View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AIS Electronic Library (AISeL)

Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2009 Proceedings

Americas Conference on Information Systems (AMCIS)

2009

Mitigating security breaches through insurance: Logit and Probit models for quantifying e-risk

Arunabha Mukkhopadhyay
Indian Institute of Management Lucklow, arunabha@iiml.ac.in

G K. Shukla
Indian Institute of Management Lucklow, girja@iiml.ac.in

Follow this and additional works at: http://aisel.aisnet.org/amcis2009

Recommended Citation
Mukkhopadhyay, Arunabha and Shukla, G K., "Mitigating security breaches through insurance: Logit and Probit models for quantifying e-risk" (2009). *AMCIS 2009 Proceedings*. 767.
http://aisel.aisnet.org/amcis2009/767

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Mitigating security breaches through insurance: Logit and Probit models for quantifying e-risk

Arunabha Mukhopadhyay
Indian Institute of Management Lucknow
arunabha@iiml.ac.in

G K Shukla
Indian Institute of Management Lucknow
girja @iiml.ac.in

Abstract
The common e-threats deterring ecommerce are identity theft, hacking, virus attack, graffiti, phishing, Denial of Service (DoS), sabotage by disgruntled employees, loss of laptop, financial fraud and telecom driven frauds. These discourage users from online transactions. Organizations spend millions of dollars to implement the latest perimeter and core security technologies, to deter malicious attackers and to ensure confidentiality, integrity and availability of data. Yet, security breaches are common. It results in loss of opportunity cost, market capitalization and brand equity for organizations. We propose e-risk insurance as a strategy to supplement the security technologies, and to mitigate these financial losses. In this paper, we propose two generalized linear models (GLM) namely Logit and Probit for quantification of the probability of an e-threat, using CSI/FBI data. We also compute the expected loss amount for organizations using collective risk model. Based on it, we ascertain the net premium to be accrued to the insurance companies.

Keywords
IS security, IS risk, e-risk quantification, security breach, e-commerce, Logit and Probit models, e-risk insurance.

Introduction

Online security breaches are adversely impacting the top and bottom lines of most companies, worldwide. Federal Trade Commission in a report states that 8.3 million US citizens had been victims of identity theft in 2005 (http://www.idtheftcenter.org/). CSI-FBI report 2007, states that the total amount of loss suffered due to malicious attack amounted to $66,930,950. Loss due to financial fraud alone contributed to $ 21,124,750, (i.e., 31% of the total loss) (Richardson, 2007). This includes compromise of e-commerce sites mostly. The data compromised from these servers are used for identity thefts. In February 2008, the eBay's Korean unit had been hacked and 10 million users private information leaked. Chinese hackers attacked South Korea's oldest and largest online shopping site, Auction.co.kr, this year. They compromised information of 18 million customers and took away financial data, too, from the organaization's servers. In another incident, Max Ray Butler, a hacker, owned Cardersmarket website. He used his ecommerce site to hack into computer networks, to steal credit card and other personal information. He later used this information himself or resold them for malicious intent. In 2007, hackers compromised Monster Worldwide Inc servers and got hold of 146,000 users' personal information. Monster Inc was running USAjobs.gov site, on behalf of the federal government. In the same year, AOL had an attack on its servers (http://www.pogowasright.org and http://www.privacyrights.org/ar/ChronDataBreaches.htm). This infected the servers with malicious programs and therein compromised confidential customer data from their databases. According to Gartner and the Ponemon Institute, the loss of a single non financial record amounts to $197. Approximately, 127 million records were lost in 2007, which amounts to a loss of $25 billion (Burger, 2008). Microsoft UK faced a graphiti attack, in 2007, when a set of hackers compromised the website's security and performed a SQL injection. The most malicious botnet of 2007 was "Storm,". It tricked people into opening an email, with the subject "230 dead as storm batters Europe". Subsequently, the Trojan worm was downloaded into the unsuspecting user's machine Garretson). Researchers at Google have reported, that 10% web pages could successfully "drive-by download" a Trojan virus onto a visitor's computer. They had used a sample size of 4.5 million websites for their study. Once the malicious software is downloaded into the user's machine, hackers can easily compromise the target machine. The common techniques used by hackers are to place malicious software on Web sites, manipulating Web server security, manipulating user-posted content, advertising and third-party widgets (Provos, McNamee, Mavrommatis, Wang and Modadug, 2008). In a response to curve bad incidents, chief technology officers (CTO) resort to implementation of perimeter and core security (Tanenbaum, 1996, Whitman and

Mattford, 2007) technologies to prevent confidentiality (C), integrity (I) and availability (A) of data (Gordon and Loeb, 2002; Gordon A, L, Loeb, P. M., Sohai, T, 2003; Dhillon and Gholamreza, 2005; Dhillon and Backhouse, 2000; Anderson, 2001; Schneier, 2000). The CSI FBI report 2007, states that 98% organizations have anti-virus software, 97% have firewall, 84% have implemented virtual private networks (VPN), 80% have anti-spyware, 69% have intrusion detection system (IDS), 51% have strong password management, 47% encrypt the confidential data stored, and so on (Richardson, 2007). These technological investments help to protect their information assets and also prevent their networks being clogged, sniffed or snooped. This in turn helps, to reduce the frequency of the bad event. Numerous research efforts have gone into developing systems and networks that are invincible. Yet complete security still remains a myth (Austin, D. R, Darby A. R. C, 2003). White House, NASA, and Penatagaon website too have been compromised, in the recent past. In reality a combination of technology, policy and use of financial instruments need to be done to mitigate the losses to online business organizations. These necessities online business organizations take a proactive initiative to (i) assess the risk of the organization and (ii) to put in place necessary business continuity (BC) disaster recovery (DR) plans.

The impact of such malicious attacks, can be broadly classified as loss of (i) opportunity cost (ii) market capitalization, and (iii) brand image (Mukhopadhyay 2007b; Mukhopadhyay, Chatterjee, Saha, Mohanti and Poddar 2005). *Opportunity cost* (OC) of any service is defined as the value of all the other services that one must give up in order to deliver it (Varian, 2003; Brealey and Myers, 2000). If the communication channel is down, then online business organizations have substantial loss of OC. For example, a Denial of Service (DoS) attack on e-bay on 11th June 1999 had resulted in a loss of revenue to the tune of $3 to $5 Million. *Market capitalization* (MC) is the product of stock price times the total number of outstanding shares (Brealey, and Myers, 2000). If the organization has too many malicious attacks, then its stock price takes a dip. That, in turn, affects the MC (Campbell, Gordon, and Loeb, 2003). For example a DoS attack on e-bay on 11th June 1999 had resulted in drop of its share price by 20% (Kesan, Majuca, and Yurcik, 2004). Recent studies have reported that the stock price of companies that had suffered a security breach in 2006 had fallen by 12 percent of its value since the breach (The *Wall Street Journal,* September 2006). *Brand* defines the attractiveness and familiarity in the market about a product/service provided by an organization. Brand equity permits companies to charge premium prices for products and services, contributing to increased profit margins. Companies invest huge amounts to develop brand equity. Too many adverse impacts tarnish brand image of an organization, and customers fear to transact with it (Campbell et al.,2003).

