# MAX PLANCK INSTITUTE FOR FOREIGN AND INTERNATIONAL CRIMINAL LAW

# ANALYSIS AND MODELLING OF CYBERCRIME: PROSPECTS AND POTENTIAL

SOUMYO D. MOITRA[1]

**Contents:**



[1] Professor of Operations Management, Indian Institute of Management Calcutta, Diamond Harbour Road, Joka Kolkata 700 104. India. This work was done when the author was at the Max Planck Institute for Foreign and International Criminal Law in Freiburg, Germany between April and June 2003 with financial support from the Max Planck Institute. Encouraged for this project from Prof. Hans-Joerg Albrecht, Director of the Institute is also greatfully acknowledged.

## Introduction

The Internet is now regarded as one of the most significant developments in information technology in recent times. Its phenomenal growth, diversity of uses and its applications for information dissemination have been widely reported in the popular press, in journal articles and in books. Thus the terms cyberspace and even cyber-society have become ubiquitous[2]. The uses of the Internet include communications, data collection and of course e-commerce. Not only has the number of people who get "on-line" increased dramatically, the use of the Internet by businesses has also grown significantly in importance. For instance, the number of people online in 2003 has been estimated to be between 450 million to 620 million. In addition to the large numbers, the growth rate has been phenomenal, compared to other technologies, as can be seen from the fact that it was 369 million on 2000 and is projected to be between 700 million and 940 million in 2004[3]. Internet penetration in Europe is now estimated to be over 40%, up from 18% in 2000. Similarly, the number of hosts on the Internet is estimated to be around 172 million in 2003, up from about 9.5 million in 1996[4].

Altogether, this growth and widespread use of networks has generated a dependence on them of many elements of society, both at the individual as well as organisational level. Individuals are increasingly using computer networks to gather information and doing personal business (such as banking), and businesses routinely exchange data and carry out transactions over such networks. In Europe alone, the number of company sites is expected to increase from about 500,000 in 2000 to an estimated 8 million by 2003, and e-commerce will account for approximately 25% of total European business revenues. Worldwide, there are now well over 3.2 million web servers from a few thousand in 1994. The number of websites has been estimated to be around 28 million in 2002 and about 16% of them are capable of conducting e-commerce. Thus the Internet has clearly become

---

[2] In addition, there has been a dramatic growth in various types of computer networks or "networked information systems" such as LANs, WANs, intranets and Virtual Private Networks (VPNs). Many of these are also connected to the Internet.

[3] The sources for these data are eMarketeer.com, Computer Industry Almanac, Global Reach, and Internetstats.

[4] Netwatch. The breakdown by country indicates US having 75 million, Japan with 11 million, Germany 6 million and the UK 4 million.

indispensable to individuals, corporations, research and educational organisations and also to governments. In fact the Internet is considered to be an important part of the National Infrastructure in many countries. For example, The European Union began its "eEurope Initiative" in 1999 to "ensure that Europe can reap the benefits of the digital technologies."

Along with this growth in scale, usage and dependence, many risks associated with these information networks, and in particular the Internet, have emerged. In addition to various technological and socio-cultural issues that have arisen, cyberspace has opened up a new arena for criminal activities. This is the result of another characteristic of the Internet, namely, that it is a very open network and it is very easy to access the computer systems on it. Thus almost anybody can get on it, including malicious hackers. Consequently, any system connected to the Internet (even indirectly) is potentially vulnerable to possible unauthorized intrusions, attacks and cybercrimes. Indeed, these intrusions and attacks, which we shall include in "cybercrimes", constitute a significant threat to the functioning of the Internet, and in particular, have been a major inhibitor of the growth of e-commerce.

Indeed, most indications point to substantial illegal/criminal activity over the information networks including many *new kinds* of crime and deviance, such as virus attacks that destroy computer files and systems, and distributed denial-of-service (DDoS) attacks that can paralyze communication links to an organisation's information system. Anti-terrorism legislation after 9/11 in many countries have upgraded certain computer-related crimes to terrorist crimes. Although some cybercrimes may have some parallels with traditional white-collar crimes or trespassing, for example, computer fraud (a typical white-collar crime) or unauthorized access (electronic trespassing), even these cybercrimes have a number of distinguishing features that make them qualitatively different from traditional crimes and hence they also need to be studied along with the new cybercrimes[5].

Computer crimes and security breaches of information systems have now been well documented[6], and it is clear that they present a threat that is seri-

---

[5] Clifford, R.D. (Ed.) 2001 *Cybercrime: The Investigation, Prosecution and Defense of a Computer-related Crime*. Carolina Academic Press, Durham, NC; Denning, D. and Denning, P. *Internet Besieged: Countering Cyberspace Surfflaws*. Addison-Wesley, 1998; Furnell, S. *Cybercrime: Vandalizing the Information Society*. Addison-Wesley, 2002.

[6] See AusCERT 2003: *The 2003 Australian Computer Crime and Security Survey*; Computer Security Institute 2002: *The 2002 CSI/FBI Computer Crime and Security Survey*; Furnell, 2002 op. cit.; Hollinger, 1997 *Crime, Deviance and the Computer*,

ous enough to warrant further study and preventive measures. The threat is serious in many ways. Increasingly, societies are becoming information-based and any threat to safe communication endangers the functioning of society. According to one scholar, "As never before, industrial societies are dependent upon complex data processing and telecommunications systems. Damage to, or interference with, any of these systems can lead to catastrophic consequences."[7] The Commission of the European Communities also notes that "information and communication infrastructures have their own vulnerabilities and offer new opportunities for criminal conduct" and that "these offences constitute a threat to industry investment, and assets, and to safety and confidence in the information society."[8]

Further, many businesses are depending on "electronic-commerce" to become more efficient and competitive, and cybercrime would seriously inhibit the growth of "electronic-commerce" and which in turn will deprive us all of its benefits. Finally national defense systems are also dependent on computer networks and hence threats to the networks are a matter of national security to most countries including those in Europe.

As a result, cybercrime has attracted the attention of most governments, policy makers and several supranational organizations. The US was one of the first countries to enact a comprehensive set of computer crime laws both at the federal and state levels. The Council of Europe has responded to threats of cybercrime with a comprehensive Convention on Cybercrime[9]. The convention covers the types of activities that fall under the concept of cybercrimes[10] and goes into the particulars of criminal investigation and criminal procedure in case of cybercrimes. One point that is becoming clear is that successful enforcement of cybercrime laws will depend on striking an adequate balance between confidentiality and security of data transmit-

---

Dartmouth, Aldershot, U.K.; Parker, D. 1998 *Fighting Computer Crime: A New Framework for Protecting Information*. New York, John Wiley & Sons, New York; Sieber, U. (Ed.) *Information Technology Crime: National Legislations and International Initiatives*. Carl Heymanns Verlag, 1994.

7 Grabosky, P. 2000 *Computer Crime: Challenge to Law Enforcement*. Law Enforcement Review, N100, 31- 38. Grabosky, P. 2000 *Computer Crime: Challenge to Law Enforcement*. Law Enforcement Review, N100, 31- 38

8 Commission of the European Communities: *Creating a Safer Information Society by Improving the Security of Information Infrastructures and Combating Computer-related Crime*. Brussels, 2001.

9 Dated 23. 11. 2001.

10 Section 1, Title 1 – Title 4.

ted through computer networks on the one hand and access to such data by police and public prosecution on the other. This is linked to obligations of service and access providers to store traffic (and partially content) data for specified periods of time. The last years have seen in the European Union member countries significant legislative activities in this field[11]. Clarification of such policy issues requires further study of the advantages and disadvantages of alternative rules on network data retention and access. Such a study has to be based on the kind of research outlined in this paper.

While there is already a substantial volume of literature on this topic[12], much research still needs to be done to get a comprehensive picture of cybercrime, and in particular how we can effectively and efficiently control it. There has hardly been any systematic empirical or analytical work involving extended data analysis or modelling although there is now data available on reported cybercrimes and also some data from surveys. Nor has there been any systematic investigation of the existing data sources to assess their validity and usefulness for analysis, or to develop decision-making models for policy analysis. Unless we specifically analyze and model cybercrime for this purpose, we shall not be able to get a sufficiently in-depth understanding of the phenomena that is needed for developing effective policies to counter cybercrime.

Further, it has become vital to disaggregate the rhetoric from the reality[13]. On one hand, the media has sensationalized cybercrime (including making unsubstantiated references to "cyber-war", "cyber-terror", etc.)[14]. Commercial security firms have also tended to exaggerate the number of attacks and degree of damage done to garner more business[15]. Thus the sense of danger has been unduly magnified in the public consciousness. On the other hand there are indeed a significant number of network attacks and

---

[11] Note of the Council of the European Union to the Multidisciplinary Group on Organised Crime, Brüssel, 20. November 2002.

[12] For example Cohen, F. *Protection and Security on the Information Superhighway*. Wiley, New York,1995; Denning and Denning 1998 op.cit.; Furnell 2002 op. cit.; Grabosky, P., Smith, R.G. and Demsey, G. 2001 *Electronic Theft: unlawful Acquisition in Cyberspace*. Cambridge University Press; Hollinger, 1997 op.cit.; Parker, 1998 op.cit.; Sieber 1994 op.cit.; Wall, D.S. (Ed.) 2003, *Cyberspace Crime*. Ashgate/Dartmouth; Wasik, M. 1991 *Crime and the Computer*. Clarendon Press, Oxford; among many others to be discussed later.

