Security Decision Support Challenges in Data Collection and Use

Ian P. Cook and Shari Lawrence Pfleeger
The RAND Corporation

Abstract
In 2005, the Bureau of Justice Statistics (BJS) surveyed 36,000 businesses about cybercrime. We contrast past cybersecurity data collection efforts with the BJS study. Then, based on a framework for evaluating applicability of data to models and decision-making, we discuss data collection and reporting common to cybercrime studies. Finally, we assess the potential for any data collection effort to fulfill the data needs of multiple security models.

Keywords
Cybersecurity, modeling, decision support

The Need for Security Data
Information security does not lack for insightful, well-constructed models of attacks and defenses (Rue and Pfleeger, 2009), of threats (both purposeful and accidental), and of explanations for why investment in security measures can be underfunded even by well-meaning organizations and managers (Boehme and Moore, 2009). Less abundant are good sources of data with which to empirically validate these models. Indeed, this gap adversely affects not only a researcher’s ability to develop or revise a model but also a practitioner’s daily decisions about resource allocation.

The Department of Justice’s Bureau of Justice Statistics (BJS) attempted to close this gap by conducting the first carefully-sampled, large-scale cybercrime survey in the United States. When the full survey data and findings are published, the research community will have a rich source of credible data with which to set and test hypotheses. This article discusses the structure and findings of the BJS study, reviewing how and where it contributes to our understanding of cybercrime in the U.S. We begin by summarizing past cybersecurity data collection efforts. Then we examine the BJS study directly, discussing general issues of data collection and reporting common to any survey. Finally, we assess the potential for data collection to meet the data needs of multiple security models.

From Data to Knowledge to Practice
Measuring security requires trustworthy, credible data about the type and distribution of attack methods, attack frequency, successful defense methods, and more. In the past, such data have come from externally-administered surveys conducted by a host of institutions and agencies (Pfleeger and Rue, 2008) as well as from internally-generated user surveys, problem reports, and other collection efforts. Various entities actively collect information about online activities related to cybersecurity, including data breaches, spam, botnet activity, and viruses. Table 1 illustrates how, by searching the
web, we can easily find cybersecurity information. The number and variety illustrate both the need for data and the lack of uniformity in data definitions, types, and detail. Despite substantial efforts, no data set (separately or in combination with others) provides a complete understanding of the threat environment or the effects of cybercrime, or serves as useful input to decision-making about cybersecurity investment or effectiveness.

Table 1: Example Sources of Cybersecurity Data

<table>
<thead>
<tr>
<th>DATA SOURCE</th>
<th>TYPE OF ORGANIZATION ADMINISTERING</th>
<th>DATA TYPES</th>
<th>COLLECTION METHOD</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CERT</td>
<td>Governmental / Non-Profit</td>
<td>Breaches, Malware, Software vulnerabilities, Original threat research</td>
<td>Field reporting, original collection, research teams</td>
<td><a href="http://www.cert.org/">http://www.cert.org/</a></td>
</tr>
<tr>
<td>Ponemon Institute</td>
<td>Private Firm</td>
<td>Data breach cost estimates, privacy metrics</td>
<td>Survey creation/fielding</td>
<td><a href="http://www.ponemon.org/research-studies-white-papers">http://www.ponemon.org/research-studies-white-papers</a></td>
</tr>
<tr>
<td>Open Security Foundation</td>
<td>Non-Profit</td>
<td>Data breach incidents</td>
<td>Breach notifications per state laws, media, voluntary review</td>
<td><a href="http://datalossdb.org/">http://datalossdb.org/</a></td>
</tr>
<tr>
<td>Arbor Networks</td>
<td>Private Firm</td>
<td>DDoS Attack Rates and Size</td>
<td>Original survey</td>
<td><a href="http://www.arbornetworks.com/">http://www.arbornetworks.com/</a></td>
</tr>
</tbody>
</table>

There are several reasons for this failure.

- **Lack of visibility.** When data are collected privately, they may not be shared with the broader community, often for fear of revealing trade secrets or losing competitive advantage.

- **Lack of applicability.** Some studies focus narrowly on particular industries, making findings hard to generalize. Other studies have methodological limitations, such as restrictive definitions, that narrow the applicability of their results. (Pfleeger et. al., 2006)

- **Lack of representativeness.** When studies reflect only a set of clients, the sample may not represent all possible industries. Even carefully-designed
samples may not achieve a response rate large enough to assure that the sample is still representative.