In the context of rising security breaches, Sarbane's Oxley (SOX) Act, Gramm-Leach-Bliley (GLB) Act and Health Insurance Portability Insurance and Accountability Act (HIPAA), have laid down stringent and mandatory procedures for organizations, to ensure privacy and security of data. This aims to minimize the losses of security breaches. They also mandate that all organizations should have a proper contingency plan in place to mitigate the risk. Similarly, Basel II accord mandates that banks should do a proper risk assessment and quantification of their operational risk, and set aside a capital charge for the financial organization. Lawsuits can arise if a third party gets affected due to non adherence to these requirements.

Of late there is a growing consensus about implementing IT governance by organizations. IT governance mandates effective management, policies, controls and procedures in place to ensure that an organization's information systems (i) support the organization's objectives and (ii) they are used responsibly and (iii) that IT-related risk is minimized. Effective IT governance (Brown and Grant, 2005; Solms 2005) is one element of compliance and corporate governance programme. Similarly, an effective IT governance programme helps to ensure efficient Enterprise Risk management (ERM) (Miccolis, 2000) process. To ensure proper information security management system (ISMS) is in place organizations need to be BS7799 certified. BS7799 in turn necessitates risk assessment and quantification.

In this backdrop, we propose that organizations implement techno-financial solution (i.e., e-risk insurance) as risk mitigation strategy (i.e., to minimize the OC, MC and brand image drop) against malicious online attacks. In this paper we propose the use of e-risk insurance as an effective mechanism for loss reduction to organization. The basic premise for use of e-risk insurance (Kesan, Majuca, and Yurcik, 2004; Reid, and Stephen, 2001; Schenier, 200a, Mukhopadhyay et al., 2005; 2005a; 2005b; 2006; 2007; 2007a; 2007b) is the quantification of the expected e-risk, in case of a bad incident. In this work, we take into account eight types of attack such as virus attack, insider net abuse, laptop theft, DoS, unauthorized access to information, financial fraud, theft of proprietary information and telecom fraud from the CSI-FBI report (i.e., 1997 to 2007), and using generalized linear model (GLM), such as Logit and Probit, formulate a probabilistic pattern of attacks, for the past years. Based on this model, we then arrive at the

expected probability of attack. Then using collective risk modeling we arrive at the expected loss amount and the premium that the organization will need to pay to be indemnified for the same.

Related work

We chronologically trace the methods and techniques used in the area of IS risk management from 1970 to till date. Figure 1 shows that the literature can be broadly split into 3 phases namely: (i) conventional IS risk management (ii) social and organizational issues related to IS risk management and (iii) the risk to online businesses (e-risk) (Mukhopadhyay, 2007b). The e-risk literature has originated post 2000. This paper proposes the use of insurance for mitigating the risk. This is in line with the developing e-risk literature which broadly aims to: (i) classify e-risk (ii) quantify e-risk, through qualitative (E-R (Quali)), quantitative (E-R (Quanti)) and hybrid (E-R (Hybrid)) methods, (iii) its impact on business, (iv) suggest optimal security solutions, (v) propose e-risk insurance solutions and (vi) analyze the existing e-risk insurance products. The late 90's to till date, multiple studies have been reported which use social and organizational aspects to assess and mitigate IT risk. The conventional IS risk management literature has its origin early as 1974, with the development of *formal methods* and sequentially the *evaluation criteria*. Efforts have been made since the late 70's to quantify IT risk (R-A (Quanti)) methods. The qualitative (R-A (Quali)) methods originated in early 80's, and the hybrid (R-A (Hybrid)) methods in the late 80's. We would critically evaluate the popular models used from each of these phase.

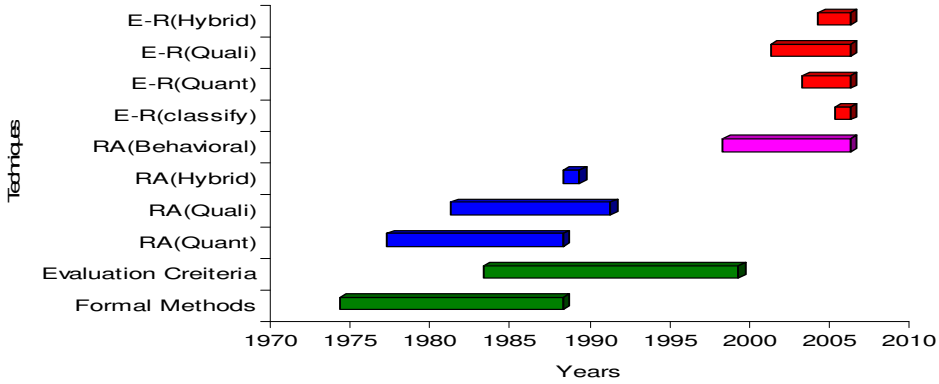

Figure 1: **IT risk literature** (E-R stands for e-risk, RA represents risk analysis)

Formal Models

They focus on developing an analytical basis for the design, specification, realization, implementation and evaluation of security systems. The objective was to come up with rigorous mathematical proofs related to IS security issues. With this aim, Bell La Padula, Biba, Clarkson- Wilson and the Jueneman, proposed their models between 1979 and 1989. Table 1 compares the models.

Table 1: Comparison of formal models

Models / ***Issue***	**Bell La Padula Model**	**Biba Model**	**Clarkson-Wilson Model**	**Jueneman Model**
Object of evaluation	Users, processors and data	Users, processors and data	Users, processors and data.	Users, processors and data
Classification basis	*Confidentiality*	*Integrity*.	*Confidentiality and Integrity*.	*Confidentiality*.
Beneficiaries	Military	Military	Business	Business across heterogeneous networks

Evaluation Criteria

These were developed to come up with a mechanism for evaluating the invincibility of the computer systems being developed. The formal methods, discussed in the previous section, were the basis of development of evaluation criteria. These techniques were equally popular in the USA and the European countries. Table 2 critically compares the models.

