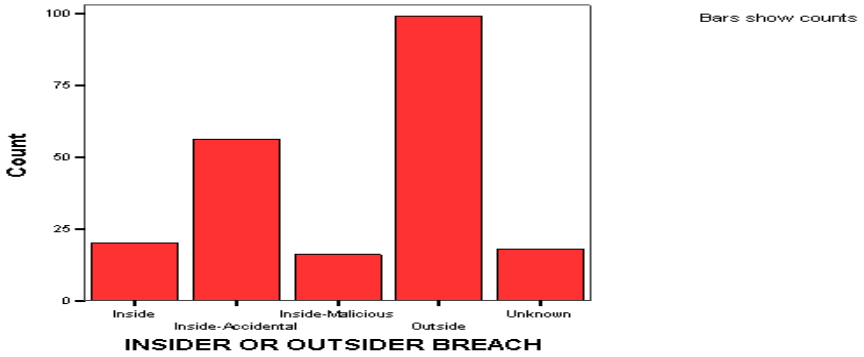

Figure 5 – Number of breach incidents by Insider or Outsider

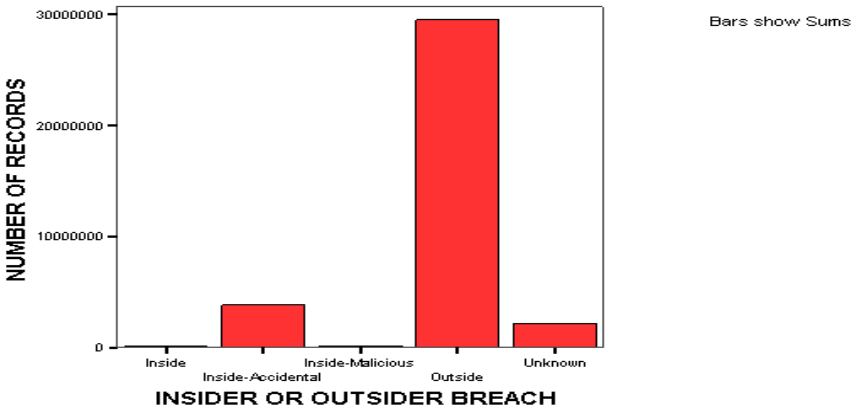

Figure 6 – Number of Records Lost by Insider or Outsider

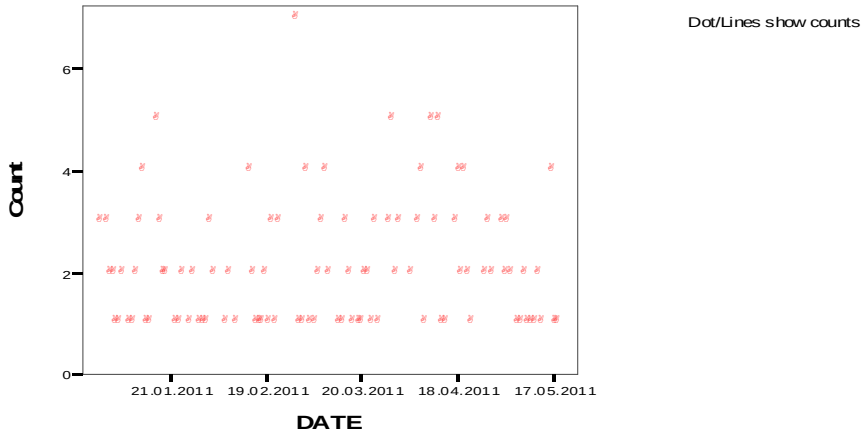

Figure 7 – Number of Breaches in Time

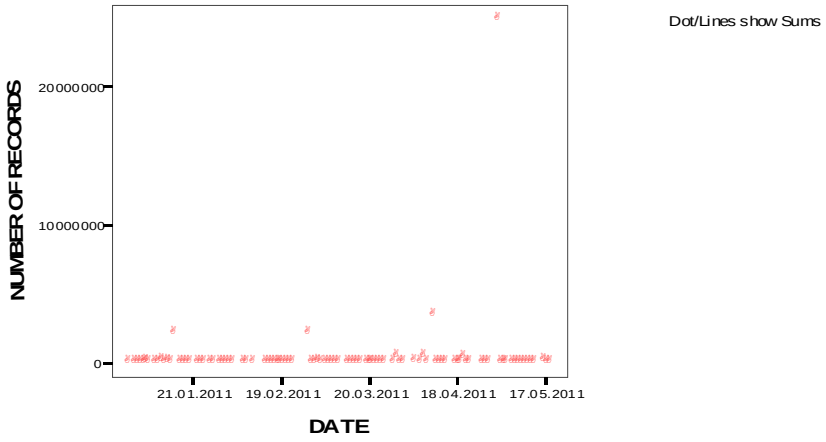

Figure 8 – Number of Records Lost in Time

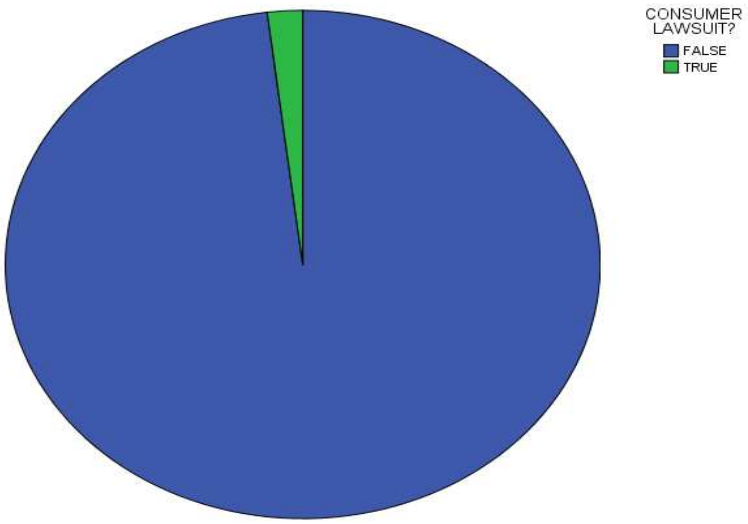

Figure 9 – Fraction of Law Suits

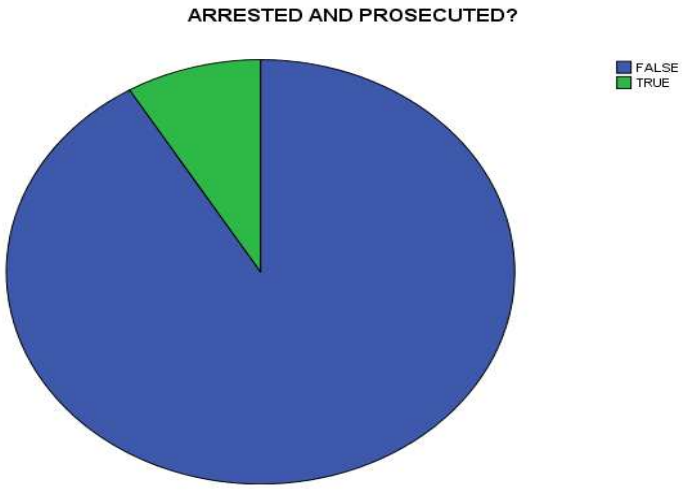

Figure 10 – Fraction of Perpetrators Arrested/Prosecuted

Table 1 - Number of Records per Incidence Summary Statistics

Descriptives

			Statistic	Std. Error
NUMBER OF RECORDS	Mean		170563.45	119467.4
	95% Confidence Interval for Mean	Lower Bound	-64958.67	
		Upper Bound	406085.57	
	5% Trimmed Mean		7059.38	
	Median		550.00	
	Variance		3E+012	
	Std. Deviation		1727120	
	Minimum		0	
	Maximum		24600000	
	Range		24600000	
	Interquartile Range		5887	
	Skewness		13.783	.168
	Kurtosis		195.113	.335

Table 2 – Number of breaches by Breach Mechanism

		Count
TYPE OF BREACH	Disposal_Document	21
	Disposal_Document/Lo	1
	Email	9
	FraudSe	22
	Hack	42
	Hack/FraudSe	1
	Hack/Virus	3
	LostDocument	10
	LostDrive	12
	LostLaptop	1
	LostMedia	3
	LostTape	1
	MissingLaptop	1
	SnailMail	12
	StolenComputer	11
	StolenComputer/Stole	3
	StolenDocument	5
	StolenDrive	4
	StolenLaptop	19
	StolenMedia	1
	StolenTape	1
	Unknown	6
	Virus	4
	Web	16