- **Lack of completeness.** Data elements represent aspects of processes, products and resources. Without a larger context for interpreting the data, it may not be possible to understand the meaning of changes in data values.

As a discipline matures, so do its measurement and data collection processes. For example, as health science evolved, it developed methodical techniques to evaluate and improve its practices, and to make more effective its decision-making and analysis. Evidence-Based Decision Making (EBDM) responded to demand for insight beyond anecdotal physician experience that could support development, review and implementation of effective medical decision-making (Cochrane 1972). By focusing on evidential quality, analyzing results across multiple studies, and emphasizing methodological rigor, many other professions have been able to use EBDM to evaluate techniques and then incorporate effective ones into common practice.

Other disciplines, such as law (Friedmann and Post 1990) and software engineering, look to EBDM as a basis for effective decision-making. For instance, software engineers evaluate the efficacy of various methodologies, based on evidence for their success. Kitchenham et al. (2004) note the success of EBDM in medicine and suggest creating Evidence-Based Software Engineering (EBSE), describing how it would create common goals across research, inform decision makers, improve software development, and provide support to skills certification. Creating EBDM in cybersecurity requires a solid foundation on which to build. The BJS survey may provide an essential building block.

**The BJS Survey**

In 2003, the U.S. National Strategy to Secure Cyberspace directed the Department of Justice to develop better data about the nature and prevalence of cybercrime and electronic intrusions. “Other data collections address some aspects of cybercrime, but no large-scale (or nationally representative) survey collects sufficient information to accurately measure cybercrime and its consequences or to develop risk factors.” (Rantala 2008) Consequently, in 2005 the BJS administered its National Computer Security Survey (NCSS). The effort was intended to be provide a representative dataset describing cybercrime conducted against or over a company’s networks. To ensure that the results were representative both of U.S. businesses and of key business sectors, the survey was sent to 35,596 private U.S. companies representing 7,278,109 businesses in 36 different industries.\(^1\) Actions encouraging completion were taken in hopes of achieving an 80 percent response rate, but responses were received from only 8,079 (a 23 percent response rate). Nevertheless, this sample is larger in size and wider in scope than any similar study currently conducted by either public or private institutions. Tables 2 and 3 summarize findings from the NCSS report:\(^2\)

---

1 For a full discussion of the sampling and weighting methods, see http://www.rand.org/pubs/technical_reports/TR544/
2 The full report is available at http://www.ojp.usdoj.gov/bjs/abstract/cb05.htm
Table 2: Prevalence of computer security incidents among businesses, by type of incident, 2005

<table>
<thead>
<tr>
<th>Type of incident</th>
<th>All companies responding</th>
<th>Companies detecting incidents</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>All incidents</td>
<td>7,636</td>
<td>5,081</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>Cyber attack</td>
<td>7,626</td>
<td>4,398</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Computer virus</td>
<td>7,538</td>
<td>3,937</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Denial of service</td>
<td>7,517</td>
<td>1,215</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Vandalism</td>
<td>7,500</td>
<td>350</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Cyber theft</td>
<td>7,561</td>
<td>839</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>7,492</td>
<td>1,792</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Experience of Crime vs. Experience of Attacks

<table>
<thead>
<tr>
<th>Highest Prevalence of Cybercrime</th>
<th>Highest Prevalence of Cyber Attacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telecommunications</td>
<td>82%</td>
</tr>
<tr>
<td>Computer system design</td>
<td>79%</td>
</tr>
<tr>
<td>Manufacturing, durable goods</td>
<td>75%</td>
</tr>
<tr>
<td>Chemical and drug manufacturing</td>
<td>73%</td>
</tr>
<tr>
<td>Manufacturing, non-durable goods</td>
<td>72%</td>
</tr>
<tr>
<td>Business and technical schools</td>
<td>72%</td>
</tr>
<tr>
<td>Publications and broadcasting</td>
<td>71%</td>
</tr>
</tbody>
</table>

Unlike customer surveys, the BJS survey was intended to be more representative of business and business sectors than other, narrower data collection efforts. Unfortunately, the “response rates were not sufficient to support national or industry-level estimates.” (Rantala 2008) For example, telecommunications topped the list of those suffering cybercrime and cyber attacks but had the second lowest response rate (16%). Telecommunications probably reflects a general unwillingness by some businesses to divulge information about its state of affairs. The low response rate indicates a larger problem: Although the response rate for the entire NCSS was higher than in any other study reviewed, the respondents comprise less than one quarter of the possible respondents.