Table 2: Comparative study of evaluation criteria

Criteria	**TCSEC**	**ITSEC**	**CC**
Common name	*Orange Book*	-	*ISO 15408*
Location	USA	European Union.	USA and UK
Object of evaluation	Classified information.	Classified information.	Operating system, computer network and applications
Classification basis	*Confidentiality*.	*Confidentiality and Integrity*.	*Confidentiality, Integrity* and *Availability*
Base model	Bell La Padula Model	Clarkson-Wilson Model	-
Security levels	6	6	7
Focus Area	Military data	Military and business data	Military and business data

Risk Analysis

Risk analysis studies, related to IT, can be broadly classified into three categories, namely qualitative, quantitative and hybrid, as shown in Figure 2.

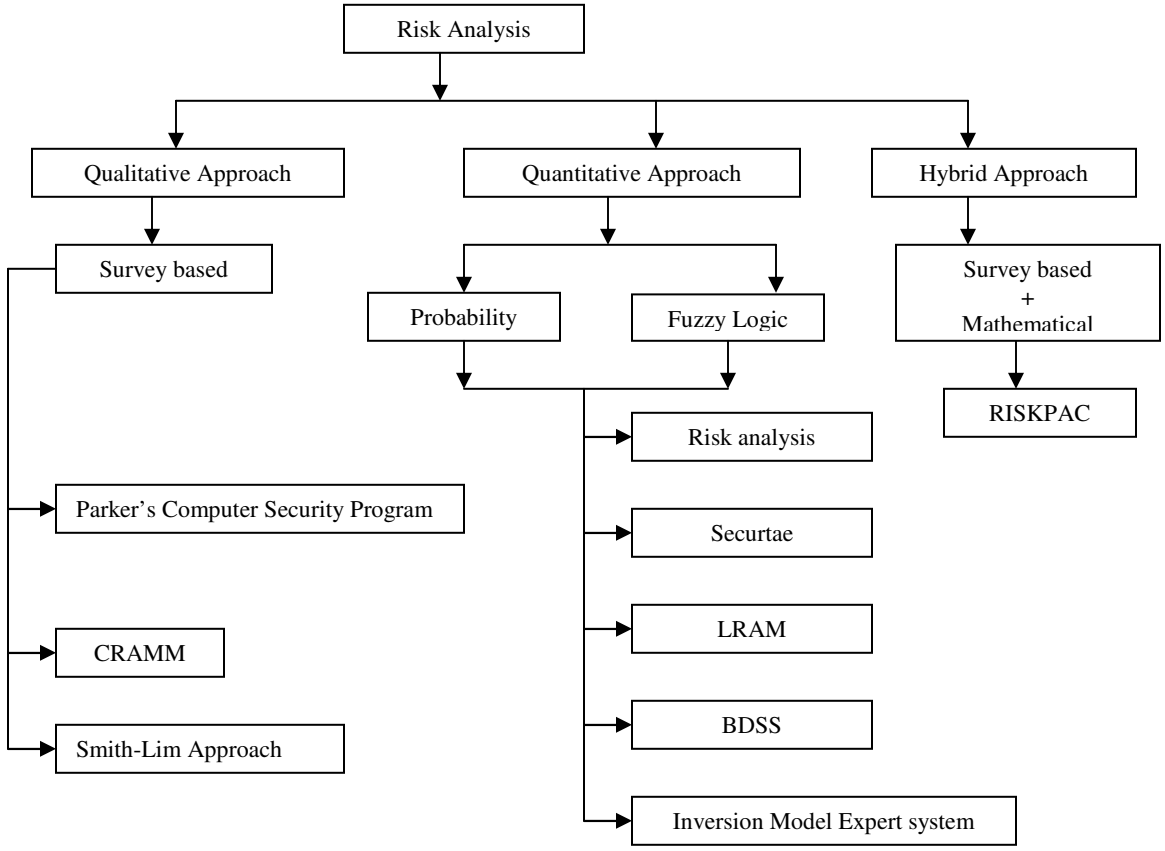

Figure 2: Classification of approaches in risk analysis study

The quantitative techniques resort to rigorous mathematical studies, involving probability theory or fuzzy theory, to arrive at a value of the risk. Qualitative approaches use questionnaires and extensive survey of organizations to arrive at the impact of a failure or sabotage. The hybrid approach captures some subjective parameters though questionnaire method and uses the mathematical technique to arrive at final risk estimation. We review some representative works under each category in the following subsections. Our work in this thesis falls under the quantitative category.

Quantitative techniques

These techniques aim to arrive at an expected value of loss due to a malicious impact on a computer system. A comparative analysis of quantitative techniques is shown in Table 3. We categorize each of the techniques based on methodology and the output produced. It is to be noted that risk analysis model by Courtney (1977), first formulated that expected loss is the frequency of the impact times the loss associated with each impact.

Table 3: Comparative analysis of quantitative techniques.

Features		**Risk Analysis (1977)**	**Securtae (1978)**	**LRAM (1987)**	**BDSS (1988)**	**Inversion Model Expert system**
Methodology	Fuzzy Logic		√			
	Bayesian			√	√	
	Expert system					√
Output	Expectation of loss	√		√	√	
	Vulnerability analysis		√		√	
	Identification of threat			√	√	√
	Safeguards for security				√	√
	Controls to be implemented			√		

(***LRAM =Livermore Risk Analysis Methodology; BDSS =Bayesian Decision Support*** System ***Qualitative techniques*****.**
√ = used method)

Qualitative techniques

These techniques use extensive interview or questionnaires to arrive at an expected value of loss due to a malicious impact on a computer system. We chronologically trace the qualitative techniques proposed for IT risk analysis.

- **Parker's Computer Security Program (1981)**

It is a five phase model, as elaborated in Table 4.

Table 4: Phases of Parker's model

1	Identification and valuation of assets	(i) Identify assets (i.e., people, supplies, hardware, and data), (ii) Form of assets (i.e., moveable property, magnetic patterns, printed paper, (iii) Location of sites (i.e., remote, internal environment, computer) (iv) Accountability (i.e., first party or third party).
2	Identification of threats.	(i) Source (i.e., employees or external vendors), (ii) Motives (i.e., human failure, irrational behavior), (iii) Act (i.e., physical, logical, covert etc); results (i.e., disclosure or destruction) (iv) Losses (i.e., monetary or denial of service).
3	Risk assessment	(i) 2 way matrix, vulnerable assets on x-axis, and the occupational nature on the y-axis. (ii) It is assumed that each of the assets have certain types of attacks (i.e., modification, destruction, disclosure, taking and denial of use).

