AMCIS 2003 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2003

An Empirical Study of Computer and Internet Security Breaches Using Sparse Data

Kallol Bagchi
University of Texas at El Paso

Godwin Udo
University of Texas at El Paso

An Empirical Study of Computer and Internet Security Breaches Using Sparse Data

Kallol Bagchi
University of Texas at El Paso
bagchi@utep.edu

Godwin Udo
University of Texas at El Paso
gudo@utep.edu

Abstract

*This study discusses some empirical findings on various computer and Internet-related crimes***.** *An appropriate diffusion model was used that encompassed modeling of two opposite behaviors—imitation and deterrence acts. The model also is adequate for handling sparse data. The model was used to analyze various types of attack. The results shows that different types of attacks follow different types of growth, that increase in imitation is related to an increase in deterrence, and that imitation rates are much higher than deterrence rates. Model validation results are also included.*

Keywords: Security attacks, Gompertz model, sparse data, deterrence, imitation, validation

Introduction

Computer and Internet-related crimes show no signs of abatement. A 2002 survey conducted by the CSI/FBI reports that ninety percent of surveyed large corporations and agencies detected computer security breaches within the last twelve months and eighty percent of them acknowledged financial losses due to computer breaches (Power, 2002). According to a CERT/CC report (2002) computer security vulnerabilities more than doubled in 2001 with 1,090 separate holes reported in 2000, and 2,437 reported in 2001. Following the same trends, the number of reported incidents also drastically increased with 21,756 documented in 2000 and 52,658 in 2002. Through the continual 24 x 7 monitoring of hundreds of Fortune 1000 companies, Riptech has discovered that general Internet attack trends are showing a 64% annual rate of growth (**http://www.riptech.com**).

Newmann (1999) mentions that the costs of cyber crime are difficult to measure, however, they are reasonably substantial and growing exponentially. Other researchers (Lukasik, 2000) claim that costs are essentially doubling each year. The problem gets even more complicated when one considers that these crimes are underreported. Ullman and Ferrera, (1998) mention that according to FBI estimates, only 17 percent of computer crimes are reported to government authorities.

However, not all attacks deserve the same attention and not all attacks may show same type of growth rate. It is important to know how these various crime rates are growing. This needs to be empirically investigated. Although estimation with a sparse set of data at an earlier stage of growth is challenging, it has been proved in past studies to be useful. In this paper, we focus on different types of attacks, how these have evolved, whether different types of attacks have evolved similarly and how deterrence effects are working.

The study is preliminary in nature. The reasons are as follows. Literature is almost non-existent on this topic. Data on different types of security breaches are sparse (Power, 2002). One of the most referred studies of security breaches, the CSI/FBI computer crime and security survey made by Richard Power, contains only a recent few years of data (1996-2002). Modeling such security breaches during the early stages of data availability is difficult but extremely critical. Analysis with sparse data is, however, not uncommon in research literature. For example, marketing literature reports forecasting of sales of new products with as few as five years of data (Mahajan and Peterson, 1985). The dynamic behavior of hundreds of good innovations has shown similar characteristics during the early phases of growth as observed across many types of products (Bass, 1969; Mahajan et al., 1985; Jepson, 1976). Previous research report that the shape of sales curves of many innovative products, during the growth phase is similar (Mahajan et al., 1985). Sales of new products in the early phases tend to grow extremely rapidly. This high growth rate tends to decrease over time and finally the diffusion matures and tapers off, as newer technologies replace older ones. Previous research have also found that in the growth phase, exponential or logistic curves are typically used for modeling purposes, which however, are inadequate to model any innovation at an earlier stage. A small error at this stage can have a large effect on later time period forecasts (Martino, 1972).

Modified Gompertz curves such as the General Sales Growth Curve (Sparse data, 2002) have been reported which describes the data well and yields good curve fitting and forecasting of new innovations in the early growth phases (Jepson, 1976; Lakhani, 1979). The Gompertz curve could be a good fit for innovations which rapidly increases in the beginning and then tapers off slowly. Its point of inflection occurs when at 33% of total potential diffusion. Such a model is used in the present study of bad innovations.