[13] Wall 2003 op.cit.

[14] Smith, G. *An Electronic Pearl Habor? Not Likely*, in Wall 2003 op. cit.

[15] These same security firms often try to instill a false sense of security by suggesting to potential customers that their products will remove the threat of cybercrime.

cybercrimes that are committed by various malicious hackers. Given the low detection and reporting rates, the actual prevalence is probably high enough to warrant some concerted response by society although the measurable damage (in many cases) may in fact be quite small (or at least, smaller than is reported). Therefore, it is critical that we have more and better information and analyses to arrive at a balanced view of cybercrime and an equitable, effective and efficient policy to control it.

Thus, in addition to other activities that are being carried out, we need an empirical approach to studying cybercrime that would include estimating the prevalence of cybercrime as accurately as possible, analyzing the patterns of cybercrime and the victimization resulting from cybercrime and developing models to explore the impacts of criminal justice policies for the prevention and control of cybercrime. A comprehensive study of this problem requires above all a multidisciplinary approach: one that includes knowledge of computer networks and telecommunications; an understanding of information systems and how systems are attacked; a criminal justice perspective; and an analytic approach to model the process, investigate the impacts of cybercrime and assess the effectiveness of preventive measures. It is this kind of multidisciplinary approach that we adopt in this paper.

## 1. Research Issues

The Internet has understandably attracted immense attention and much has been written on the subject already and from a variety of perspectives. Here we limit ourselves to noting the main themes in the literature on cybercrime and then we shall focus on the analytical issues in understanding cybercrime from an empirical and criminological perspective[16].

Much of the current literature has been concerned with the legal issues of computer crime and cybercrime with the focus on the laws related to computer crime (computer crime subsuming cybercrime) and on specific cases

[16] The interested reader will find many further references in the works cited in this paper.

rather than empirical analysis[17]. Also, articles on computer law and cyber law are continually appearing in numerous legal journals and it would be impossible to cite them all here[18]. A fairly large body of sociological literature has discussed the novelty of cybercrime with respect to traditional crime, 'identity' in cyberspace and the nature of cyberspace communities. There is also a related literature that is concerned with the impact of the Internet on cultural matters and as a communications medium, but not much has been written about how to analyze cybercrime data or model cybercrime for the purpose of developing crime control policies[19].

A substantial literature now exists in computer science on the network security issues, vulnerability of computer systems to attacks and the nature of intrusions and attacks on systems. However, the issue there is the defense of systems against attacks and not on the criminological nature of the illegal or unauthorized activity[20]. Finally, the matter of cybercrime is wide-

---

[17] Hollinger, 1997 op. cit. has an extremely useful collection of readings: for example, Michalowski, R.J.; Pfuhl, E.H. - *Technology, property and law;* and Forscht, K.; Thomas, D.; Wigginton, K. - *Computer Crime: Assessing the Lawyer's Perspective.* Legal and governance issues are discussed in Kahin, B. and Keller, J.H. 1997, *Coordinating the Internet,* The MIT Press, Cambridge, MA, and Kahin, B.; Nesson, C. 1998. *Borders in Cyberspace*. The MIT Press, Cambridge, MA. Additional discussion can be found in Edwards, L. and Waelde, C. (Eds.) 1997, *Law and the Internet*, Hart, Oxford; Akdeniz, Y., Walker, C.P. and Wall, D.S. 2000, *The Internet, Law and Society*, Longmans, London; Wall, 2003 op. cit.

[18] See in particular Criminal Law Review 1998 (Special edition) and the International Yearbook of Law, Computers and Technology. Journals that regularly publish such articles include: Harvard Journal of Law and Technology, International Journal of Law and Information Technology, International Review of Law Computers and Technology, Journal of Information, Law and Technology, Rutgers Computer and Technology Law Journal, and Santa Clara Computer and High Technology Law Journal.

[19] Among the many books in this area, a few of the salient ones are: Bell, D. and Kennedy, B.M. 2000, *The Cybercultures Reader*, Routledge, London; Castells, M. 1996, *The Rise of the Network Society*, Blackwell, London; Slevin, J. 2000 *The Internet and Society*, Routledge, London; Taylor, P.A. 2001 *Hackers: Crime in the Digital Sublime*, Routledge, London; Turkle, S. 1997, *Life on the Screen: Identity in the Age of the Internet*, Simon & Schuster; Wallace, P. *The Psychology of the Internet*, 1999, Cambridge University Press, Cambridge; Wallace, J. and Mangan, M. 1996 – *Sex, Lies and Cyberspace*, Henry Holt, NY; additional references may be found in Wall 2003, op. cit.

[20] The interested reader may consult the following selection of books and papers in this area: Cohen 1995 op. cit.; Denning and Denning 1998 op. cit.; Doll, M.W., Rai, S. and Granado, J. 2002, Defending the Digital Frontier: A Security Agenda, John Wiley and Sons, New York; Ellison, R.J., et al. 1997, *Survivable Network Systems:*

ly discussed in the e-commerce/e-business literature, where it is taken as a grave threat to the potential of doing business over the Internet. Given the enormous potential of e-commerce, the inhibitory impact of cybercrime (if not controlled) is indeed a major economic issue[21]. The International Chamber of Commerce, for example, notes in a report that "Cybercrime threatens the "brave new world" of e-commerce" and it "is casting a large shadow over an otherwise remarkably positive development." Another writer comments "the security and privacy risks still present major stumbling blocks to realizing the twenty-first century vision of a robust digital economy."[22]

There are some unique aspects of cybercrime that make it a complex topic to study. Firstly, it is a relatively new phenomenon for law enforcement and many members of the criminal justice system are still unfamiliar with it. Secondly, the definitions of cybercrime are not always clear or uniform across countries. Thirdly, cybercrime is characterized by a remoteness between the perpetrator and the victim that can be quite extreme and transnational. Fourthly, the probabilities of detection and reporting are far lower than most traditional crimes. Fifthly, the nature of the evidence is very different from traditional crimes, and finally, we have not yet had the time or experience to properly comprehend the nature and implications of cybercrime.

In this paper we use the term *site* to denote a set of inter-connected computer systems belonging to the same organization at one physical location. As is frequently the case in the literature, we use the terms *computer*, *computer system*, *information system* and *system* interchangeably. We also use *attacks*, *incidents*, *intrusions* and *crime* as similar terms, although there are differences: an *attack* is a malicious, undesired action committed on a computer system; an *incident* may comprise one or more attacks that are related to each other and generally includes some additional reference to

*An Emerging Discipline.* CMU/SEI-97-TR-013; Moitra, S.D. and Konda, S. 2000a, *A Simulation Model for Managing Survivability of Networked Information Systems*. SEI/CERT Report – CMU/SEI-2000-TR-020, Carnegie-Mellon University; Northcutt, S. and Novak, J. 2002, *Network Intrusion Detection*, Que Publications; Parker 1998 op. cit. Additional references can be found in these books.

[21] Boni, W.C. and Kovacich, G.L. 1999, *I-Way Robbery: Crime on the Internet*, Butterworth-Heinemann, London; Garfinkel, S. and Spafford, G. 1997 *Web Security and Commerce*, O'Reilly & Associates; Quarantiello, L.E. 1997, *Cyber Crime: How to protect yourself from Computer Criminals*. Limelight Books, Lake Geneva.

[22] Ghosh, A.K. 2001 *Security and Privacy for E-Business*, John Wiley and Sons, NY.

the impact on the victim system; an *intrusion* is any unwanted set of messages or any unwanted attempt to interfere with a computer system through a network; a *crime* is some illegal activity that is committed on a computer and/or the data stored in it. These are all approximate terms since no consensus has yet emerged on a taxonomy for cybercrime.

The criminological approach is relatively less well represented in studies of cybercrime[23]. However, the nature of cybercrime, the problem of classification of the different types of crime that comprise "cybercrimes" and the social control of cybercrime has engaged a number of researchers[24]. However, as in the case of other disciplines, the current criminological literature has not delved into the quantitative or analytical issues related to cybercrime. There are many modelling approaches in criminal justice, but they have hardly been applied to cybercrime[25]. Further, cybercrime may need different models that would capture its uniqueness. In any case, to develop useful models that can be estimated and validated, we first need to assess the available data, to be sure that the data needed for the models are available in the first place.

While developing the data collection methodology, we need to concurrently consider the different modeling approaches that would be possible and useful, and then select those models that can realistically be developed, estimated, validated and applied given the available and potentially available data. Thus the investigation of data on cybercrime goes hand in hand with the exploration of potential models. The approach involves an iterative

---

[23] This is noted by Wall, 2003 op. cit.

[24] Grabosky, P. 2001 *Computer Crime: A Criminological Overview*. Forum on Crime and Society. 1, 1, 35-53; Grabosky, P. and Smith, R.G. 1998 *Crime in the Digital Age: Controlling communications and cyberspace illegalities*. Federation Press; Mann, D. and Mike Sutton (1998) *NETCRIME: More Change in the Organization of Thieving*. British Journal of Criminology, 38, 2, 201-229; Speer, D.L. 2000. *Redefining borders: The challenge of cyber crime*. *Crime, Law and Social Change* 34, 259-273; Thomas, D. and Loader, B. D. 2000, *Cybercrime: Law enforecement, security and surveillance in the information age*. Routledge, London; Wall, D.S. 1998. *Catching Cybercriminals: Policing the Internet*. International Review of Law Computers. 12,2. 201-218; and Wall, D.S. (Ed.) 2001 *Crime and the Internet*. Routledge, London.