		(iii) The exposure of each of the assets, with respect to occupation is assessed on a scale of 5 (e.g., 5 indicate up to 100%, 4 indicate up to 80% etc). (iv) A value is attached to each of the assets. This provides a risk exposure data for each of the asset. (v) The top management using the concept of rational thinking, decide to implement security, for each of these assets (Baskerville 1993).
4	Planning for Security	This decision is based on the cost–benefit analysis, minimal human interference, total reliability etc
5	Implementation of the security Safeguards	It is a continuous process and safeguards need to be continuously monitored, as the vulnerability of the assets keeps changing.

This model takes into account also social and human factors (such as motives, acts, sources of threats) associated with security (Baskerville 1993). The study is broadly qualitative in nature.

- **Smith-Lim Approach(1984)**

The model consisted of two matrixes for risk evaluation. The approach is detailed in Table 5. For example very high vulnerability and very low impact, implied low risk for and organization. The rules would be customized depending on the organization (Baskerville, 1993).

Table 5: Smith-Lim Approach

It assumed 3 generic threats and 4 generic targets/assets.

Threat-target matrix.

Threat / ***Target/Assets***	***Natural hazard***	***Direct human***	***Indirect human***
Facility			
Hardware			
Software			
Documents			

A rule base was developed, indicating the impact of risk arising from 2 parameters,
(i) vulnerability (i.e., absence of safeguards) and
(ii) impact (i.e., severity of impact).

Vulnerability impact analysis

Impact / **Vulnerability**	***Lo***	***Hi***
Lo		Low risk
Hi		

- **CRAMM (Farquar 1991).**

This qualitative method was used by Britain's Central Computer and Telecommunication Agency (CCTA). This model has three stages, as illustrated in Table 6. CRAMM model output a set of recommendations for the information security of an organization. The main criticism of this model was that it churned highly technical reports. The solutions provided were in terms of safety measures. There was very little focus on social and human factors (Baskerville 1993).

Table6: CRAMM model

	Objective	**Methodology**
1.	Identification of assets and assignment of monetary values.	The monetary value associated based on impact in case of failure or sabotaged.
2.	Grouping assets based on their	The owners of the asset asked to evaluate the vulnerability

	vulnerabilities	associated with it (5 point scale used).
3.	Evaluate the existing controls present in the organization	

Hybrid techniques

This approach uses a combination of qualitative and quantitative techniques. A questionnaire method is used for data collection. This data is then fed into rigorous mathematical models.

- **RISKPAC (Computer Security Consultants 1988).**

It was a mixture of qualitative and quantitative approach. There were 12 types of questionnaire used to collect data from the end –user, as shown in Table 7.

Table 7: RISKPAC

Questionnaires		Other Inputs	Output
Business issues	The product and services the organization produce	Risk profiles of Organization. Controls already present in the organization	Safeguards for the organization
	Necessity of computerization		
MIS and business	Applications used and impact to business		
	Personal computers and its related configuration		
	Computer systems and their operating systems		
Risks	Computer applications, the related operational risk and sensitivity to business		
	Network associated risks		
IT Audit	IT audit and its necessity		
	Type of security evaluation criteria to be used		
	Physical security of the organization		
BCP/DR	Exposure to natural disaster		
	Type of backup and disaster planning.		

(MIS= management information Systems; BCP/DR= Business Continuity and Disaster recovery)

The basic tenants of RISKPAC are summarized as follows: (i) it was based on Utility theory, (ii) it uses qualitative techniques for estimating security (i.e., linguistic variables, as in fuzzy sets), (iii) user friendly and could be understood by non-professionals too; (iv) used Courtney's model (Baskerville, 1993) to estimate the expected loss. The basic drawbacks included: (i) limited capability to model; changes in organizational issues, (ii) it did not provide a mechanism of the viability of two or more security strategies (Baskerville 1993).

Value based assessment of information system.

A value focused approach was used to study the fundamental objectives for information system's (IS) security and the means to achieve it. Data was collected through in depth interviews of IT mangers in various organizations. The main objective was to judge the values of people regarding IS security. This data was in turn validated by security experts. It was noted that social, human and interpersonal issues also contributed to towards IS security. These issues need to be looked into, for a proper risk assessment strategy. For example issues such as level of employee motivation, work ethics, level of personal privacy, capability level of an individual, knowledge of the business process, level of control etc were looked into (Dhillon et al., 2005).

RITE model for IS security.

Technology issues, coupled with social and organization factors, give a holistic picture of the vulnerability in an organization. The security studies should go beyond issues of confidentiality, integrity and availability. Social issues, such as responsibility (R) of roles, integrity (I) of employees, trust (T) and ethicality (E) too, need to be looked into. The behavioral factors need to be monitored in an organization to get a proper understanding of the risk perceptions (Dhillon et al., 2000).

End –to- end planning for information security

All computer systems are vulnerable. An end to end planning solution for risk mitigation associated with IT comprising of 4 stages are discussed here (Straub & J.Richard, 1998). Table 8 summaries their work.

Table 8: Strategies for end –to end planning

1.	Recognition of security problem	(i) Overall risk perception of the industry; (ii) Controls imposed in the organization; (iii) Local risk perception, (iv) Formation of the managerial perception of the system risk.
2.	Risk analysis	Managerial perception is used as the basis for the risk calculation.
3.	Generation of alternatives to mitigate risk	The alternatives include (i) deterrence; (ii) prevention; (iii) detection; (iv) Remedies. The focus is to maximize prevention and minimize remedial action.
4.	Finding out a viable strategy for implementation of security	(i) Chalk out a planning decision for security implementation. This involves developing a counter matrix to evaluate the effectiveness of strategy vis-à-vis the alternative strategies discussed in stage 3. (ii) Implement the best security strategy, depending on cost –benefit analysis and the risk perception of the system

All these follow a feedback loop and decisions are taken in an iterative fashion. This mechanism is innovative as it discusses a planning horizon solution to the risk mitigation issue.