Types of Security Breaches

Some of the important security breaches in 2001 were the results of the following attacks (CERT/CC, 2003): multiple vulnerabilities in the Internet Software Consortium's Berkeley Internet Name Domain [BIND] server, Sadmind/IIS worm (a worm that exploits a vulnerability in Solaris systems and subsequently installs software to attack Microsoft IIS web servers), CodeRed worm (a self-propagating malicious code that exploits IIS-enabled systems), SirCam worm (a malicious code that spreads through email and potentially through unprotected network shares), and Nimda blended threat (a combination of worm, viruses, etc. that propagates itself via several methods, including email, network shares, or through an infected web site). Security breaching techniques have come a long way from early hacker-induced attacks of 1970s. Sophisticated attacks include superior software techniques that are increasingly difficult to separate out from normal network traffic. An automated such sophisticated attack consists of four phases: (1) scanning for potential victims, (2) compromising vulnerable systems, (3) propagating the attack, and (4) coordinated management of attack tools (Householder et al., 2002). To increase attack efficiency, scanning and attack tools are integrated and attack cycles are also automatically initiated. Distributed attack tools are common.

The main types of reported popular Internet-based attacks are denial of service, worm and virus, domain name system (DNS), router attacks and web defacement. The denial of service attack prevents legitimate users in using the service typically by flooding a network or disrupting connections or services. An example of such attack was Mafia Boy attack from February 7-9, 2000 on web sites such as Yahoo.com, CNN.com, and Amazon.com. The web sites went out of service for more than two hours costing $1.2 billion in loss in businesses (CCITS, 2002). A worm is a self-propagating malicious piece of code and is highly automated. An example of a worm is Robert Morris' Internet Worm, which was released on November 2-3, 1988. The worm crashed more than six thousand computers on the Internet. Top-level domain servers are potentially vulnerable and any attack on these can cause widespread problems. There are viruses which can be spread by executing infected programs. An example of a destructive virus is “I love you,” which appeared in May 2000, took five hours to spread and cost some $10 billion in damages and lost productivity (CCITS, 2002). Sometimes when the infected program runs, it may wipe out the hard disk and do other damage. Routers, which are devices used to direct traffic on a network, can be attacked in several ways such as by denial of service or by using the router as an attack platform. It is no surprise that many corporate network professionals cite e-mail parasite (62%) and spam (17%) as the two most damaging types of external security attacks. CSI/FBI report (2002) talk about many other types of computer and Internet-based attacks or misuses such as financial and telecom fraud, telecom eavesdropping, sabotage, laptop misuse, active wiretap, insider abuse of net access, to name a few. We also treat Web defacement as a separate type of attack, because of its importance and recent frequency (more than 50 a day in 2002, **http://www.cnetnews.com**). Reasons for web defacement are similar as other types of hack attacks: electronic graffiti, getting attention, intellectual challenge etc. to name a few. Domains such as .gov, .mil, .com are frequently targeted for web defacement attacks. Mirror web sites such as Alldas.de, attrition.org and safemode.org have chronicled this phenomenon. They have also been closed down by hacker attacks.

Deterrent Methods and Tools

As security breaching techniques got refined, so also was security attack detection and prevention techniques. There are several tools available to firms to combat security attacks. Firewalls can provide ongoing protection to a firm. Firewalls are placed between the company network and the Internet. They can deny suspicious traffic. To inform companies when they are under attack, another system called the intrusion detection system (IDS) is needed. An IDS examines all packets and prepares a log file. The security administrator examines the log file to look for suspicious patterns and generate messages for possible attacks. If an attack packet passes through the firewall, the next line of defense is to prepare the host from possible attacks by installing vendor-specific current patches for known weaknesses in the system. A large number of attacks emerge from known weaknesses in popular software. Security systems are also designed to prevent eavesdropping attacks. Secure communication is ensured when the checks for authenticity, integrity and confidentiality are maintained. Many techniques such as biometrics, digital IDs, encrypted logins, anti-virus software, access control mechanism are additionally used to prevent attacks (Power, 2002). Not all of them are universally effective or popular in use.