[25] Models of criminal careers are especially relevant here. For example: Barnett, A., Blumstein, A. and Farrington, D.P. *A Prospective Test of a Criminal Career Model*, Criminology, 27 (2) 373-388, 1989; Cohen, J. *Criminal Career Research: Its Value for Criminology*, Criminology, 26, 1988; Greenberg, D. *Modeling Criminal Careers*. Criminology, 29(1), 17-46, 1991.

approach that alternates between identifying data requirements for modelling on one hand and modelling based on the available data on the other hand. It is possible that currently there are not sufficient data to estimate some models that might otherwise be very useful. If it is deemed possible to collect the needed data, further surveys should be designed, fielded and the resulting data analyzed. Thus we first have to assess currently available data and then consider the design of new surveys. It could also be the case that reliable and generalizable data that is necessary for a comprehensive picture of cybercrime just cannot be obtained. In such cases, we could investigate possible simulation models that can be run based on reasonable assumptions[26]. The data obtained from the various sources can be used as inputs after proper validation and accounting for biases. Doing sensitivity analysis with the simulation model can fill gaps in the data. However, research in these directions does not seem to have been undertaken so far, and it is these issues that we discuss in this paper.

In this paper we focus on cybercrime only. By cybercrime, we mean illegal or deviant activity on the Internet involving both a computer (or computers) as the tool to commit the action *and* a computer (or computers) as the target of that action. Thus we can refer to these as "C2C" incidents. Further, we restrict ourselves to empirical and modeling issues in cybercrime. The paper is in three parts: A) review of current data, B) development of data collection instruments and C) modeling cybercrime through simulation. We briefly introduce each of these areas next.

## *A. Assessment of available data on the prevalence of cybercrime*

There are a growing number of apparent sources of data on cybercrime as any search on the Internet will reveal. However, there is really no systematic catalog of these databases on cybercrime. While some search results have links to websites which purportedly have data on cybercrime, a visit to those sites have revealed that there is actually almost no data that could be used for developing models[27]. Therefore we have searched for major data

[26] Moitra and Konda, 2000a op. cit.

[27] Moitra, Soumyo D. 2002a. *Internet Crime: Towards an Assessment of its Nature and Impact*. (Under revision for the International Journal of Comparative and Applied Criminal Justice).

sources and have started a "metadatabase" of those sources that appear to have valid data. While this is a huge task because virtually nothing has been done yet, it is important to make a start in order to have a systematic idea about what data are available. It is expected that this will be of help to future researchers who wish to have an idea of what websites actually have some data and which do not.

In addition to knowing what is available, we need to know as much as possible about the details of the databases if we are to assess their possible uses and applications. That is, we need to assess their potential for research and how relevant they are for estimating various statistical measures and models. For example, we would like to know if the data includes reports of cybercrimes by type of crime, whether the victim system is identified or whether there is any information on the perpetrator, and so on[28]. We also need to assess the overall quality, potential biases, and what inferences can be legitimately drawn from them and what cannot. This is an attempt to link the available data to potentially useful analytic models that can be estimated with them. Very little of such assessment of Internet data has been done to date, but it is important for distilling information that could be useful for Internet policy. Again, while such a task can never be completely done, the exercise is important given the paucity of useful cybercrime data and we hope it will benefit future studies of cybercrime. Section 3 reports on our initial attempt at this.

## *B. Data collection for the study of cybercrime*

As will be clear later in Section 3, there is an outstanding need for further, systematic data collection on cybercrime in order to answer many outstanding questions of interest, to estimate statistical models of prevalence of cybercrime and to understand the resulting victimization of information systems, individuals and organizations. This would open up a new victimological perspective in that cybervictims have not been studied very much and there are many new aspects to victimization through cybercrime, such as the remoteness between the perpetrator and the victim, the fear of crime in cyberspace and the fact that actually there are two dimensions to the vic-

[28] Moitra, S.D. and Konda, S. *The Survivability of Network Systems: An Empirical Analysis*. SEI/CERT Report – CMU/SEI-2000-TR-021, Carnegie-Mellon University, 2000b.

timization: the damage done to information and the (sometimes intangible) damage done to the victim. Victims could be individuals or organizations and they will have different behaviours, resources, attitudes and responses. All these aspects will be addressed by the surveys we have proposed in Section 4. It is important that these surveys be methodologically rigorous, and for this reason we have based them on past, established surveys and survey methodologies while at the same time extending the previous cybercrime surveys to include items of information that are of interest to empirical criminology and policy analysis. At the same time they have to be practical instruments that are not excessively long and that can be replicated across countries and over time.

In addition, we shall also discuss and develop methods for recording reported data. This relates to reports from computer sites that have experienced intrusions or crimes over the Internet. With well-designed instruments, such data can also provide useful information on cybercrime such as trends. However, analyses of survey and reported data will necessarily give only a piecemeal view of deviance in cyber space. It would be unrealistic to believe that even the best surveys that could be *practically* undertaken would provide us with sufficient data of satisfactory quality that would answer all our questions about cybercrime. In order to construct a more comprehensive picture and to fill in the gaps in existing data, we propose developing a simulation model of the cybercrime process.

## *C. Development of simulation models for cybercrime*

There are several advantages in using simulation models[29]. For example, they can be very useful in exploring cases where there is insufficient data by running different scenarios and observing the results. By considering a range of parameter values, we can partially overcome some of the data limitations. Provided it can be reasonably validated, the results of the simulation model can provide insights into the cybercrime process, estimates of the prevalence and patterns of cybercrime, and impacts of changing poli-

[29] Law, A.M. and Kelton, W.D. 1999 *Simulation Modeling and Analysis*, McGraw-Hill, New York.

cies[30]. We therefore propose a framework for a simulation model that utilizes available data while overcoming current data limitations to the extent possible and, at the same time, can yield insights into the implications of alternative policies. A simulation model would be particularly appropriate for such analysis at this stage of our knowledge because we can use the outputs of the model to check the validity of our assumptions and see if the model corresponds to what we know about cybercrime. With a reasonably validated simulation model, we can then estimate the impact that can be expected when there is a change in some policy or other variable. Thus we can use the model to explore a variety of questions and even test for various biases in the data we have on cybercrime. A preliminary model has been developed[31] but that needs to be extended to include more details of the cybercrime generation process. Such a model will allow us to move from a reactive approach to cybercrime to a proactive approach. The simulation model will complement the knowledge we derive from surveys and reported data and help us to construct a more accurate picture of crime in cyberspace. As there will always be uncertainties and lack of data about cybercrime, simulation models can guide us in overcoming these gaps in the data and in arriving at a reasonable view of how cybercrime might evolve under different policies.

## 2. Prevalence of Cybercrime

### *A. Data sources on cybercrime*

The Internet itself is a useful (though limited) source of information on such data sets. Many agencies, institutes and organizations have posted data or information on cybercrime on the Internet websites. However, there are limitations to these sources. One is that very often they have no actual data, even though they have so advertised themselves, and secondly, very rarely are there original data from surveys or the actual data that has been reported. Usually when there is any useful information, it is a summary of

[30] Caulkins, J.P., Crawford, G. and Reuter, P. 1993 Simulation of Adaptive Response: A Model of Drug Interdiction, Mathematical and Computer Modelling, 17 (2), 37-52; Cohen, F. *Simulating Cyber Attacks, Defenses, and Consequences*. Fred Cohen & Associates. 1997.

[31] Moitra and Konda, 2000a op. cit.

the data or sometimes simply some discussion on the data. However, it is possible that such organizations that have data might be contacted for additional data that may not be on websites, especially the actual or raw data. The data may be available for research purposes on certain guarantees of anonymity and confidentiality. This is an important area for further research. The published literature in journals was also searched, and some papers reporting survey results on issues related to cybercrime were found[32]. However, very few of the surveys were designed to elicit data for the purposes of policy analysis, and none of them were designed to understand individual cybervictimization.

There are two broad data collection methods in the study of cybercrime. One is through surveys where a sample of potential respondents is selected and the received responses are analyzed. The other method is to collect data that is voluntarily reported to some authorized agency. Even here we can distinguish between data reported to agencies such as the CERTs which are mostly on intrusions from the Internet that have been detected by a system or site, and the data reported to agencies such as the IFCC which get reports mostly from individuals on alleged cybercrimes such as fraud in an Internet transaction.

## *B. Assessment of some available data*

An initial template for a metadatabase on cybercrime is given in Table 1 in the appendix. This needs to be modified and extended in the future. In this section we shall discuss what useful information that can be derived from the data cited in the metadatabase and the extent to which they can help with understanding the prevalence and patterns of cybercrime and with modelling the cybercrime process. We shall also discuss briefly analyses that could be done with the data but was not, and finally we shall indicate some of the important omissions in the survey questions (or reporting formats) themselves.

[32] An effort was made to contact researchers in this area for leads to data sources, and some had very kindly responded with additional sources. Thanks are due to James Backhouse of the London School of Economics, Peter Grabosky of the Australian National University, Mike Sutton of Nottingham University, S.E. Thatcher of the London School of Economics and David Wall of Leeds University. However, any omissions or mistakes are entirely due to the author.