A generalized linear model for e-risk quantification

Our aim is to quantify the probability of occurrence of bad events such as virus attack, insider net abuse, laptop theft, DoS, unauthorized access to information, financial fraud, theft of proprietary information and telecom fraud based on data available from the CSI-FBI report (i.e., 1997 to 2007). We, use *generalized linear model* (GLM) (McCullagh and Nelder, 1989, Dobson, 2002) to fit the CSI –FBI attack data. The CSI survey is the most widely quoted set of statistics in the industry, and is currently in its 12^{th} year. In the 2007 survey, they had 494 respondents. This number has varied over the years. The sample space was spread across both the government and private sector and multiple industries. The representation of the major sectors was as follows: financial sector (20%), followed by consulting (11%), education (11%), information technology (10%), and manufacturing (8%). 41% of the respondents were "people responsible for enterprise security". We have chosen the GLM where, the expectation (μ) of the dependent variable Y, is a member of the exponential family. Y is connected to the linear combination of the variables $X^T\beta$, where X denotes time. Equation (1) defines the relationship between time X and Y variables.

$$\mu = E(Y|X) \; ; \; \eta = X^T\beta \; ; \; \eta = G(\mu_j) \; ; \qquad (1)$$

$$E(Y|X) = G^{-1}(X^T\beta) \qquad (2)$$

Where we assume, $X^T\beta = \beta_0 + \beta_1 * t$ and $X^T\beta = \beta_0 + \beta_1 * t + \beta_2 * t^2$

Here we assume G to be either a linear or a quadratic Logit *link function,* and t denotes time. Therefore equation (2) can be written as shown in equation (3).

$$\text{Logit}(p) = \log(p / (1 - p)) = G(\mu) \qquad (3)$$

$$p = 1/(1+\exp^{-\eta}) \quad (4)$$

Equation (4) denotes the probability (p) of a malicious attack. η can take any value in the range from $-\infty$ to ∞. The exponential function $\exp^{-\eta}$ is always positive, and so p of each attack maps to the range [0, 1] i.e., $0< p <1$.

Our proposed GLM models for studying the attack patterns are as shown in equation (5, 6, and 7)

Model 1: $p = 1/(1 + \exp(-\beta_0 + \beta_1 * t)$ (5)

Model 2: $p = 1/(1 + \exp(-\beta_0 + \beta_1 * t + \beta_2 * t^2)$ (6)

Similarly, we propose a linear probit model to fit the data points.

Model 3: $G(\mu) = \Phi^{-1}(\mu) = \beta_0 + \beta_1 * t$, (7)
where Φ is normal cumulative density function.

We assume all the eight types of attacks, as surveyed by CSI follow a binomial distribution, have probability mass functions as shown by equation (8).

$$f_i(y_i) = {}^{n_i}C_{y_i}\, p_i^{\,y_i}(1-p_i)^{n_i - y_i} \quad ; i = 1 \text{ to } k \text{ years} \quad (8)$$

Here n_i denotes the sample size of respondents interviewed in the i^{th} year, y_i is the number of respondents who had felt the brunt of the attack in the i^{th} year , p_i is the probability of particular type of attack occurring and $(1- p_i)$ is the probability of a particular type of attack not occurring. The parameters of the GLM are estimated by the maximum likelihood estimate (MLE) method. For a given probability distribution specified by $f(y_i; \beta, F)$ and observations $y = (y_1, y_2 \ldots y_n)'$, the log-likelihood function for β and *F*, expressed as a function of mean values $\mu = (\mu_1, \ldots, \mu_n)$ of the responses $\{Y_1, Y_2, \ldots, Y_n\}$, has the form shown in equation (9).

$$LN(\mu_i; y_j) = \sum_{i=1}^{k} LN(f(y_i)) \quad (9)$$

The MLE of the parameters $\boldsymbol{\beta}$ is obtained by iterative re-weighted least squares (IRLS) (McCullagh and Nelder, 1989).

We choose the GLM model, as opposed to the linear regression models (i.e., $E(Y|X) = X^T\beta$ or $Y= X^T\beta +\varepsilon$), for the study, as the error term (ε) needs to be continuous in nature and hence it implies that the response Y must have a continuous distribution as well. Hence, linear regression model fails to deal with discrete Y's as that are studied in this case. Also, analysis of the attack data we find that it follows a non-linear trend. Figure 3, shows the number of respondents, who have agreed to the fact that they faced any of this eight type of attacks of the years (i.e., 1999 to 2007). It is also evident that insider net access and financial fraud are on the rise, whereas the other six types are on the decline. A linear model will not be a good fit for this set of data points.

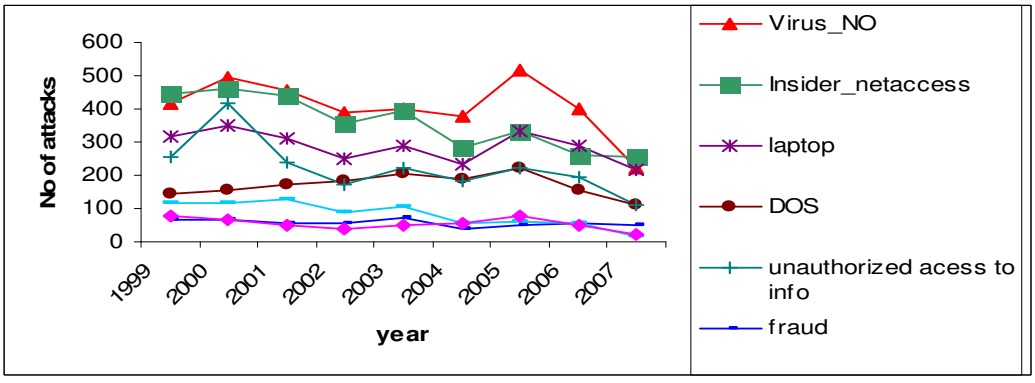

Figure 3: The number of attacks from 1999 to 2007

The main reasons sighted for the decline of the attacks are (i) implementation of security technology (Bagchi and Udo, 2003), (ii) mandatory compliance to SoX, GLB, HIPAA or (iii) BS7799 norms (Mukhopadhyay et. al.,2006, Richardson, 2007) and so on.

We find a similar type of work by Bacghi et al., 2003, where they had proposed a Gompertz model for predicting the growth of the malicious attacks. They had used the model shown in equation (10).

$$N(t) = \exp[(-c/q)\exp(-qt) + K] \quad (10)$$

where t = time, N(t) = cumulative number of attacks incidents at time t, c= net rate of instigation to attacks, q= rate of inhibition to attacks.

