Experiences and Challenges with using CERT Data to Analyze International Cyber Security

Stuart Madnick
Xitong Li
Nazli Choucri

Working Paper CISL# 2009-13

September 2009

Experiences and Challenges with using CERT Data to Analyze

International Cyber Security

Stuart Madnick
Sloan School of Management
Massachusetts Institute of
Technology
Cambridge, MA 02139 USA
<smadnick@mit.edu>

Xitong Li
Sloan School of Management
Massachusetts Institute of
Technology
Cambridge, MA 02139 USA
< xitongli@mit.edu>

Nazli Choucri
Political Science Department
Massachusetts Institute of
Technology
Cambridge, MA 02139 USA
< nchoucri@mit.edu>

Abstract

With the increasing interconnection of computer networks and sophistication of cyber attacks, it is important to understand the dynamics of such situations, especially in regards to cyber international relations. The Explorations in Cyber International Relations (ECIR) Data Dashboard Project is an initiative to gather worldwide cybersecurity data publicly provided by nation-level Computer Emergency Response Teams (CERTs) and to provide a set of tools to analyze the cybersecurity data. The unique contributions of this paper are: (1) an evaluation of the current state of the diverse nation-level CERT cybersecurity data sources, (2) a description of the Data Dashboard tool developed and some interesting analyses from using our tool, and (3) a summary of some challenges with the CERT data availability and usability uncovered in our research.

Keywords: Cybersecurity, Computer Emergency Response Teams, Data Dashboard, Country Comparisons

1. Introduction

The growing global interconnection and interdependency of computer networks, in connection with increased sophistication of cyber attacks over time, demonstrate the need for a better understanding of the collective and cooperative security measures needed to prevent and respond to cybersecurity emergencies. In 1988, the United States Department of Defense and Carnegie Mellon University formed the Computer Emergency Response Team (CERT) to lead and coordinate national and international efforts to combat cybersecurity threats. Since then, the development of collective and collaborative cybersecurity has been officially underway for more than twenty years and the number of CERTs worldwide has grown dramatically, leading to the potential for a sophisticated and coordinated global cybersecurity response network.

The Explorations in Cyber International Relations (ECIR) project is a collaborative effort involving the MIT Political Science department, the MIT School of Management, the MIT Electrical Engineering & Computer Science department, the Harvard Law School, and the Harvard Kennedy School of Government. The ECIR Data Dashboard Project is an initiative to gather the cybersecurity data publicly provided by the worldwide, but individual, CERTs and

provide scholars, policymakers, IT professionals and other stakeholders with a comprehensive set of data on nation-level cybersecurity, information technology and demographic data to allow analysis and comparisons across countries. The Dashboard allows these stakeholders to observe transnational, chronological and multivariate trends and correlations that can lead to insightful explanations of the approximate causes and predictions of the potential future trends of global cybersecurity issues.

As a milestone, this paper presents three results that have been achieved so far: (1) it summarizes the current state of data from country-specific CERTs, (2) it introduces the ongoing Cybersecurity Data Dashboard project which was developed to help these stakeholders track potentially critical trends in relevant cybersecurity data, including attacks, threats, vulnerabilities, and defenses, etc. along with some sample analyses, and (3) it identifies some shortcomings regarding the availability and usability of CERT data and areas for future development. Increasing stakeholders' access to summary and analytical data should significantly enhance the efficacy of cybersecurity efforts at all levels, including individual and institutional defense, corporate and national policymaking, and high-level coordination and cooperation. The flexibility of the Data Dashboard is especially important in ECIR since it is a research experiment in interdisciplinary discourse and interaction and the different disciplines have different stakes in this Dashboard and the security issues.

The remainder of the paper is organized as follows. Section 2 presents a brief history, summary, and analysis of national-level CERT activities and their publicly available cybersecurity data. Section 3 introduces the ECIR Data Dashboard prototype and illustrates some analyses. Section 4 identifies some of the challenges that we have experienced during the investigation and development. Section 5 presents a number of related works and finally, section 6 concludes the paper.