The most widely cited survey on cybercrime is the Annual CSI/FBI Computer Crime and Security Survey[33]. The 2002 survey is the seventh conducted so far, and it is a poll of information security professionals at US organizations. The response rate has been around 14% for the last few years. Only three demographics of the organization are collected: industrial sector, number of employees and gross income. In the report on the survey results, it is noted that there will be changes in the eighth survey, so it can be hoped that more demographic data will be available on the respondents. Unless we have more information, we shall not be able to see if different types of organizations have different experiences nor can we extrapolate to all organizations generally. The only question asked about the computer systems is on the security technologies used. However, it would be very useful to know what their systems are used for, the applications run on them and the extent of usage. Then the organizations are asked about attacks and intrusions: how many in the last year, from where and of what kind. For analytic purposes, a key piece of information would have been the *times* of the attacks. This is probably not very difficult, since very likely there are records of these attacks in the organization. There are some questions asked specifically on the WWW sites. 97% of respondents have WWW sites and 47% use them for e-commerce. The organizations are asked about unauthorized access or misuse, how often they occurred, from where and of what type. The most significant finding is that a high proportion of these attacks came from outside (60% from outside and 32% a combination of outside and inside). This once again highlights the issue of the "remoteness" of cybercrime. However, there are only 12 type of attacks or misuse listed. These seem to be incomplete with respect to the range of possible cybercrimes, and also the list does not seem to have been systematically developed. This highlights the need of a standard taxonomy for cybercrime. On the issue of hiring reformed hackers as consultants, the majority of organizations are opposed to it (69%). Finally the organizations are asked about what actions were taken following the attack, and reasons why it was not reported (if not). It appears that negative publicity and fear of competitors are the key reasons for not reporting.

A major thrust of these CSI/FBI surveys has always been to try to quantify the financial losses due to the attacks. This is done again in 2002 and

[33] Computer Security Institute 2002 op. cit.

the responses seem to indicate very substantial losses as well as sharp increases from the previous year. However, these are all estimated and reported by the organizations themselves, and should not be taken at face value. There can be a number of reasons for these rather high and possibly inflated numbers. First of all, these organizations (that voluntarily report these numbers) may well have suffered higher than average losses, so these numbers would be biased in any case. Secondly, (since they are willing to report them) it is more than likely that these organizations are planning to get compensation through insurance, and naturally they would try to use the highest possible numbers. Thirdly, they may also be planning legal action against other entities, and again they would obviously ask for the highest possible damages, as is frequently the case in litigation. Finally, it may be the case that the organizations are vastly over-estimating their information systems and proprietary informational assets. Without a deeper analysis of the situation and an independent verification of the valuation method, it is probably better to wait before accepting these figures.

In the conclusion of the report a number of additional limitations are mentioned such as the self-selection by the 14% who have responded, the veracity of recalled data and the inevitably low detection rates which result in many attacks going unnoticed. However, in spite of these shortcomings, the survey has collected valuable data. Unfortunately, it has not been analyzed to anything near it full potential. For example, with the one exception where losses are computed against type of attack, there are no analyses involving cross-tabulation. Nor are any other statistical methodologies used to obtain a clearer picture from the survey data, although many other analyses are certainly possible. We hope that further analysis will be done in the future.

An improved version of the CSI/FBI Survey was developed and fielded by AusCERT, the Australian Federal Police and other agencies in Australia[34]. Overall, the results are similar to the CSI/FBI findings with a sharp increase in financial losses in 2003 over 2002. However, overall levels of incidents appear to be lower. This survey included a number of additional interesting questions such as the security policies and procedures used. They also queried organizations on in-house security qualifications and experience as well as the satisfaction levels, and they asked about the suspected motive, (although the answers must be guesses in the most part).

[34] AusCERT 2003 op. cit.

Another valuable item of information gathered is the time lost in recovering from the attack. In most cases, it was either less than a day or 1 to 7 days. Here a finer breakdown would have been very useful. Finally, organizations were asked about the outcome of incident allegations. While the summary is not quite clear, it appears that the most common outcome (57%) was that the allegation was investigated but no one was charged. It further appears that in 21% of the cases someone was charged. In spite of its enhanced question-set, this survey has similar deficiencies as noted above for the CSI/FBI survey. Other limitations are also noted in the report itself and which equally apply to the CSI/FBI survey. Thus there could be "discrepancies associated with the respondents' interpretation of questions," and "difficulties in determining which of the answer options were actually applicable." Finally, the data were clearly not utilized to its full extent since many potential paths of analyses[35] were not followed up.

Currently there are a number of efforts underway by several government agencies to collect information on cybercrime. However, not much information is available yet to make any assessment. For example, the *2001 US National Crime Victimization Su*rvey has questions on computer use, experience and computer-related incidents. Similarly, the *2001 British Crime Survey* also has some cyber-crime related questions. In Canada, the *General Social Survey* and the *Household Internet Use Survey* have begun to include a few questions related to cybercrime. However, in all cases, the scope is extremely limited with respect to gaining insights into the prevalence and patterns of cybercrime as a whole. It is clear that the tactic of appending a few questions to a traditional survey is not at all adequate for the proper study of cybercrime. More specialized surveys will have to be conducted.

There have been a number of other, smaller surveys to learn about experiences and attitudes of individuals with respect to cybercrime. For example, in one study of Swedish organizations, it was found that viruses were the single most common problem (63% of incidents), most cases of unlawful access occur over the Internet (77%) and about 39% of IT offenses are reported to law enforcement[36]. Furnell[37] cites some of his own sur-

[35] For example cross-tabulation and correlations.

[36] Korsell, L.E. and Soderman, K. IT-related Crime: old Crimes in a New Guise, But New Directions Too! Journal of Scandinavian Studies in Criminology and Crime Prevention, 2, 5-14. 2001.

[37] Furnell 2002 op. cit.

veys where he probes the attitudes of individuals in academic settings on computer crime and abuse. He also tried to rank the seriousness of different types of cybercrimes. He cites further surveys by the UK Audit Commission and the UK National Computing Centre. Also in the UK, the Department of Trade and Industry conducts the "Information Security Breaches Survey." These are telephone interviews of individuals responsible for information security in (mostly) business organizations[38].

In another survey to test Social Learning Theory, undergraduates were asked about their sources of imitation (if any), association with others, attitudes towards certainly and severity of punishment for cybercrime and their normative attitudes[39]. The main result found was a positive correlation between use of Bulletin Boards and commission of computer crime. Cyberlaws appeared be have symbolic rather than deterrent value. Yet another survey found that students (in 2 US universities) relate more to other students than to faculty, university employees or the immediate community as regards software piracy[40]. Finally, a survey, exploring the attitudes of IS managers towards network attacks, found some approval of cybervigilantism or "striking back" but also complacency about organized attacks[41].

These last few are all relatively small surveys and focus on a particular issue, for example, attitudes towards software piracy. Hence the information we get is highly fragmented. In order to get a broader picture of cybercrime, we need more comprehensive surveys, over larger samples (organizations and individuals) and across multiple countries. It is this issue that is

---

[38] Information Security Breaches Survey 2002: Technical Report, DTI, UK.

[39] Skinner, W.F. and Fream, A.M. A social learning theory analysis of computer crime among college students. Journal of Research in Crime and Delinquency, 34(4), 495-518. 1997.

[40] Ramkrishna, H.V., Kini, R.B. and Vijayaraman, B.S. Shaping the moral intensity regarding software piracy in university students: immediate community effects. Journal of Computer Information Systems, 41(4), 47-51. 2001. See also Rahim, M., Seyal, A.H. and Rahman, M.N.A. Factors affecting softlifting intentions of computing students: and empirical study. Journal of Educational Computing Research, 24(4) 385-405. 2001, Hinduja, S. Correlates of Internet software piracy. Journal of Contemporary Criminal Justice 17(4), 369-382. 2001, and Hollinger, R.C. Crime by Computer: Correlates of software piracy and unauthorized account access. Security Journal, 2(1), 2-12. 1992. for additional survey results on software piracy.

[41] Hutchinson, W. and Warren, M. Attitudes of Australian information systems managers against online attackers. Information Management & Computer Security, 9(2-3), 106-111. 2001.

taken up in the next section. However, in all cases there is the additional need to fully analyze whatever data has been obtained from a survey, and as noted above, this has rarely been done. An additional possibility that arises is utilizing information from one survey or database to correct for biases in the results from another survey or database. With a large number of databases, we might also consider appropriate meta-analyses.

Apart from surveys, we have data on cybercrimes that are simply collected upon being reported. There are now many agencies in many countries that collect reports of various types of cybercrime. The Internet Fraud Complaint Center (IFCC) in the US collects information on complaints received by it and refers them to the appropriate law enforcement or regulatory agency[42]. Although it concentrates on Internet-related fraud, it gets many different types of complaints. While the number of fraud complaints has increased steadily since 2000, those numbers cannot be taken as any valid trend in Internet fraud, since the reporting base is not known. The Internet itself has grown. The number of users has grown enormously. Further, awareness of IFCC as an agency to report an Internet fraud case to must also have grown. There may have concurrently grown a willingness to report a fraud to the IFCC for a variety of reasons. Thus the reporting base has most likely grown significantly, and unless that base is estimated, we cannot conclude anything about the trend. The base has to be ascertained from a separate survey. In 2002, the top three complaint categories were Auction fraud (46%), non-delivery (31%) and credit card fraud (12%). IFCC also collects reported financial losses from complainants, but those numbers cannot be verified, and hence must be considered tentatively. IFCC also asks about the gender and residence of the perpetrators and the gender, residence and age of the complainant. The report remarks on the special case of Internet fraud (and which is true for Internet crime in general) that it is truly borderless with the perpetrator and the victim often very far apart.