We argue that the probability models, such as Logit and Probit are a better choice, as we are modeling a totally uncertain event. It is always a win- win situation for the malicious intruder. A smart hacker needs only a single point to break into an organizations network, whereas the CTO needs to know all the gaps, to prevent his entry (Anderson, 2001). Similarly, a zero-day attack can compromise any organization's core and perimeter security and cause enormous damage. The number of attacks in a given year is totally independent of the previous year, as CTO's implement security measures at the earliest to avoid further losses. Thus, use of cumulative number of respondents as in equation (6) is not justified. The use of probabilistic model such as Bayesian Belief Networks (BBN) is also justified, as experienced CTO's, can provide prior belief regarding the incidence of malicious attacks against information assets (Jensen, 1996). These beliefs can be modified as more data is obtained. It is also noted that the use of security technology merely lowers the probability of the attack and not the actual number of attacks. It is of outmost importance that organizations have a contingency plan, to mitigate the monetary loss, whenever an attack happens. Such business continuity plans needs to budget for the expected loss of revenue, due to a malicious attack, as opposed to the actual number of attacks. We suggest that use of e-risk insurance (Kesan, Majuca, and Yurcik, 2004; Reid, and Stephen, 2001; Schenier, 200a, Mukhopadhyay et al., 2005; 2005a; 2005b; 2006; 2007; 2007a; 2007b) as an instrument to hedge such uncertain risk. The basic premise of insurance rests on the assumption that the expected exposure to risk can be computed. The use probabilistic models such as Weibull, Poissson, Negative Binomial etc to model the frequency of the risk and also the claim amount distributions, in case of an expected calamity (Hossack, Pollard and Zehnwirth, 1983) is common. Thus, under all circumstances, use of a probability is well justified for the model.

Using our 3 models, defined in equation (5, 6, 7), we compute the probability of attack. We use collective risk model (Hossack et al., 1983), to compute the expected loss. Collective risk model assumes that the number of attacks (N) and the loss amounts (L) are both stochastic in nature. Let us assume that there are N malicious attacks, such that each attack leads to a loss of L_i. Each of these losses are independent and identically distributed (iid). The total loss (S) for virus attack, to an organization in a given year, is shown in equation (11).

$$S = L_1 + L_2 + \ldots\ldots\ldots\ldots\ldots + L_N \quad (11)$$

The expected value of the loss due to an attack is computed using equation (12).

$$E(S) = E(N) * E(S) \quad (12)$$

$$E(S) = n_i * p_i * MA_i$$

Where n_i = denotes the sample size of respondents interviewed in the i^{th} year, p_i= probability of attack in the i^{th} year, MA_i= monetary impact of an attack in the i^{th} year . To compute the premium, we need to compute the variance of the total loss (S), given by the equation (13).

$$Var(S) = E(N) * Var(L) + \{E(L)\}^2 * Var(N) \quad (13)$$

Then based on the expected loss arrived at in equation (11) the premium is defined as the expected loss E ($LAmt_i$) times the times The premium is arrived at by using equation (14), after taking into account overhead loading (OV) and the contingency loading (k).

$$Premium = (1+OV) * E(S) + k * \sqrt{Var(S)} \quad (14)$$

Results and Discussion

Figure 4a and 4b illustrates the Logit and Probit curves that fit onto the probability of attack data points. The red line indicates the Logit curve, while the green line depicts the Probit curve. We did not plot the quadratic Logit equation, since we found that there was significant improvement to the curve fitting using it. The Logit and the Probit data curves almost superimpose on each other. The expected probability curves reflect a decrease in all types of attacks from 1999 to 2007. It is to be noted that the decrease has occurred following a smooth exponential manner pattern, over time. This is very clearly evident for the theft of propitiatory information, in Figure 4b. This corroborates to the fact that there will always be some amount of threat of an attack, and that it can never become zero. In Figure 3a, especially, the data points of DoS attacks is so widely dispersed, that it is not possible to fit any probabilistic curve (such as Logit or Probit) to it.