Sometimes, an attacker succeeds by breaking all systems. This is called a security incident. Companies need to have good plans for incident handling (also called incident response). This includes stopping the attacks, restoring the system to its pre-attack state, and possibly prosecuting the attacker (or punishing an employee attacker administratively). In case when a firm's security administration fails, Internet security sites as mentioned above can provide help. Organizations such as CSI, CERT, NIPC, IEEE task force (IEEE, 2002) on security and privacy, lend enormous deterring efforts to stop hacking that maliciously damage the academic, government and business activities.

The security infrastructure and security providers no doubt act as a deterrent to attempts of such breaches by sustained organized efforts. Security laws and regulations of a nation additionally aid the deterrent side of the equation. The U.S. government already has some regulations in place. These are related to computers, access devices and communication lines, stations and systems. As an example, the computer fraud and abuse statute 1030 can be cited. The statute states that if anyone knowingly or intentionally accesses a computer without authorization or exceeds authorized access, he/she is liable to be punished (NSI, 2002). International efforts are also not lacking. Forty-one European countries, plus the U.S., Canada and Japan attended a recent convention on cyber crime. These nations signed a treaty that supplies a legal framework aimed at the protection of society against cyber-crime (Conventions, 2002).

A Model of Security Breach/Attacks

Many researchers have studied the Computer/IS security issues (Atkins, 1996; Parker, 1983; Straub, 1990). Straub and his research partners have used general deterrence theory in IS environment (Straub, Carlson, and Jones, 1994; Straub, 1990; Straub and Nance, 1990; Hoffer and Straub, 1989). The basic argument in this body of work is that information security actions can deter potential computer abusers from committing illegal acts. They also have found empirical evidence that security actions can lower systems risk.

However, previous studies lack in empirical results on how different types of attacks grows or providing reliable models of such attack growths. This is important, as some attacks enormously and rapidly disrupt the Internet infrastructure for a length of time, thus resulting in millions of dollars in loss. Take the example of the "Code Red Worm" virus. It infected more than 250,000 systems around the globe in 9 hours on July 19, 2001 and its estimated total global economic impact was $2.6 billion (Householder et al., 2002).

The growth process can be studied from an innovation diffusion perspective (Rogers, 1991). Innovation diffusion literature is usually concerned with good innovations and thus biased towards good innovations. The study of bad innovations such as security attacks can alert readers to the fact that innovations are not always good and what could be done to prevent such bad innovations. This is done in the present study. There are four main elements in the diffusion process: (1) the innovation (good or bad), (2) channels of communication, (3) time, and (4) the social system. In the present case, examples of channels could be direct word-of-mouth or mass media communication channels including the Internet/Web. Time is the rate at which the innovation is diffused and the social system is the system of all potential and existing attackers.

Ideally, a growth model is needed that can capture both deterrence as well as imitation activities to model the security breaching incidents. However, traditional diffusion models do not provide the necessary explanatory power to analyze the attack phenomenon adequately (Mahajan and Peterson, 1985).

The Gompertz Model

The modified Gompertz model used by Pitcher et al. (Pitcher et al., 1978) assumes that the probable causes for the outbreak of such incidents are imitative as well as inhibitive in nature. The imitative aspect is based on incident news spread via the Internet as well as by word-of-mouth; the inhibitive aspects can also be spread via Internet/Web sites and related stories. However, people only engage in security attacks when they feel threatened or are motivated by some economic or other gain and have observed the success of earlier attackers (Bandura, 1986). Traditionally, the challenge or threat to such attackers has been mostly an intellectual one: to break a system. To quote a hacker expert, "It's the sheer challenge (to crack a code or break a system) rather than any (criminal intent). They see it as an intellectual challenge and a prize, (and) they look at the success of what they have done rather than the consequences of the lives of people they have affected" (Dreyfus, 2002). Of course, as mentioned earlier, other types of challenges come, for instance, from making money or taking economic or political advantage. The more successful the earlier attackers are, the more aggressive the behavior of the present attacker becomes. Each such incident is an imitation of previous behavior and a behavioral model for others to imitate. On the other hand, the increase of security activities and success stories about preventing such attacks could reduce the number of attacks. Thus, a combination of imitation and inhibition as assumed by the asymmetric model could provide a realistic background in modeling such incidents.