The Bundeskriminalamt of Germany has been collecting data on reported computer crimes. It uses eight categories of computer crimes, and by far the dominant type is fraud with credit or cash cards (61.3% of 79,283 cases in 2001). It also records a spectacular rise of 40% from 2000 to 2001 in all computer crimes taken together. However, a large part of this increase is due to computer fraud in general and fraud related to telecommunications services. It is not clear the extent to which higher awareness on the

[42] *IFCC 2002 Internet Fraud Report*, The National White Collar Crime Center.

part of the reporting public, better reporting or more efficient collection of reports has played a role in the observed trend. Apart from the fact that it is a relatively new crime type and hence will require some time for modifications and adjustments in the reporting procedures as more is learnt about it, the typology does not appear to be satisfactory. Some categories are too broad, and some cybercrime types are not included.

In the US, the Uniform Crime Reports (UCR/NIBRS) have begun to include statistics on "Computer Crime Offenses by Type." However, only in two places is it noted in passing whether it was a computer-related crime: in "property description" and in "offenders suspected of using (among other options) 'computer equipment'." The breakdown of the frequencies by "type" follows the categorical breakdown used for traditional crimes, and gives no insight into the cybercrime situation. The National Crime Victimization Survey (US) has also started to include questions on cybercrime, but the data were not yet available at the time of writing. In Poland, the TPSA (Polish Telecom) has collected some data on network incidents that show increases in scanning and probing, and the Polish Police Computer Crime Unit is also collecting data on computer crimes as well as sentencing. It appears that most of the sentencing has been for computer fraud[43]. A number of private and semi-private organizations also record reported cybercrimes such as viruses, but they all appear to suffer from a variety of methodological shortcomings, including the ones mentioned earlier[44].

Around the world, Computer Emergency Response Teams (CERTs) have been designated (usually by their national governments) as a central point for reporting computer network attacks, intrusions and crimes. The original one is CERT/CC based at the Software Engineering Institute, Pittsburgh, USA, and it has been collecting data on network incidents for a number of years. This data basically consists of records of reported incidents between 1988 and 1995. For each incident, the variables recorded are: SD (start date), ED (end date), NS (number of sites involved), NM (number of messages), LV (level of the incident), MO (a vector of methods of operation used), CA (corrective action), NT (notes), RS (a vector of re-

[43] Adamski, A. Computer crime in Poland: Three years' experience in enforcing the law. Available on the Internet.

[44] Some of the better-known ones are included in the metadatabase in the appendix.

porting sites), and OS (a vector of other sites involved). The data has been described and analyzed in detail elsewhere[45]. The information is given by the site experiencing the attack, and is essentially based on network traffic data. However, for a criminological investigation or for policy development, we would like to have much more detailed information. A recommendation has already been made to include additional information when recording network incidents[46]. The main recommendations regarding further data collection include:

1. Maintain incidents-data in a standard database for easy access and analysis after sanitizing (to ensure anonymity).
2. Map details of incidents into broad categories in terms of costs, impacts, and survivability of the victim system.
3. For each incident: order or rank the MOs by some criteria (such as seriousness or skill-level).
4. Collect data on inroads made for each incident & the "end" state of systems after the incident.
5. More detail in the NT and CA fields.
6. Trace data on perpetrators whenever possible.
7. Identification of attacking site(s) and number of attacking sites.
8. Model learning on the part of attackers and on the part of victim sites.
9. Data on various reactions and response times at the victim sites.
10. Long-term precautions that victimized sites take.

In the next section we address the problem of improving data collection instruments and developing suitable surveys and methodologies.

[45] Howard, J. 1995 *An Analysis of Security Incidents on the Internet (1989-1995).* Ph.D. Dissertation, Carnegie-Mellon University; Moitra and Konda 2000b op. cit.

[46] Moitra and Konda 2000a op. cit.

## 3. Patterns of Cyber-Victimization

It is self-evident that to understand a phenomenon satisfactorily, it is important to have sufficient and reliable data on it. Since cybercrime is a relatively new phenomenon, it is not too surprising that we do not yet have the kinds of data we need for a proper understanding. We have seen the problem in the previous section and in fact, many commentators on cybercrime have noted this deficiency[47]. In this section we discuss some further issues related to cybercrime that we would like to study, and suggest some data collection instruments to obtain the additional data needed to overcome this deficiency to the extent feasible.

### *A. Summary of limitations discussed previously*

It has been noted that there are two classes of data sources: those actively collected through surveys and those reported by victims to a designated center. It has also been pointed out that the available data (of both types) have major flaws and biases. For one, there is no standardized conceptualization or taxonomy, so the reported numbers from different sources cannot be compared or aggregated to get an overall picture of cybercrime[48]. In fact there is still a debate on what exactly are cybercrimes. For another, no systematic reporting and no uniform recording methodologies have been developed. We also know very little about the victims, how they have been victimized and how they perceive the incident themselves[49].

Some of the limitations are common to both the types of data (survey and reported) and in addition the current data of each type have their own limitations. A common limitation is that many variables that would be of interest are not included, for example, details of the impact of cybercrime on the victims. Also, the research design and the hypotheses to be tested are not spelled out in most of the cases. Yet another problem is the lack of validity checks on the data that have been collected.

As far as surveys are concerned, there are several problems (some already mentioned) that should be resolved. One is the composition of the

[47] Grabosky, 2000 op.cit.; Sieber, U. The International Handbook on Computer Crime. John Wiley & Sons, Chichester. 1986; Wall, 2001 and 2003 op.cit.

[48] For a discussion on the issue of taxonomy, see Moitra 2002a op. cit.

[49] Wall, 2001 op.cit.

sample. A possible self-selection on the part of the respondents can give rise to a number of biases. For example, those organizations that are better at detecting network intrusions may have a higher probability of reporting. As a result, we may be overlooking many intrusions against other organizations. There can be errors of both types in the reported damage. Some organizations may wish to exaggerate the damage in order to claim compensation from insurance companies. Other organizations may not admit any damage in order to protect their reputations. Thus the data should be examined very carefully and these possible errors should be taken into account. These surveys have not yet focused on individuals or households (although there are a few exceptions) and the demographic variables collected are limited. The issues of attitudes toward cybercrime, fear of crime in cyberspace, and self-reports on dubious activities have hardly been investigated. Finally it is always important to investigate the non-respondents as far as possible, and this does not appear to have been done at all.

The reported data are again of two types. One type is the set of reports on actual or suspected crimes by the victims and the other is the data collected by designated centers (CERTs) which tend to have network and traffic details about intrusions but not the eventual crimes committed (if any) after an intrusion takes place. Thus we have very little knowledge of the impacts and the outcomes of these incidents. A general problem is that trends cannot be estimated since the reporting base is usually not known. For example, the trends reported by the security industry or other organizations can be biased since the reporting base could have (and probably has) increased dramatically. Thus, a 40% rise in reported incidents may actually signify a *decrease* if the user population has increased by 50% in the same time period! Finally, there may well be a self-selection bias in the reported data if those who have experienced more crimes or more serious crimes have a higher probability of reporting them (similar to the problem in surveys).

These are just a few among many other possible biases. Such potential biases have been well discussed in the literature on research methodology. Future data collection instruments should attempt to rectify these problems at the very least. Furthermore, the instruments should be extended to address additional issues in cybercrime that we shall discuss below.

## *B. The victimological perspective*

Cybercrime impacts both systems and their users. The victims of cybercrimes are computer/information systems and users, the latter including

both individuals and organizations. Thus cybercrime has opened up a new area in victimology. It adds new dimensions to the nature of victimization which include the relationship between information and its user/owner and people's view of cyberspace. The study of cybervictimization will provide a vital key to understanding the patterns of cybercrime and its overall impact. From this point of view, previous studies have further serious limitations in scope, both in terms of the items of information collected and of the respondents surveyed.

This analysis of victimization will be at the "micro" level compared with the "macro" level examination of aggregate reported data. To properly understand the patterns of cybercrime and develop effective and equitable policies, we need to understand the cybercrime process as experienced by victims and the impact it has on victims directly or indirectly. The issues to be studied in cybervictimization include the experiences, attitudes and behaviour of users of the Internet.

There are a number of questions concerning patterns and trends that can be further pursued and many insights into the patterns of cybercrimes and victimization histories can be gained. Organizations and well-known sites often experience repeat victimization. For such sites, we can construct the pattern of victimizations over the sequence of crimes that they have experienced. For example, we might be able to estimate the rates of occurrence of different types of incidents and the patterns in successive types of incidents. This would be similar to developing a model of "victimization histories" and would be analogous to "criminal careers". In criminal careers we model the criminal activities of an individual over time[50], whereas here we would be modelling the victimization experience of a network site (or set of sites belonging to the same organization) over time. However, the models will have a similar structure. As in criminal careers, we could use the concept of a "crime-switch-matrix[51]." to analyze the successive types of crimes that victims experience. We can also examine whether certain types of sites are prone to certain types of attacks. With such data, it is possible analyze the times between incidents and see if there are any variables that

[50] Blumstein, A., Cohen, J., Das, S. and Moitra, S.D. *Specialization and Seriousness During Adult Criminal Careers*. Journal of Quantitative Criminology, 4(4), pp. 303-345. 1988; Cohen, J. op. cit.