Moreover, the survey’s definitions limit its interpretation. A careful, coherent definition of terms is essential for internal validity and therefore for proper analysis. In common parlance, “cybercrime” is used in numerous contexts, often including vandalism against computers (such as viruses), crimes against people (such as identity theft or cyber bullying), degradation of corporate Internet access (such as distributed denial of service), and more. The NCSS considers cybercrimes to be “crimes in which the computer system is the target,” “crimes in which a computer is used to steal money or things of value,” and any other incident involving a computer. These additional incidents include “spyware, adware, hacking, phishing, spoofing, pinging, port scanning, and theft of other...
information, *regardless of whether the breach was successful or damage or losses were sustained as a result.* Emphasis is added for two reasons:

1. The NCSS appears to focus on monetary loss and system downtime as the metrics for damage to a company, and
2. The counting rules capture assaults regardless of success, an element not always present in other surveys.

The NCSS’s definitions are not necessarily externally valid, because their counting rule is ambiguous. For instance, the nature of viruses makes it difficult to classify an incident’s characteristics: Does the presence of a single virus type, such as Sasser, constitute a single incident? After all, once a fix for a given virus type is identified, every other case of the virus can be neutralized with an easily repeatable solution. Or should the number of incidents reflect the number of infected computers? The latter counting scheme reflects the degree of IT support needed to rectify the problem. The distinction is important not only for the survey (different companies may have taken different approaches, thus rendering the values for the same question across respondents incommensurate), but also for its potential use in security investment decisions about virus-related problems: fixing one infected computer can be far less expensive than fixing many.

Attempts to gain system access are subject to the same criticism. Should a count of repeated attempts by a single entity reflect the number of attempts? Are multiple attack vectors employed by a single attacker considered individual or multiple incidents? What if they test the defenses of multiple assets? Should the system be viewed as a whole, or should subsystems be counted separately? Answers of both “yes” and “no” have valid arguments supporting them. But the ambiguity points to a need to make precise the terms and counting rules it uses, supporting the intention to which the data will be put. This ambiguity reflects the tradeoffs between simplicity of response and differentiation of result. For example, when potential responses are made simple (assuming that a more detailed survey is more laborious to complete and, because it is voluntary, risks overburdening the responder), the responders may be more likely to complete the survey. However, when the responses require more detail, smaller or more secure companies (i.e. fewer incidents) may be overrepresented in the set of completed responses.

The NCSS excludes many organizations that are not considered to be private firms: schools, publicly-run utilities, private households, and public administration groups, for example. This exclusion limits the applicability of the survey’s results to only a subset of organizations facing computer crime. Thus, the resulting dataset offers researchers and practitioners limited help in understanding whether certain attack methods are more prevalent in one kind of organization than in another.

The size and scale of the NCSS demanded considerable time for BJS’ analysis. Released in 2008, the NCSS includes data from reporting companies based on their individual positions in 2005. If typical, the time lag is worrying. The speed of innovation on the Internet suggests that reactions now to the findings from 2005 may be ineffective. For example, a report from 2008 indicates that, for the U.S., malicious code signatures increased by 23% over 2007 (Symantec Enterprise Security 2008). Similar reports suggest that companies operate in highly dynamic threat environments. For instance, a

---

3 Emphasis added.
Government Accountability Office (2007) report summarized the existing literature on data breaches and identity theft, indicating that while there is clearly an increase in the number of breaches occurring every year, it is difficult to know how that number translates into cases of actual identity fraud that cost both consumers and businesses money. Translating escalating attacks into monetary costs to inform resource allocation decisions is difficult even with up-to-date incident information; data from prior years have less value. Although the NCSS is a baseline for repeated collection, the trends may be old news by the time the analysis is released.

**Call and Response**

Both the unprecedented study size and its administration by a federal agency should ensure the NCSS a healthy response rate. However, all surveys suffer because the quality and accuracy of self-reported information varies. Companies are not legally obliged to respond, and some respondents may use less than adequate care in providing answers. Gathered data may not cover all categories of participants equally. Neither are all the responses likely to be equally precise. Sometimes, to help fill gaps or assess reasonableness, the basic responses can be supplemented with data collected elsewhere. For example, the Verizon Data Breach Investigation Report supplemented its survey responses with information collected during investigations of data breaches.