The model can be expressed mathematically as:

$$\frac{dN(t)}{dt} = c \cdot e^{-qt} \cdot N(t)$$

where, t = time, N (t) = cumulative number of attack incidents at time t and c, q are parameters of the models. The parameter c denotes the net rate of instigation to attacks and q denotes the rate of inhibition in such attacks.

We model the growth process as a combination of such influences as well as preventive efforts by various agencies to curb such incidents. Our analysis suggests that the growth was indeed influenced by a combination of factors: attacks by like-minded peers (hackers or crackers) and attack-preventive measures put forward by various governments, academic and security agencies. The results have implications for everyone - from security professionals and merchants associated with on-line trade over the Internet to academics, professionals and other day-to-day users of the Internet/Web.

Hypothesis Formulation

Although imitative and deterrence acts constitute the background of any attack scenario, the rates of imitation and deterrence may not be same. When the rate of instigation increases it may mean an overall increase in deterrence rate as more and more security products will be developed. As these products come into the market, attackers find ways to bypass these products and refine their attacks, which in turn makes the security products more refined. This cycle of reinforcing each other (i.e., attack and deterrence) continues (Pitcher et al., 1978).

Hypothesis 1. *Relative increase in instigation rate is related to relative increase in deterrence rate.*

Sofaer et al. (2000), observe that "the risks of cyber terrorism and cyber crime vastly outweigh our abilities to control those risks by technological means, although technology can help and should be vigorously pursued." Thus preventive measures are assumed to be thoroughly outweighed by attacks. Therefore it is expected that value of c, the rate of instigation will be much higher than the value of q, the rate of deterrence or inhibition.

Hypothesis 2. *Values of c, the net rate of instigation will be much higher than values of q, the rate of inhibition for computer and Internet-related bad innovations, i.e., digital crimes and security breaches.*

Data

Richard Power (2002) has gathered data on some aspects of cyber crime from 1997 onwards. The survey called "Annual CSI/FBI Computer Crime and Security Survey" gathers data based on survey questionnaire sent to information security practitioners in US organizations. The response rate for the 2002 survey was 14% with 503 practitioners responding. This data set is used for the present study. We used total annual loss data which was available in US dollar value. The data were converted to 1996 US dollar value by dividing by the price deflator for each year.

For the 2002 survey, questionnaires were distributed to 3,500 information security professionals with a 14% response. The responses were anonymous. Job titles of respondents ranged from corporate information security manager and data security officer to senior systems analyst. Organizations surveyed included corporations, financial institutions, government agencies and universities.

For web-defacement, we use monthly defacement incident data from January, 1995-July 2000, based on data gathered by a mirroring firm named Attrition (**http://www.attrition.org/mirror/attrition/stats.html**) Attrition began actively mirroring defaced sites since 1995. However, it had to close down in 2002, due to hacker attacks.

Data Analysis Method

The data analysis method used in this paper is a non-linear least square regression scheme. We used SPSS to design and run the non-linear model described above, with different sets of data. Non-linear equations are sometimes known to be difficult to converge. The convergence problem is handled with suitable initial values of parameters. The Levenberg-Marquardt algorithm (SPSS, 2003) is mostly used to determine parameter values of interest, q and c.

Results

Hypothesis 1. *Relative increase in net instigation rate is related to relative increase in inhibition rate.*

Figure 1 shows the result for hypothesis 1. The figure captures the fit of the power function of the relationship between q and c. The function is: $q = .075c^{(1.95)}$ ($R^2 = .58$). An increase in instigation rate is greater than the corresponding relative increase in inhibition rate. This is consistent with results obtained from other types of crimes (Pitcher et al., 1978). The moderate fit and the positive value of c supports hypothesis 1.

Hypothesis 2. *Values of net instigation rate, c, will be much higher than values of inhibition rate, q, for computer and Internet-related bad innovations, i.e., computer crimes and security breaches.*

Table 1 contains results from running the model for various types of computer crimes and security breaches. The R^2 value from the model fits are very high (.96-.99). The values of q and c are very different, for each type of security breach, with values of c much higher than q. When c > q, it means that overall impact of net instigation is more than inhibition rate and vice versa.