[51] Albrecht, H-J. and Moitra, S.D. *Escalation and Specialization : A Comparative Analysis of Patterns in Criminal Careers*. Crime and Criminal Justice, 36, pp 303-345. 1988. Max Planck Institute, Freiburg.

explain them. For example, in a previous study, it was found that the inter-incident times depended on the type of site and on the types of incidents[52]. The above discussion simply highlights some of the possible questions regarding cybervictimization that can be researched. In fact there are many other related issues that could be studied once we have sufficient and reliable data[53].

> We would also like to relate the analysis of victimization of network sites to the new area of locational analysis in criminology[54]. These are a class of spatial models where "the place of crime" is the unit of analysis. These models may be particularly relevant for cybercrime since the victim systems are points in cyberspace where the criminal activities occur, and there might be certain characteristics of the "place" that can be modelled and which could yield insights into the cybercrime process. In general, we can consider the "site" as the place of crime where the "site" could be any computer system (even a server) that is involved in any act of cyber-deviancy.

For these analyses, we need data at the victim site level including information on the times and types of crimes/attacks/intrusions that have occurred at a site/system. Some data of this kind are being collected by the CERTs as discussed above and also by the CSI, etc. However, the collection/survey methodologies need to be enhanced if we are to answer some of the important questions and that is why we are suggesting the following instruments. It is quite conceivable that we shall be able to combine the different types of information from these diverse sources, and be able to conduct a much richer analysis. In addition to organizational sites, we need to learn about the victimization experience of individuals and households who use the Internet. We would be interested in knowing about their experiences on the Internet, their attitudes towards cybercrimes and some details about their usage as well as their "behaviour" on the Internet. Thus, the surveys we propose will gather information on their experience with viruses, harassment through email, and so on. Further, we would like to learn about their responses to being victimized: whether they reported the incidents, asked

---

[52] Moitra and Konda 2000b op. cit.

[53] Blumstein, Cohen, Das and Moitra op.cit.; Cohen op. cit.; Greenberg op. cit.; and Wittebrood, K. and Nieuwbeerta, P. 2000 *Criminal Victimization During One's Life Course: The Effects of Previous Victimization and Patterns of Routine Activities*. Journal of Research in Crime and Delinquency, 37 (1), 91-122.

[54] Eck and Weisburd, 1995 op. cit.

for help from friends, retaliated, etc. Respondents will be asked about their perceptions of the risks in using the Internet and cybercrime in general. Some demographic data will have to be collected of course. Among them, it would be interesting to know their attitudes towards risks in general. It is also important to understand the extent of fear of crime in cyberspace. For instance, do people (or organizations) restrict their use of the Internet out of fear? Do they avoid visiting certain types of sites such as chat-rooms because they are perceived as unsafe?

These surveys should cover all segments of society and include people with different levels of knowledge, involvement and Internet usage. They should include students, professionals (and within these, computer specialists, lawyers), law-enforcement personnel, households (with multiple users) and members of the general public who have very little (or no) experience with computer networks and therefore might not have been victims of cybercrime. Finally, we would like to include some open-ended questions about their opinions on what should be done and their recommendations regarding cybercrime. With such data, we could relate victimization/harassment experiences to their usage patterns. We could also correlate their experiences with their demographic characteristics, "behaviour" and attitudes.

> The network traffic data collection at sites is an additional issue that is very important. It is synergistic with Intrusion Detection Systems (IDSs) which monitor the traffic for signs of cybercrimes[55]. However, the data collected is also important for network forensics that attempts to trace the perpetrator and gather evidence for prosecution[56]. Given its importance for investigation and prosecution, we propose to study this aspect of data collection and how it can be improved to help the criminal justice system. This will involve the consideration of the advantages and disadvantages of alternative rules on network data retention and access with respect to balancing the needs of individual privacy and the needs of law enforcement[57]. The tradeoffs involve the various costs and benefits of storing excessive data versus too little data.

---

[55] Bhaskar, T. and Moitra, S.D. 2002, *Genetic Connectionism for Intrusion Detection System*, IIMC Working Paper.

[56] Casey, E. 2000 *Digital Evidence and Computer Crime*. Academic Press, New York; Kruse, W.G. II, and Heiser, J.G. 2001 *Computer Forensics: Incidence Response Essentials*, Addison-Wesley, New York; Mandia, K. and Prosise, C. 2002 *Incident Response: Investigating Computer Crime*. Osborne/McGraw-Hill.

[57] Clifford, 2001 op. cit.

## *C. Data needs and methodological considerations*

The goal of the survey research on cybercrime is to build on previous work in criminal justice modeling, especially criminal careers (as analogous to victimization histories) and victimization studies[58]. In making inferences from the data, we also need to understand the process by which data on crimes and incidents are generated. It is a "twice-filtered" process, since first the crime must *detected*, and then it must be *reported*. Thus we have to contend with both detection and reporting rates. Again this is analogous to detection, reporting, arrest, conviction or sentencing rates that have been studied in criminal justice models[59]. In the case of cybercrime, these rates are exceptionally low at present. One of the interesting issues is how and why they vary across reporting units. An investigation of these questions could help in developing policies that might increase these rates and this could in turn help in controlling cybercrime and providing better data in the future.

In fielding the surveys it is extremely important to ensure sufficient sample sizes and a wide variety of respondents. As already explained, we are considering C2C cybercrimes only. An additional methodological issue for future research is the updating of a standard taxonomy for cybercrime, so that when crime types are compared across surveys and over time, the comparisons are valid. There are of course a number of other methodological issues related to analyzing the data and discussions on them can be found in the literature. Here we restrict ourselves to developing the surveys with a view to estimating models for cybercrime and cybervictimization that build on models in traditional criminal justice. There is currently a notable lack of thorough statistical analyses of cybercrime data, and it is hoped that the proposed survey will allow more sophisticated analyses. One other methodological issue is the examination of the representativeness of the sampled base and the non-responses.

> As noted earlier, the fielding of the survey and the analysis of the data will represent a new area in victimology where both information and computer systems as well as individuals and organizations have been victimized remotely, and often in new ways, over the Internet. With the proposed surveys and data collection methods, we will be to develop and estimate more detailed models of the cybercrime process. The kind of data currently

[58] The relevant literature has already been cited in earlier notes.
[59] See above.

being collected are not enough, since they were not designed to answer such questions, and a systematic plan for data collection on cybercrime is necessary. We need regular, on-going and well-designed data collection methodologies to obtain information along the different dimensions of cybercrime. It is important that the data collection be statistically rigorous and uniform over countries and time so that reliable inferences can be made about the nature of cyber-victimization in its various aspects. Such data will provide more relevant information regarding the details of cybercrime, will help answer the questions of interest more precisely, and suggest how effective policies could be developed to reduce and control it.

## *D. Overview of the surveys:*

The above discussion suggests a need for developing new questionnaires for further surveys along with some more detailed specification of data collection methods. Given the plans and activities in progress to gather data on cybercrime by many governmental agencies and other organizations, these surveys discussed below should be of help to them in their efforts. Therefore, we have developed data collection instruments at four levels that are described below.

1. We have developed a carefully designed questionnaire to survey individuals and households. This will collect data on demographics, victimization experiences and general attitudes regarding the Internet, cybercrime and risks. The design will draw upon the current state of the art in victimology and survey design[60].
2. We have also developed another questionnaire for organizations that will include questions about their computer systems, applications, Internet use and systems management issues including security measures[61].

---

[60] The US National Crime Victimization Surveys; Wittebrood, K. and Nieuwbeerta, P. op.cit.; Van kesteren, J., Mayhew, P. And Nieuwbeerta, P. 2000 *Criminal victimization in 17 industrialized counties: Key findings from the 2000 International Crime Victims Survey* (Dutch Ministry of Justice). See also Fowler, F.J. 2001 *Survey Research Methods*. (3rd ed.) Sage Publications, Beverly Hills; Nardi, P.M. 2002 *Doing Survey Research: A Guide to Quantitative Research Methods*. Allyn and Bacon, Boston, for methodological details in survey design.

[61] These are based on earlier surveys by Computer Security Institute, 2002 op. cit. and AusCERT, 2003 op. cit.

3. We have proposed a data collection system at sites so that relevant variables associated with the network traffic at the site is collected. This involves setting appropriate filters at specific nodes of the networked system. This has been recognized as an important issue, and the Commission of the European Union has noted a need for developing efficient methods for such data collection. Such data is needed first to detect network attacks and crimes and second, to investigate the cybercrime. This area of network forensics is an important new and unique area in cybercrime and it is essential to study this for more effective investigation and prosecution of cybercrime. Given the transnational nature of cybercrime, it is very important that the methods be uniform across countries.
4. We have also proposed a broader system of data collection for designated reporting centres. Some such centres have already been established, and are often known as CERTs (Computer Emergency Response Teams) or something similar. They collect data that is voluntarily reported to them regarding network incidents, and we incorporate ways of improving this system of reporting to and collection by the CERTs in our suggested system. This data would be based on the data collected in the previous level (that is, at sites with computer systems connected to the Internet) and would consist of the details of any attack or intrusion that have been detected at that level and reported by the sites to designated centers. Again it is important that the methods be uniform across countries and over time.

These four instruments are given in the appendix. They will of course have to be fully formulated, formatted and perhaps edited before being fielded. It cannot be expected that they will suffice immediately. Rather, they will have to be tested, tried and refined over time. As in all survey research, we shall have to achieve a balance between the length of the survey and the amount of information desired from respondents.