These problems highlight more general issues about data quality. For example, consider the incentives and disincentives for disclosing information outside the organization suffering an attack or loss. The impact of data breach laws has been analyzed, as has the effect of identity theft (Romanosky et al., 2008). Disclosing the existence of a data breach may have monetary costs and reputational effects that drive away potential customers. Sometimes, the incentive not to disclose is high enough that failure to report has been criminalized. The first such state law was enacted in California in 2002 (California Civil Code §1798.82); federal guidelines were stipulated in the Gramm-Leach-Bliley Act, requiring federally supervised institutions to report incidents to regulators (Pub. L106-102, V(A), 113 Stat. 1338 (Nov. 12, 1999)).

Of course, such conflicts could be independent of a single company’s business skill. At a given time, in a large collection of competitors, only some companies may experience a particular kind of attack. With each attack is associated a probability that a breach will occur; a rise in the number of attacks may increase the probability of a successful breach somewhere in the group of competitors. Some companies may value some types of information more than others; for instance, the loss of a telephone list (which can easily be restored from backups) has less impact than the loss of plans for the next new product design. An increasing probability of loss of something of great value could change corporate behavior (e.g. by causing implementation of more controls over highly-valued information). Moreover, when customers perceive some kinds of potential security problems (e.g. when their personal data may be disclosed), they may curtail or cease business with the company perceived to be putting their data at risk. Thus, disclosing items such as the level of attack activity may have effects similar to those of a full data breach. As part of its good business practice, each company assesses the value of each transaction as well as the likelihood of a successful attack. In this situation, two extremes bound the possible results. On one hand, it is possible that no company chooses to report a breach; then, the unaware customers do not change their behavior.
Alternatively, should all the companies be attacked and choose to disclose (all else being equal), customers have no good alternatives and are still not likely to alter their behavior; the reputation cost is still avoided.

However, these endpoints are not equally feasible options. Consider the case where all but company X choose to disclose security incident information. In this situation, X may realize some reputational benefit from not disclosing attack or breach information, giving it a relative advantage over its competitors. But it is unlikely that only a single company would make this choice. It is more likely that all companies are induced not to disclose. Thus, the all-disclosing equilibrium is tentative at best, without some additional enforcement rule. Had mandatory reporting been present (to raise the response rate to statistically acceptable levels), the NCSS could have estimated cybercrime impact by industry, evaluating trends later as the survey is repeated over time. Even informal decision making about security investment could be helped by more comprehensive rates for attacks and losses by industry. Writ large, such trends are analogous to health incident reporting used for tracking and responding to epidemics.

In addition to number and types of attack, the NCSS also gathers data on monetary loss. Other, similar reports express this information as the number of records compromised, but they rarely estimate the associated financial risk. Difficult to estimate, financial risk is needed for deciding how to balance the investment in security controls with the expected loss. Thus, the NCSS is an improvement over other data sources. But because the threat environment is constantly changing and assessment techniques are highly variable, great uncertainty must be associated with each monetary value in the survey response.

**Model Building**

In addition to making problems more visible, the NCSS data can support model building and use. Rue and Pfleeger (2008) discuss a range of published cybersecurity economic models that represent many approaches to modeling threats, vulnerabilities, and appropriate responses to attacks. Each approach can be assigned to categories that capture the salient features of the models and the problems they address. Table 4 summarizes the model evaluation framework. The left-most column itemizes the types of inputs models use in producing some measure supporting a security investment decision, and the column to its right contains several example inputs. The authors note that the current likelihood of getting accurate, credible information is poor for the first three input types but good for the second three.

Before using this framework to understand which models can use the BJS data, it is helpful to examine another feature of models. Each model type usually frames its variables and relationships in terms of an exchange between a single attacker (free to chose single or multiple methods of attack) and a defender. The preferred modeling methods borrow heavily from game theory and require probability estimates of various sorts. Thus, the models represent the activity of only a subset of the larger threat environment, because one attacker is a specific instance of a certain set of attacks or attack types. But the survey solicits information broadly across organizations, so data on cybercrime represent the larger set of circumstances (as when, for example, the organization experiences multiple, simultaneous attacks). Data capturing likelihood of a breach, value to the attacker of success, or costs of data loss are aggregates for which the
underlying probability distributions are unknown. Thus, predicting the characteristics of a single instance from data describing the larger environment presents a difficult challenge.