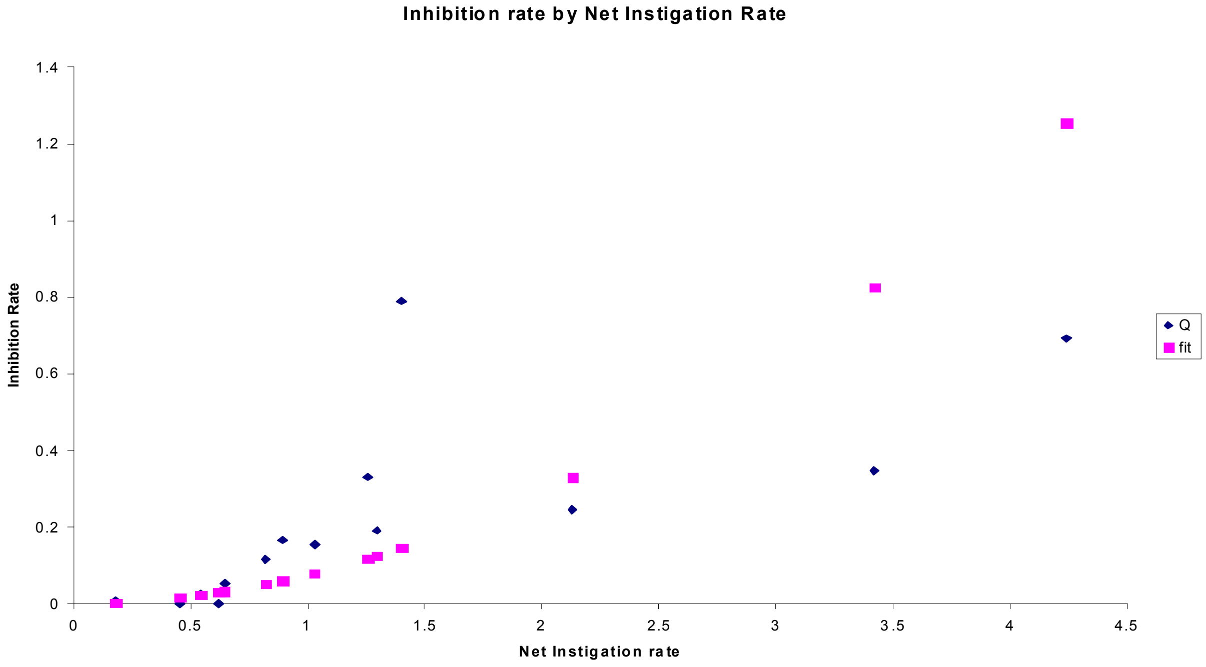

Figure 1. Inhibition Rate (q) by the Net instigation Rate (c)

Table 1. q and c Values from the Model of Various Attack Types

Items	Upper limit of cost	R^2	q	c
Theft of proprietary info.	5.94E+07	0.99	0.1555	1.029
Sabotage of data of networks	2.41E+07	0.95	0.1165	.8211
Telecom eavesdropping	8.12E+10	0.96	0.001	.457
System penetration by outsider	5.31E+12	0.99	0.002	.62
Insider abuse of Net access	8.62E+06	0.99	0.246	2.132
Financial fraud	1.74E+10	0.99	0.054	.644
Denial of service	2.79E+06	0.99	0.191	1.295
Virus	1.85E+14	0.98	0.025	.545
Web defacement	1.096E+11* (*unit is number of incidents)	.98	.0073	.1794

The difference in mean values of q and c is significant (t=-4.19, p<.001), thereby suggesting that q and c are statistically different. The results are again consistent with the results obtained from other types of crimes (Pitcher et al., 1978). This confirms hypothesis 2.