## *E. Issue of confidentiality*

With respect to research on cybercrimes, the confidentiality of the data from various sources needs to be given due importance. It is quite possible that some of the data ideally needed for research are confidential or private. This could be an item asked in a questionnaire or data on a network intrusion that has been reported on the basis of confidentiality. In many cases,

it will be possible to strip the data of sensitive identifiers and thus render the data completely anonymous. This has in fact been done in many contexts in scientific research and in particular with data on network incidents[62]. The term "sanitized" is also used to describe this process, as a result of which the researchers (and the readers of the research results) will **not** be able to identify the sensitive data items. However, it will still be possible to conduct the desired research on patterns and correlations among variables derived from the anonymous subjects.

## 4. Modelling Cybercrime

In this section we discuss a simulation model for studying cybercrime. As argued earlier, a simulation model offers a number of potential advantages: it is a useful method for investigating complex systems or processes (and so is widely used in many areas); it allows us to overcome the limitations of data availability (through sensitivity analysis) and it enables us to analyse policy impacts (by exploring different scenarios).

The cybercrime process is clearly extremely complex and we need a way to model the interactions among hacker behaviour, computer security, victims' responses, informal Internet controls, legal policies, law enforcement activities and socio-economic conditions. Simulation is the ideal tool in such a case where the process is too complex to be handled by other analytic techniques[63]. Secondly, it is almost certain that many gaps exist, and that some gaps will always exist, in our knowledge of cybercrime. Simulation offers a way to overcome the data limitations we have since a "simulation model permits us to incorporate many sources of data of varying quality and to fill in blanks with educated guesses where there are simply no data."[64] These guesses in turn can be partially validated by running many scenarios and assessing the results to see if they make sense. That is, we can

---

[62] Howard, 1995 op. cit.; Moitra and Konda, 2000b op. cit.

[63] For a more detailed discussion on simulation see Law and Kelton op. cit. For simulation applied to computer networks see Sinclair, J.B. *Simulation of Computer Systems and Networks: A Process-oriented Approach*, Cambridge University Press, 2004; Xiao, X. et. al. *A Practical Approach for Providing QoS in the Internet Backbone*, IEEE Communications Magazine, 40 (12), 56-62, 2002 and MASCOTS 2002 (10th International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems, IEEE)

[64] Caulkins, Crawford and Reuter op. cit.

check whether the results conform to what we do know about cybercrime. Simulation can partially compensate for uncertainty about various parameter values through sensitivity analysis. This sensitivity analysis will also indicate which variables are important and which variables have relatively little influence on the process. Further, the exercise of modeling in itself points out precisely what additional data are required to understand the process more fully. Therefore on one hand we can fully utilize the information we can get by using the results from collected data as inputs, and on the other hand compensate for unavailable data through sensitivity analysis.

A third advantage is that it enables us to explore alternative scenarios and the implications of different policies. We can observe how changes in a set of parameters (say Internet policy) can impact the extent and nature of cybercrime. We can control one or more sources of variation to understand the effect of variation in another source.

Simulation makes it possible to experiment with different scenarios and to observe the results very quickly by changing the inputs and assumptions and rerunning the model. It also permits us to compress time and we can observe long-term trends and effects by running the model under the conditions we are interested in. All of this will help in better understanding the policy implications of alternative prevention and control measures.

Finally, the simulation model can help capture important aspects of the complex cybercrime process. It can be used to test various assumptions and can examine the *implications* of the assumptions we make. It can provide many insights into the process. It would lead to a clearer understanding of the problems and uncertainties we face as we try to achieve a deeper understanding of the whole phenomenon. It could also be used for forecasting, if the model is found to be reasonably valid. At a minimum, it will provide us with a framework for an initial view of the process, and which can be extended, modified and refined further as we acquire more information in the future. In summary, the simulation model would support governments, criminal justice/law enforcement officials, Internet policy makers, and major organizational users of the Internet arrive at better decisions regarding risk management in cyberspace.

## *A. Outline of the model*

In light of the usefulness of a simulation model for cybercrime, we develop an outline for such a model. We highlight the major issues and parameters

that have to be taken into account and indicate what results we can expect from the simulation. As in models of criminal careers[65], we shall start with a set of active criminals and assume they commit crimes at a certain rate (which could vary across individuals). Here we need to consider the type of crime as well. Initially, we shall consider only the major types of cybercrimes to keep the model relevant and tractable, and expand the crime-type-set in the future. As clarified earlier, we shall be modeling "C2C" crimes only. Next we can proceed to model the incidence of cybercrime by taking a distribution of the crime-commission rate across the cybercriminals by crime type. The model will take into account the dependencies and functional relationships between the individual crime rates and law enforcement policies.

Initially we can start with a relatively simple model and then extend it to consider more complex phenomena and interactions later. As we proceed, we shall have the opportunity to validate the current model with the latest results on cybercrime prevalence and update the model as necessary.

A special characteristic of "C2C" crimes is that the target computer system is also selected. In actuality, this selection depends on a number of factors related to the attack as well as the victim system. These factors include motivation, opportunities and skills, as in the "Routine Activity" paradigm. We can start with the simplifying assumption that, at any given time, a set of systems are on the Internet and it comprises a heterogeneous mix of systems with varying degrees of vulnerabilities and attractiveness as targets.

Having simulated an attacker-victim pair, we shall then consider the impact of that attack (or crime) on the victim. This will essentially depend on the attack type (including the skill of the hacker) and the defenses or security mechanisms that the attacked system possesses. As mentioned earlier, this has been modeled previously[66]. What is needed for a cybercrime simulation model is an extension that considers other aspects of victimization. Modelling the victimization process should incorporate the findings from the various cybervictimization surveys and studies. Based on these findings, we can formulate the assumptions regarding the patterns of crimes and victimizations needed to develop and run the simulation model[67].

---

[65] Barnett, Blumstein and Farrington 1989 op.cit.; Cohen 1988 op.cit. and Greenberg 1991 op.cit.

[66] Cohen, F. op. cit; Moitra and Konda, 2000a op.cit.

[67] Actually the response of the victim is also a relevant factor that will be included later as the model is refined.

The analysis will provide us with estimates of the prevalence and impacts of cybercrime based on whatever current information we happen to have and the assumptions we make. The power of this tool is that we can *change* those assumptions as appropriate, and see the resulting effects on prevalence and impact. Since the assumptions will include the functional relationships between criminal behaviour and Internet-related policies (including responses of victims and other net users), we can explore the effects of different policies and responses on the behaviour of malicious hackers and hence on the prevalence of cybercrime. Since one of our goals is to identify effective policies, this exploration is an important step in doing that.

Another extension that should be incorporated into the simulation is the phenomenon of multiple attackers (or multiple attacks by one or a few attackers) on possibly multiple systems. One of the advantages of the simulation approach is that extensions like these are fairly easy to incorporate into the model. The issues here are the distributions of the number of attackers and the number of attacked systems (per attack), and we already have some initial data on these[68]. As further data are gathered and analyzed, we can easily update the distributions accordingly.

This proposed simulation will have several limitations, but it would mark a beginning for this kind of investigation. The actual context of cybercrime is very complex, and any model or analysis represents a balance between tractability and realism. While there are necessarily approximations in any model, a sufficiently carefully constructed one can provide us with estimates of variables of interest and insights that can be extremely valuable. Further research and model development will help overcome many of the limitations and provide us with an understanding of cybercrime that would be continuously improving over time.

[68] Howard 1995 op.cit.

## 5. Summary and Conclusions

This paper has discussed three areas of empirical research in cybercrime: the exploration and assessment of available data on cybercrime, the design of further surveys to collect additional data on cybercrime and cyber-victimization that would be relevant for criminological and policy analysis, and the development of simulation models for Internet management and criminal justice policy development. It has assessed some of the data available on cybercrime. In recognition of the need for more data, two surveys have been proposed. In addition, the paper has proposed a traffic data collection procedure and an enhanced system of recording reported network incidents. The paper has also discussed further data analysis and modeling with the data from a victimological perspective that would be available from the proposed surveys. Finally, it has outlined a simulation model that could complement the data analysis and further our understanding of cyber crime.

The main conclusions arrived at are:

- the available data have serious limitations and biases
- we need much more empirical analysis of cybercrime
- further surveys are needed in this area
- the traffic and incidents data collection should be enhanced
- a simulation model would be very useful for furthering our knowledge

The benefits of this discussion has been:

Knowing what data are available and their relevance for modelling

- Identifying key data needs for future research
- Developing data collection instruments to gather data in the future
- Knowing more about the usefulness of models for Internet policy analysis

- Providing a basis for future research on cybercrime that would include collection and analysis of relevant data, a new kind of victimological research and the development of a simulation model for cybercrime.

Some important directions for future research would be:

1. Extension of the metadatabase on cybercrime.
2. More detailed data analysis of existing data.
3. Developing a taxonomy for cybercrime.
4. Fielding the proposed surveys and analyse the data obtained.
5. Build the proposed simulation model.