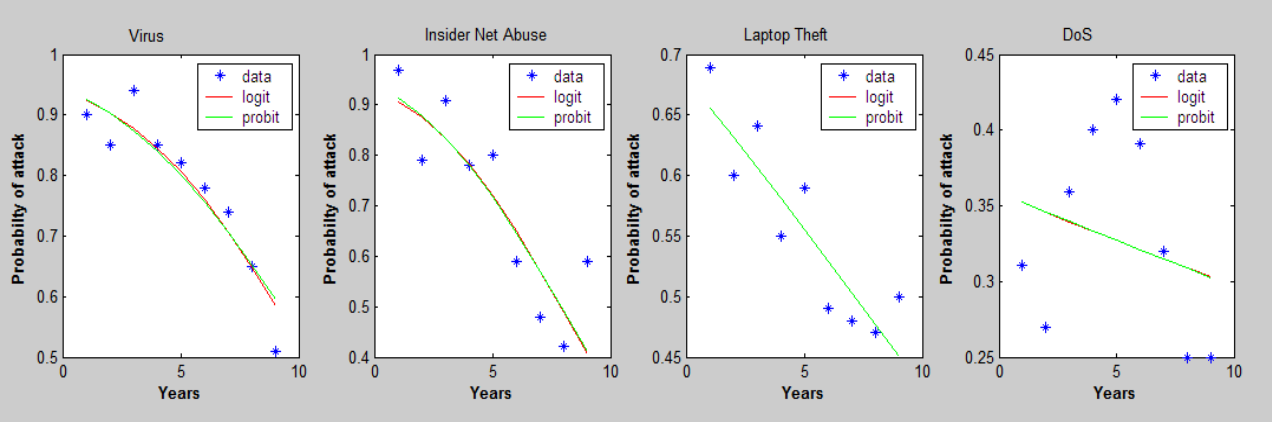

Figure 4a: Logit and Probit curves fit to the probability of attack data

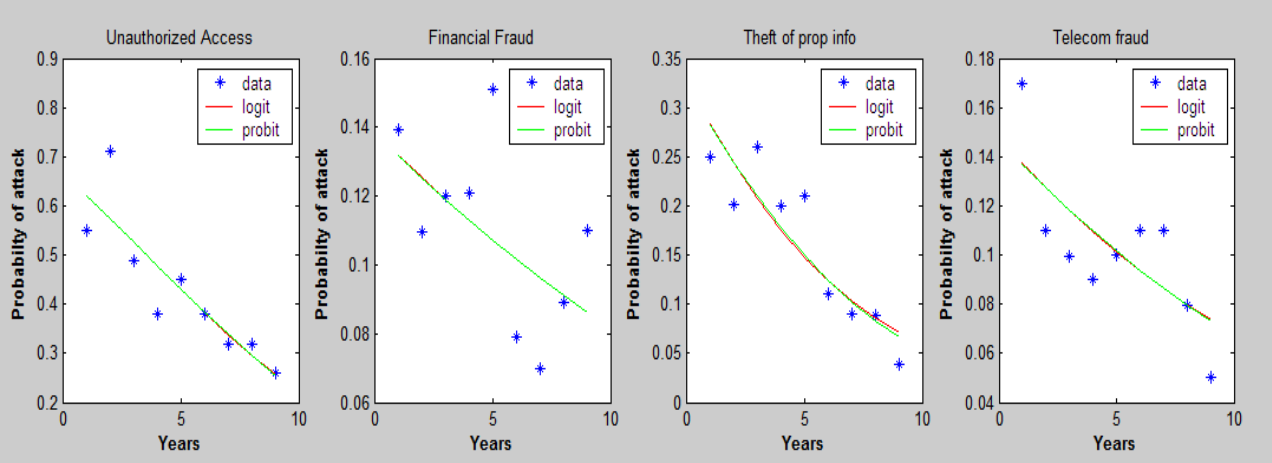

Figure 4b: Logit and Probit curves fit to the probability of attack data

The coefficients of the Logit function, and the the chi square and deviance values are shown in Table 9. The values of the Probit curve are also almost similar. The chi-square values (degrees of freedom 7, at significance level 5%) are then compared to the tabulated ones. Deviance is computed using equation (15).

$$D = 2\sum_{i=1}^{k} y_i \log\left(\frac{y_{ii}}{y_i}\right) + (n_i - y_i)\log\left(\frac{(n_i - y_i)_i}{n_i - y_i}\right), \text{ k= 1 to 9} \qquad (15)$$

Table 9: Coefficients of the linear Logit function and tabulated Chi-square values

	Virus	Insider net abuse	Laptop theft	DoS	Unauthorized access	Financial fraud	Theft of prop info	Telecom fraud
β_0	2.773	2.6104	0.74906	-0.58134	0.68956	-1.8243	-0.72308	-1.7477
β_1	-0.27036	-0.33225	-0.10541	-0.028015	-0.19522	-0.059452	-0.20579	-0.08724
Tabulated Chi square value	14.414	42.941	16.211	19.203	40.831	16.036	23.206	14.159
Deviance	56.78	200.85	20.75	74.48	79.27	25.57	42.04	18.96

Based on the chi-square values, the Logit curves fitted to the data on virus and telecom fraud show a good amount of fitness. The mean values of the other Logit curves also fit well too. Over dispersion in the data points, creates problem in fitting the Logit and Probit curves (McCullagh, et al., 1989).

Our probability based generalized linear model (GLM) such as Logit and Probit, perform much better in comparison to Gomperterz and Logistic curves fitted to the cumulative attack data points, by Bagchi et.al, (2003). The predicted number of attacks and the curve fitting was proper only for a short duration of time (i.e., 12 months only) (Bagchi et.al, 2003).

Proof of concept

From Figure 5a and 5b it is clear that the expected and the observed values of losses for the eight types of e-threats are very close. The expected loss amount has been computed using equation (12). Here, we have used the expected values of the probability computation as obtained from the Logit model, and multiplied it with the total number of respondents interviewed, and the loss amount per attack.

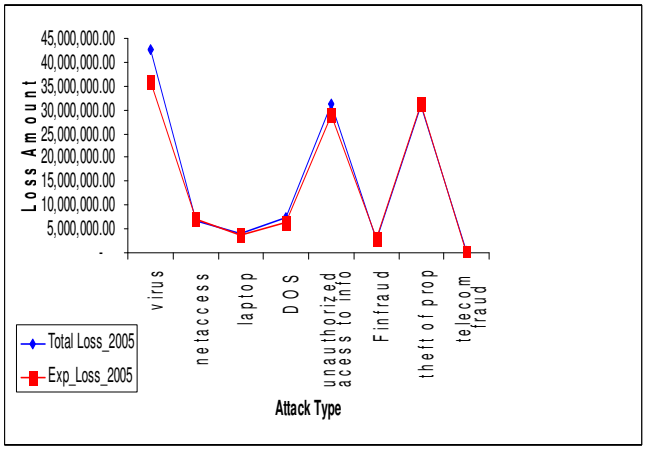

Figure 5a : Expected Loss and Observed Loss during 2005

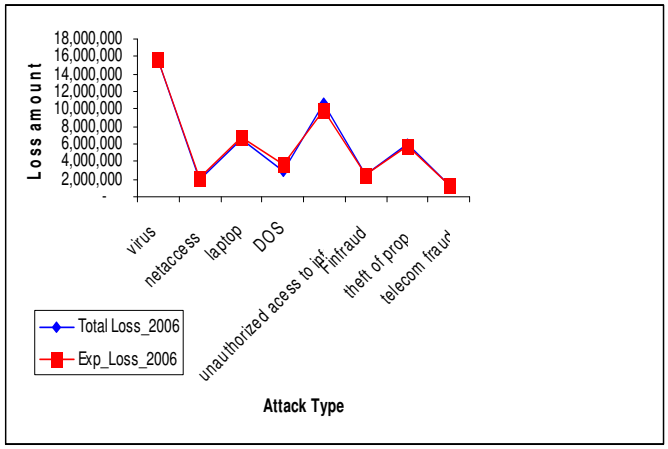

Figure 5b: Expected Loss and Observed Loss during 2006

Premium computation

Organization's can pay a constant premium amount to the insurance company and get indemnified against any losses later. We use equation (14) to compute the total premium that would have accrued to the insurance industry, in 2005, if all the expected attacks actually happened (Figure 6). We assume the overhead to and the contingency loading to be 10% respectively. The final premium fixation would take into account also the amount of risk that is passed to the insurance company and also on the fact, whether the organization is risk averse or a risk taker. The presence of such a large pool of revenue would surely motivate insurance companies to offer e-risk insurance products.

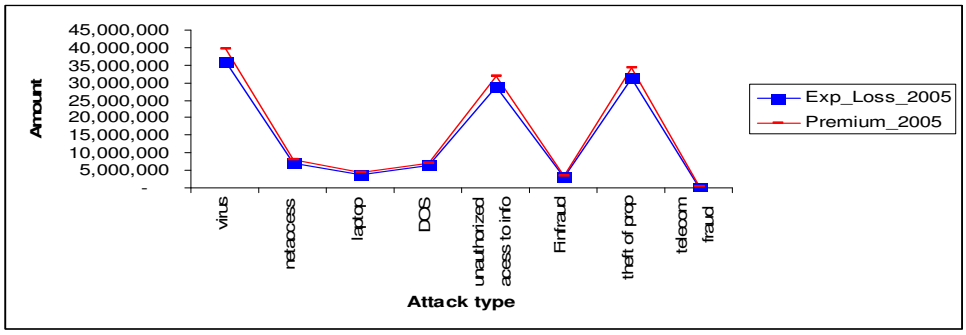

Figure 6: Premium to be paid for loss indemnification

Conclusion

E-risk assessment and quantification, forms an important component of information security management. In this paper, we have developed two probability models (i.e., Logit and Probit) to quantify the frequency of malicious attacks. Our model performs better than the existing one (Bacghi et al., 2003). The work can be further extended by assuming other parameters along with time. The basic assumption of the attacks and their corresponding losses are independent can be relaxed. A copula model, in that context would be another alternative to model the attacks. The sparse CSI/FBI data set is also a major constraint to modeling.