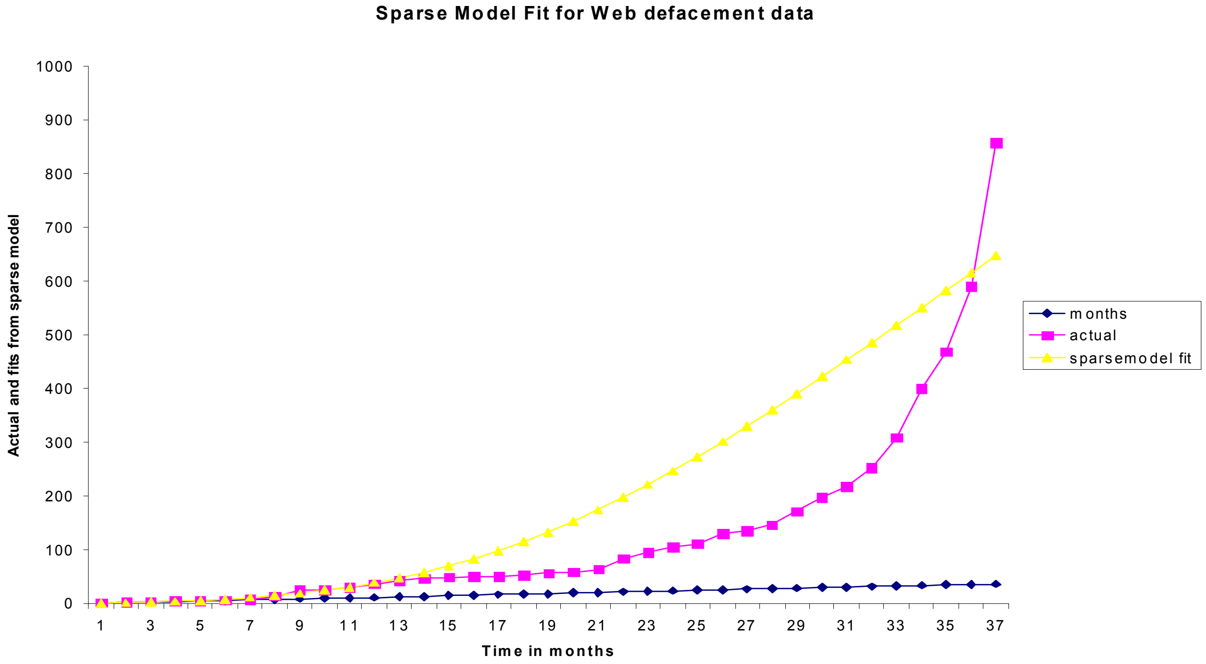

Figure 2. The Web Defacement Data and Predictions

Discussions and Conclusions

How robust is the model fit using the sparse data model? We could check that only for Web defacement case, as it had enough monthly data. Monthly data are more susceptible to fluctuations than yearly data. We used 10 data points for building the model and 10 successive data points for prediction, using the Gompertz model. Predictions up to 27 additional months are reasonable, after which the estimates deteriorate. For the first 10 forecasts, predicted value exceeded three times 100% of the actual value. Thus 30% of forecasts were off by more than 100%. The average error for the 10-month forecast was 66%. For the entire 27 months, predictions exceeding 100% greater than actual values were 14 in number, or roughly half the time as shown in figure 2.

The exponential or the logistic fits would likely to perform much worse, when compared with this model. Thus this model can act as a reasonable tool for short-term predictions.

Of the two hypotheses explored in this study, both were strongly confirmed while the remaining two were partially confirmed. In summary, the results of this study have led us to conclude that:

- Relative increase in net instigation rate is related to relative increase in inhibition rate which implies that the increasing attack incidences will force organizations and governments to come up with means of preventing or reducing them;
- For computer and Internet-related attacks (bad innovations), the values of net instigation rate is higher than values of inhibition rate which implies that much more efforts and resources need to be applied toward inhibiting attacks; and
- Different computer crimes and security breaches grow at different rates, which implies that these crimes should not receive the same level of attention because some crimes are likely to spread more rapidly than other.

This article is a first attempt to identify the nature of growth of various computer and Internet-related crimes, using a sparse set of data. First, a model was selected for bad innovation modeling which can represent both imitative and inhibitive behaviors in attacks. Next, the model was used to predict and compare various types of attacks with a sparse set of data. Although the model used is an ideal one for such purposes, it may still yield forecast errors that can only be refined with the progress of time. So the results should be interpreted with caution. However, our objective is to obtain and compare preliminary growth estimates of various attacks and this paper does indicate how different such crimes are growing.