# Appendix A:

Table 1: An Initial Metadatabase of Cybercrime Data Sources[1]

| Data set name and owner/source | Description of contents | How collected (reported or surveyed) | Availability/Conditions/Format | Details of the variables | No. of cases / time-frame | Summary statistics /available analysis | Possible inferences and deductions | Potential inputs for simulation model | Observations |
|---|---|---|---|---|---|---|---|---|---|
| CSI/FBI Computer crime survey | responses from security practitioners | Survey of organizations | n.k. | attack rates, types and financial losses | 503 cases; annual | detailed report available free online | attack rates and losses; sample biases | depends on segmentation of respondents | should be considered; check availability |
| AusCERT Survey | responses from security practitioners | Survey of organizations | n.k. | All of CSI/FBI variable plus others | 214 responses; annual | detailed report available free online | Several useful estimates can be derived | depends on segmentation of respondents | should be considered; check availability |
| Information Security Breaches Survey: Department of Trade and Industry, UK | Reported security breaches | Survey of UK organizations | n.k. | Related to security in the business environment | 1000 in 2002; annual | Technical report available on request | Several useful estimates can be derived | Can be useful for business segment | should be considered; check availability |

[69] This is a short, preliminary selection of many surveys and reports on cybercrime. It is to be viewed as a template only, and not a comprehensive or representative set of databases. Regularity was one of the criteria used to select these.

| Data set name and owner/ Source | Description of contents | How collected (reported or surveyed) | Availability/Conditions/Format | Details of the variables | No. of cases / time-frame | Summary statistics /available analysis | Possible inferences and deductions | Potential inputs for simulation model | Observations |
|---|---|---|---|---|---|---|---|---|---|
| CERT Incident Reports: CERT/CC, SEI | reports on network attacks and incidents | voluntary reporting from sites | Some old data analyzed in detail in CERT.Org website; annual CERT/CC Trends reports. | Several incident descriptors; described in the reports | Varies by year | Moitra & Konda, Howard: CERT website | discussed in CERT reports. | some useful parameters estimated | Data needs to be preprocessed |
| IFCC 2002 Internet Fraud Report | fraud reports to IFCC | reports from consumers | n.k. | Type of fraud, loss and complainant age and gender | 48,252 | summary report available free online | n.k. | n.k. | Not useful in its present state |
| Bundeskriminalamt<br><br>Annual Reports | Reported computer crimes | Part of routine data collection | N.A. | Eight types of crimes included | Annual | German Police Statistics | Trend Analysis | Validation for prevalence rates | |

| Data set name and owner/source | Description of contents | How collected (reported or surveyed) | Availability/Conditions/Format | Details of the variables | No. of cases / time-frame | Summary statistics /available analysis | Possible inferences and deductions | Potential inputs for simulation model | Observations |
|---|---|---|---|---|---|---|---|---|---|
| Riptech Internet Security Threat Report | Attack trends | Monitoring many organizations | n.k. | | quarterly | Information available online | n.k. | n.k. | |
| ICSA Labs Annual Virus Prevalence Survey | Prevalence of virus attacks | Survey | Internet Risk Impact Summary.<br><br>Available online. | Viruses | quarterly | Online: little detail | n.k. | n.k. | |
| Internet Security Sytem's Internet Risk Impact Summary Report | Security violations | Reported data | Report available online | Variety of attack types | quarterly | Online: Little detail | n.k. | n.k. | |

# Appendix B:

*Survey of Computer Usage, Network Security and Cybercrime.*[70]

Demographics:

| Occupation: | Age: | Education: | Income range: | Sex: | Residence type |
|---|---|---|---|---|---|
| Part of a Household? | How many? | Ages: | Knowledge of computers: self | Others? (Range) | Number using computer? |
| Current location? | Originally from? | Own mobile phone? | Own DVD player | {hi-tech items} | {life style variables} |

Computer ownership, usage and behaviour:

| Home PC? Number? Laptops? | {hardware} | Modem/Internet {ISP?} How long? | Software (main applications) | Shared by? | Usage details |
|---|---|---|---|---|---|
| Use PC outside? Where? | Internet usage / email. | Purchases? Any other e_commerce activities? | What is stored on your PC? | What do you download from the Internet? | What is the value of information and data you have stored? |
| How do friends use computers? | Specific interests? Read Ads? | Security tools? | Ever sent a virus? | Ever harrassed anyone? | Misused trust? |

[70] Open-ended questions that might also be included: Self-reported activities and behaviour on the Internet: Malicious Hacking, time spent on them, knowledge of hackers, etc.; Opinions and suggestions.

Experiences: (in the last twelve months)

| | | | | | |
|---|---|---|---|---|---|
| Intrusions or attacks? | Viruses? What kind | Data theft? | Hardware damage? | Remote usage? | Other crimes? |
| At what times? | Damage from Viruses? | Valuation of damage? | Monetary loss? | Responses to attacks? | Learnt anything? |
| Suspects? | Best source for help? | Internet Fraud? | Reported to …? | Outcomes. | Others' experiences? |

Attitudes:

| | | | | |
|---|---|---|---|---|
| Awareness of cybercrimes. | Worry about cybercrime? | Fear of being a victim (by crimetype) | Perceived sense of seriousnesses. | Sense/degree of illegalty of various acts |
| Types | Value of privacy | How should it be legally protected? | Punishments for violations? | Attitudes towards hackers |
| Prevalence | Precautions | Avoid sites? | Belief in deterrence. | Take any risks? |

## Appendix C:

### *Cybercrime Survey: Organizations*

Demographics:

| | |
|---|---|
| Position of Respondent | Industry sector/main business |
| Location | Other locations? |
| Number of employees | Annual revenues |
| Total company assets | Market capitalization |

Computer & Information Systems:

| | | | | |
|---|---|---|---|---|
| Details of the computer system … | MIS department? Size? | Level of qualifications of MIS staff? | Internet connection(s) | Major software packages and applications |
| Number of users | Responsibilities of MIS dept? | Average level of experience in MIS dept? | Intensity of Internet usage | Internet applications |
| How maintained? | Role of vendors | MIS budget? | ISP? How long? | Near-term plans? |

Information System (IS) functions and services/security measures.

| | | | |
|---|---|---|---|
| Major functions and services of the IS of the firm. | Audit of information assets. | Security technologies used | Security policies? Standards? |
| IS priorities of the firm? | WWW site? How used? | Cost of security technologies? | Known vulnerabilities? |

Experiences and responses: (last 12 months)
(The answers will remain strictly confidential)

| | |
|---|---|
| Types of Incidents experienced? | How many and at what times? |
| Point of attack? (internal/external) | Likely source of attack? Motive? |
| Unauthorized access or misuse of website | How many and at what times? |
| Responses? | Reported to …? Which intrusions? |
| Outcomes? | If not reported, why? |
| Lessons learnt.<br>Recovery time and problems? | What characteristic of the attack caused most harm? (E.g. hacker's skill, volume) |

For network attacks: damages (answers will remain strictly confidential)

| | |
|---|---|
| Exactly what were the damages in terms of (CIA) functionalities? | Hardware damage? |
| Damage to firms reputation? | Specific failure of security mechanisms? |

Financial losses: Yes/No; If yes, amount by type of attack/crime/misuse:

*Open-ended questions:*

General Policies and attitudes regarding Internet security:

Rank-ordering of cybercrimes by seriousness:

Future plans: Priorities for improvements?

Recommendations for Internet policy?

## Appendix D:

*Network traffic data monitoring method.*

Traffic data that should be monitored for the detection of possible intrusions, attacks or crimes at some node connecting all the computers at a site (or in each sub-network) to the Internet. The filters in the collectors at routers or switches may be set to collect the traffic data. Below is the minimum set of variable that should be collected for the purpose of detecting intrusions (attacks or crimes) and also the variables that should be stored for possible forensics if a crime is suspected. Some optional variables are also listed.

The following data items correspond to fields in TCP dump data and it is at the connection level. For other traffic data (for example, NETFLOW data), similar data items should be collected.

*Variable List:*

Start tine; End time; Duration.
Protocol; source & destination port numbers.
Source & destination IP addresses.
Network service on destination.
Bytes from source to destination.
Bytes from destination to source.
Number of packets transferred from source to destination.
Number of packets transferred from destination to source.
Number of failed login attempts.
Number of compromised conditions.

*Further optional variables (if available):*

Number of messages or flows.
Number of urgent packets.
Number of operations on access control files.
Number of outbound commands in an ftp session.
Number of connection to the same host in past 2 seconds.
Number of connections to same service in past 2 seconds.

This list should be reviewed with the responsible system managers and revised periodically.

## Appendix E:

*Collection of information from a site reporting a network incident or cybercrime.*

Information to be collected by designated centers such as the CERTs on intrusions and attacks reported by sites having networked computer systems.

SD (start date),
ED (end date),
NS (number of sites involved),
NM (number of messages),

LV (level of the incident),
MO (a vector of methods of operation used),
CA (corrective action),
NT (notes),
RS (a vector of reporting sites),
OS (a vector of other sites involved).

Further data that need to be collected include:

11. Details of incidents mapped into broad categories in terms of costs, impacts, and survivability of the victim system.
12. For each incident: order or rank the MOs by some criteria (such as seriousness or skill-level).
13. Data on inroads made into the sytem(s) for each incident & the "end" state of the victim systems after the incident.
14. More detail in the NT and CA fields.
15. Trace data on perpetrators whenever possible.
16. Identification of attacking site(s) and number of attacking sites.
17. Model the learning on the part of attackers *and* on the part of victim sites.
18. Data on various reactions and response times at the victim sites.
19. Long-term precautions that victimized sites take.

A set of guidelines for reporting network incidents can be found at the CERT/CC website www.cert.org.