We have also proposed use of insurance for indemnification of the loss. We have even computed the expected loss for each type of attack, suggested the premium to be charged for indemnification of loss. The use of e-risk insurance by organizations will promote ecommerce transactions.

References

1. Anderson, R. Why information security is hard—An economic perspective. In Proceedings of 17th Annual Computer Security Applications Conference (ACSAC), New Orleans, La. Dec.10–14, 2001
2. Baskerville, L. R. "Information Systems Security Design Methods: Implication for Information Systems development", ACM Computing Surveys, vol. 25, no. 4, 1993, pp. 375-414.
3. Bagchi, K., Udo, G. "An Analysis of the growth of the computer and internet security breaches", Communications of the AIS, vol 12, 2003, pp 684-700.
4. Brealey, A. R. Myers, C. S, "Principles of Corporate Finance", McGraw-Hill Higher Education, 2000
5. Brown, E. A., Grant, G. G, "Framing the frameworks: A review of IT Governance Research, Communications of the AIS, vol 15, 2005, pp 696-712.
6. Burger, K. A.The Cost of ID Theft, Part 1 and 2: Beyond Dollars and Cents, E-Commerce Times, http://www.ecommercetimes.com/story/61515.html. 2008, Last accessed 5-6-2008.
7. Campbell, K., Gordon, A. L., Loeb, P. M. "The economic cost of publicly announced information security breaches: empirical evidence from the stock market". Journal of Computer Security, 11, 2003, pp 431-448.
8. Dhillon, G., Backhouse, J. "Information System Security Management in the New Millennium". Communications of the ACM, 43(7), 2000, pp 125-127.
9. Dhillon, G., Gholamreza, T. "Value -focused assessment of information system security in organizations", Information Systems Journal, 2007.
10. Dobson, Annette J., "An Introduction to Generalized Linear Models", 2nd Edition,, Chapman & Hall, 2002.
11. Garretson, C. "Storm: the largest botnet in the world?", Network World, 09/28/2007, http://www.networkworld.com/news/2007/092707-storm-largest-botnet.html
12. Gordon A, L, Loeb, P. M., Sohai, l. T. "A framework for using insurance for cyber-risk management", Communications of the ACM, 46(3), 2003.
13. Hossack B I, Pollard J, Zehnwirth B, "Introduction to Statistics with applications to general insurance , Cambridge University Press",1983.

14. Identity Theft Center. http://www.idtheftcenter.org/. 2007, Last consulted 5-6-2007.
15. Jensen, V.F "Introduction to Bayesian Networks" Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1996
16. Kesan, P. J, Majuca P. R, Yurcik, "The Economic Case for Cyberinsurance**",** Securing privacy in the Internet Age, Standford University Press, 2005
17. McCullagh, P. and J.A Nelder, "Generalized Linear Models", 2nd Edition, Chapman & HaI/~CRC, 1989.
18. Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A, Chakrabarti, B. B., Podder, K. A. "Security breach losses in e-commerce through Insurance". Proceedings of 4th Security Conference, Las Vegas, Nevada, 2005.
19. Miccolis, J. A., "All Together Now", Best's Review (February 2000), pg. 122.
20. Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., Podder, K. A. "e-risk: A case for insurance". Proceedings of the Conference on Information Systems and Technology, New Delhi, India, 2005a.
21. Mukhopadhyay, A., Saha, D., Mahanti, A., Podder, K. A. "Insurance for cyber-risk: A Utility Model". Decision, Vol 32(1), 2005b, pp 153-170.
22. Mukhopadhyay, A, Chatterjee, S., Saha, D., Mahanti, A, Sadhukhan, K. S. "e-Risk Management with Insurance: A framework using Copula aided Bayesian Belief Networks". Proceedings of the Hawaii International Conference on System Sciences 39, Hawaii,USA, 2006.
23. Mukhopadhyay A, Chatterjee S, Saha D, Mahanti A, Roy R, Sadhukhan S K. "Insuring big losses due to security breaches through insurance: A business model", Proceedings of the Hawaii International Conference on System Sciences 40, Hawaii, USA, 2007.
24. Mukhopadhyay A, Chakrabarti B B, Saha D, Mahanti A. "e-risk management through self-insurance: An option model", Proceedings of the Hawaii International Conference on System Sciences 40, Hawaii, USA, 2007a.
25. Mukhopadhyay A. "A novel framework for mitigating e-risk through insurance". Phd Thesis, Indian Institute of Management Calcutta, 2007b.
26. Pogowasright , http://www.pogowasright.org, 2008
27. privacyrights.org, http://www.privacyrights.org/ar/ChronDataBreaches.htm, 2008.
28. Provos, N, McNamee D, Mavrommatis P, Wang K, Modadug N (2007), "The Ghost In The Browser Analysis of Web-based Malware", Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets , Cambridge, MA, 2000.
29. Reid, C. R., Stephen, F. A. "Extending the Risk Analysis model to include market-insurance". Computers & Security, 20(4), 2001, 331-339.
30. Richardson, R. "2007 CSI Computer Crime and Security Survey". San Francisco**:** Computer Security Institute Inc., 2007, pp. 1-28.
31. Austin, D. R, Darby A. R. C, "The Myth of Secure Computing", HBR OnPoint Enhanced Edition, 2003.
32. Schneier, B. "Secrets and Lies: Digital security in a Networked World". John Wiley & Sons, 2000.
33. Schneier, B. "The insurance Takeover". Information Security., 2000a
34. Solms, V. Basie. "Information Security Governance - Compliance management vs operational management". Computers & Security 24(6), 2005, pp-443-447.
35. Straub, W. D., J. Richard, W. "Coping with Systems Risk: Security Planning Models Risk Management Decision-Making". MIS Quarterly, 22(4), 1998, 441-469.
36. Tanenbaum, A. "Computer Networks", Prentice Hall, 1996.
37. Varian R Hal, "Intermediate Microeconomics: A modern approach", 6th edition, W.W. Norton & Company, New York, 2003.
38. Whitman, E. M, Mattford J. M. "Principles of Information Security", Thomson Course Technology, 2007