This problem, of attempting to make predictions at the individual’s level based on data collected at the group level, was recognized in an early study of voting behavior among women (Ogburn and Goltra, 1919) and was dubbed “ecological inference.” Not particular to cybersecurity, it has been addressed in other disciplines. For example, studies of election behavior (King 1997) and a medical treatment’s effectiveness (Freedman 1999) deal with this issue. Although statistical methods can moderate the effects of making the jump from the specific to the general, tension may still arise when moving back to the particular: trying to draw inferences from general data to particular organizations. In other words, if the data inputs are not representative to begin with, even the most sophisticated methods and models will provide unusable output.

Given these problems, in what models can the NCSS data be of benefit? We can add sample proxy measures and data sources for some of the inputs, as listed in the rightmost columns of Table 4. Values could be supplied in several ways, including internal organizational data collection, industrial surveys, government data collection, and more. The full set of data elements may not be available in a single place, but multiple collection and analysis efforts may be worthwhile if the resulting full data set is more appropriate for modeling and prediction. The cautions presented in Pfleeger et al. (2006) can act as guidelines for revising existing data collection efforts or initiating new ones.

Table 4: Model Input Types with Proxy Measures and Data Sources

<table>
<thead>
<tr>
<th>Input Type</th>
<th>Examples</th>
<th>Proxy Measures</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of attacks</td>
<td>• E [number of attacks per time period]</td>
<td>• [Weighted measure of # attacks reported by industry]</td>
<td>• Arbor Networks</td>
</tr>
<tr>
<td></td>
<td>• E [number of external attacks]</td>
<td></td>
<td>Worldwide Infrastructure Security Report</td>
</tr>
<tr>
<td></td>
<td>• E [number of internal attacks]</td>
<td></td>
<td>• NCSS</td>
</tr>
<tr>
<td></td>
<td>Pr [breach during a given time period]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pr [an external attack on a system succeeds]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pr [an internal attack on a system succeeds]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• E [number of reliability failures]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• E [number of security failures]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Aggregate Study Sample N reporting: have firewalls AND breach] / [Aggregate Study Sample N]</td>
<td></td>
<td>• NCSS</td>
</tr>
<tr>
<td></td>
<td>[Aggregate Study Sample N reporting: have firewalls AND breach] / [Aggregate Study Sample N]</td>
<td></td>
<td>• Verizon DBIR</td>
</tr>
<tr>
<td></td>
<td>[Aggregate Study Sample N reporting: have firewalls AND breach] / [Aggregate Study Sample N]</td>
<td></td>
<td>• DatalossDB</td>
</tr>
<tr>
<td>Vulnerabilities</td>
<td>• E [number of bugs in a given manufacturer’s component]</td>
<td>• Software patch rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• E [number of failures of a given component]</td>
<td></td>
<td>• Project Quant patch management cost modeling?</td>
</tr>
<tr>
<td>Costs</td>
<td>Cost of security</td>
<td>• Analysis of data from DataLossDB by type of failure/industry/lost data (financial, credit card, etc)</td>
<td>• Project Quant patch management cost modeling?</td>
</tr>
<tr>
<td></td>
<td>• Cost of deploying a security device</td>
<td></td>
<td>• DatalossDB</td>
</tr>
<tr>
<td></td>
<td>• Cost per unit effort of attackers and defenders for each point of vulnerability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of failure</td>
<td>Payoffs (not including security functions)</td>
<td>Cost of Data Breach estimates</td>
<td>Ponemon Institute</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------</td>
<td>------------------------------</td>
<td>-------------------</td>
</tr>
</tbody>
</table>
| • E [loss from a successful attack]  
• E [loss from loss of availability]  
• E [loss from loss of confidentiality]  
• Variance of loss from a successful attack | • Attacker’s payoff for a success  
• Penalty for attacker if caught | • Monetary loss/system downtime averages (assuming gain is symmetrical to loss) | • NCSS |

**Strategic Data Collection**

Every organization must consider its security posture, meet internal and external security standards, and set and enforce security policies. Security professionals, technology officers, financial officers, and organizational managers need frequent, good, consistent information to support decisions about technology, policy and process investments. Multiple data sources can paint a clear, detailed picture of the threat environment. The NCSS data can form the basis for decision-making but be supplemented with other data sets. Because the threat environment evolves, as does the set of responses to it, the data must be current. The NCSS should be repeated, but BJS should make a significant attempt to shrink the lag time between capture and reporting.