Acknowledgement

The authors wish to acknowledge the helpful comments of the reviewers which improved the quality of the manuscript.

References

Atkins, D., *Internet Security Professional Reference*, Indianapolis, IN: New Riders Pub., 1996.
Bass, F.M., "A New Product Growth Model for Consumer Durables," *Management Science* 15, 1969, pp. 215-227.
Bandura, A., *Social Foundations of Thought and Action*, Englewood Cliffs, NJ: Prentice-Hall, 1986.
A CCITS/Infosech Presentation on Internet Security, 2002.
CERT/CC Web Site, **www.cert.org**.
Convention on Cybercrime **http://conventions.coe.int/Treaty/en/Treaties/Html/185.htm**, 2002.
Dreyfus, S., "Cracking the Hackers' Code," August 2002, **https://f2.com.au/**.
Ford, R., "No Surprises in Melissa Land," *Computers and Security*, 1999, Vol. 18, pp. 300-302.
Hackers: a Canadian Police Perspective, **http://www.rcmp-grc.gc.ca/crim_int/hackers_e.htm**, 2002.
Householder, A., Houle, K., and Dougherty, C. "Computer Attack Trends Challenge Internet Security, Security and Privacy", Supplement to Computer, IEEE Computer Society, 2002.
Jepson, C., *E. I. DuPont de Nemours & Co., Inc, Internal Presentation*, 1976.
Katz, M., and Shapiro, C., Technology Adoption in the Presence of Network Externalities, *Journal of Political Economy* 94, 1986 pp. 822-841.
Lakhani, H., "Empirical Implications of Mathematical Functions Used to Analyze Market Penetration of New Products", *Technological Forecasting and Social Change*, 15, 1979.
Lukasik, S. J., "Protecting the global information commons," *Telecommunication Policy*, 24 (6-7), pp.519-531.
Mahajan, V., Muller, E., and Bass, F. M. "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing*, 54, 1990, pp. 1-26.
Mahajan, V., and Peterson, R. "Models for Innovation Diffusion", Sage University Paper series on Quantitative Applications in the Social Sciences, (2nd Ed.), Beverly Hills and London: SAGE Publications, Sciences, (2nd ed.). Beverly Hills and London: SAGE Publications, 1987.
Martino, J. P., "The Effect of Errors in Estimating the Upper Limit of a Growth Curve," *Technological Forecasting and Social Change*, 4 (1972) pp. 77-84.
Newmann, P. Information System Adversities and Risks, **http://www.oas.org/juridico/english/information_system_adversities_a.htm**

Computer Fraud and Abuse Act (18 USC 1030), 1986. Fraud and related activity in connection with computers, Title 18, Section 1030, http://nsi.org/Library/Compsec/cfa.txt, 2002.

Parker, D.B., *Fighting Computer Crime*, New York, Scribner's, 1983.

Pitcher, B., Hamblin, R., and Miller, J. "The Diffusion of Collective Violence," *American Sociological Review*, Vol. 43, February 1978, pp.23-35.

Power, R., CSI/FBI Computer Crime and Security Survey, *Computer Security Issues and Trends*, Vol.8, No. 1, 2002, pp. 1-22.

Rogers, E., *The Diffusion of Innovation*, New York: Free Press, 1991.

Sofaer, A.D., Cuellar, M., et al., (Eds), *The Transnational Dimension of Cyber Crime and Terrorism,* Hoover National Security Forum Series, 2000.

Forecasting with Sparse Data: Applying the *General Sales Growth Curves*^{SM,} http://www.lieb.com/NEWS11/forecast.htm, SPSS 10 Syntax Reference Guide, 2003.

Straub, D.W., "Effective IS Security: An Empirical Study'', *Information Systems Research*, Vol. 1, Sept. 1990, pp. 255-276.

Straub, D.W.and Welke, R. "Coping with Systems Risk: Security Planning Models for Management Decision-Making" MISQ April 1999.

Ullman, R., and Ferrera, D. "Crime on the Internet," *Boston Bar Journal*, Nov./Dec., n.6., 1998.