On the other hand, more is not always better. Amassing any and all data can paint a misleading picture, especially when it is not clear which data are applicable to a given situation or problem. It is important for organizations to develop criteria for determining when and whether a data set is relevant.

Many analysts are eager to mine open sources of data, not only for ease of access but also to reduce data collection costs. As new data sets become openly available, they can be evaluated for appropriateness and timeliness. In addition, the continual inclusion of new information in data sets such as DataLossDB can help to ensure timeliness and relevance (McNamara 2008).

However, there are significant problems with open source data. One need only reference the history of Wikipedia to see cases of politically charged submissions, erroneous data being proliferated, and outright maliciousness. Securosis runs Project Quant, a research effort studying software patch management strategies and costs. It has fielded an open survey via the Internet, requiring little or no pre-screening for submission (though Securosis researchers are reviewing data used in the analysis) (Securosis 2009). How credible are these data? Using convenience surveys to gather data from whoever is willing to provide it, the creators of some open source data sets have no sampling strategy; thus, we cannot determine the population the data represent. Even with submission screening, the voluntary nature of the data provision admits the possibility that a volunteer will “poison” the data to mislead or hide a dangerous or embarrassing situation. Moreover, there is no punishment if the poisoning is discovered, so an organization using the data relies on providers wholly outside their control. Additionally,
the data made available to one organization is available to all others; competition may
provide incentives to supply misleading data. And the open information is available to
potential attackers, too. Even knowing that an organization is making investment
decisions based on a particular data set may provide insight into resource allocation, thus
allowing a potential attacker to choose her methods to maximum effect.

Closed, private data acquisition also has costs and benefits. Obvious costs are
associated with data collection instruments and tools, a repository for data storage and
retrieval, extraction of information from external sources, and the personnel time and cost
associated with these efforts. But dedicated collection makes possible highly focused
sampling that is relevant to an organization’s particular interests. Although response rates
are still an issue, the organization can ensure consistency in definitions, relevance to
organizational needs, common understanding of purpose, and application of appropriate
measures. Further, the model guiding investment decisions can be kept private.

In reality, most organizations combine open and private data. The open data paint
a general picture that is then made more particular and relevant with private information.
The larger issue is the continued inability to define the threat environment. In part, this
problem reflects strategic choices about disclosure, as well as variance in the distribution
of threats faced by each organization.

**Recommended Next Steps**

BJS’s work in constructing, fielding, and analyzing the NCSS has raised the level
of general inquiry into cybercrime’s reach and impact. Its focus on economic loss
associated with system downtime provides measures important both for understanding
the potential threat and for communicating cybersecurity information to multiple
audiences of varying technical backgrounds. Moreover, the NCSS has given
cybersecurity a prominence on par with other criminal justice issues. With repetition, the
NCSS may support the derivation of national or business sector estimates of
cybersecurity impact, one of the rationales for its creation. But in its present form, the
NCSS may not satisfy the needs of those seeking to make routine security investment
decisions. Yearly reporting based on data from several years prior enables depiction of
trends but provides only a foggy window on the current cybercrime environment.

Even if security professionals make little use of the NCSS, national policy-makers
can use its findings, particularly about the information infrastructure. For instance, in
testimony before the U.S. Senate, Richard Power cited the “conventional wisdom” that
insiders perpetrate 80% of cybercrime; he then provided conflicting evidence (U.S.
Congressional Testimony, 1996). The NCSS’ definitions of incidents, loss by industry,
and attacks from various sources may provide more credible support about national
indicators to support policy decisions.

The disappointing NCSS response rate may prompt discussion about disclosure
and use of security data beyond breach notification. Credible, timely data can strengthen
the ability of consumers, governments, and industries to guard information. On a larger
scale, questions of national security arise. Power grids, transportation infrastructure,
health care systems, and financial data are critical to modern existence; all are potentially
vulnerable. Opaque security threatens more than just immediate safety; it poses a danger
to long-term competitiveness. Because basic research on security requires a broad view,
understanding the environment in which security investments are made requires trustworthy data. The NCSS, albeit tentative, is a step in the right direction.

References