2. Computer Emergency Response Teams (CERTs)

2.1 History and Purpose of CERTs

The first CERT, at Carnegie Mellon University (CMU), was launched in 1988 (Kimberland 2003) as a response to the Morris Worm (Orman 2003) attack which took down perhaps 10% of the Internet during November, 1988. The CMU CERT, during the 1990s, began to help other countries develop their own CERTs and maintains to this day a formal Computer Security Incident Response Team (CSIRT) development program, including for the United States. The CMU CERT is now officially known as the CERT Coordination Center (CERT/CC)[1], as many other response teams have chosen the name CERT (where others have chosen CSIRT). CERT/CC is the ordination hub for all global CERTs and is responsible for setting standards, best practices, and policies. Although the available statistics are not as detailed as nation-level CERTs, they are highly aggregated and serve as a useful indicator of global CERT effectiveness. CERT/CC works closely with US-CERT (Kimberland 2003), the latter of which is an indirect branch of the Department of Homeland Security.

---

[1] http://www.cert.org/

The CERT mandate is to develop and promote best management practices and technology applications to "resist attacks on networked systems, to limit damage, and to ensure continuity of critical services" (Software Engineering Institute 2008). CERT/CC works in the following fields which provide a guideline for the work of other national CERTs and CSIRTs around the world:

- *Software Awareness*: Searches for, receives, analyzes, and reports major software security vulnerabilities and malicious code. Publishes advice on responses to vulnerabilities and threats, and helps create software more secure to attack.
- *Secure Systems*: Engineering of networks that have high situational awareness and high response speed to deal with coordinated attacks. Goal is to create networks that can survive attack and continue functioning.
- *Organizational Security*: Encourages and helps develop implementation of proper security management and software in individual organizations, and advocates government policy that increases security of national, corporate, and private systems.
- *Coordinated Response*: Helps create and train response teams for different organizations, governments, and companies, including the Department of Homeland Security (US-CERT), and the National Computer Security Incident Response Team (CSIRT) of Qatar.
- *Education and Training*: Provides public training seminars, certification training/testing, as well as collegiate degrees at CMU.

The interconnected nature of modern computer networking assures that major failures in the security of a single institution have the potential to create larger damage to other institutions, or even large portions of the Internet. To solve the collective action problem, CERTs were designed with decentralization and coordination in mind. Ideally, the national CERTs would overlook and coordinate an array of CERTs at various levels below, such as industry sectors and individual companies. These lower level CERTs would work with each other under the auspices of the national CERT in order to offer both robust prevention and monitoring capability and a decentralized, distributed response to emergencies and attacks that may arise. This ideal configuration would lead to an efficient coordination between organizations ranging from semi-government, non-profit to private/corporate to ensure both collective and individual security.

2.2 Current Status and Breadth

In reality, the CERT security structure remains in its infancy in most countries that do have national CERTs, and the ideal CERT network (as explained above) is not even fully developed in the CERT's origin nation, the United States. Many countries do not have CERTs, but significant progress has been made over the past two decades in increasing the population of national CERTs and other CERT institutions, especially in countries with a large Internet user population or Internet-centric economy.

While there is no single authoritative and centralized list of national CERT programs, Table 1 lists the 54 countries that the authors have found[2]. There are certainly other countries with

---

[2] These countries are found from http://www.first.org/about/organization/teams/ and

some sort of cybersecurity teams, but the CERTs listed are more specifically national-level, cooperative, educating, and responsive organizations.

Most large enterprises, both local and global organizations, have dedicated IT security teams, some of which are called CSIRTs or even CERTs (but many of which are not). These enterprise-level cybersecurity teams are often used for solicited surveys from the CERTs for collecting incident information and are the points of contact for dissemination of best practices and threat alerts.

Table 1. Countries with National CERTs

Argentina	Australia	Austria	Bangladesh	Brazil	Brunei
Canada	Chile	China (PRC)	Croatia	Czech Republic	Denmark
Estonia	Finland	France	Germany	Greece	Hong Kong
Hungary	Iceland	India	Indonesia	Ireland	Israel
Italy	Japan	Latvia	Lithuania	Malaysia	Mexico
Myanmar	Norway	Pakistan	Philippines	Poland	Portugal
Qatar	South Korea	Russia	Singapore	Slovenia	Spain
Sri Lanka	Sweden	Switzerland	Taiwan (ROC)	Thailand	Tunisia
Turkey	UAE	United Kingdom	United States	Uruguay	Vietnam

2.3 General Data Availability from CERTs

Many of the national CERTs collect information on a number of cybersecurity issues in their countries by year, quarter, or month. Information collection, in general, is conducted by surveys: organizations voluntarily (although often by solicitation) disclose attack types, defenses, and shortcomings within the organization. In addition, some CERTs have performed data collection through passive probes in their national networks. CERTs often aggregate the data to present nationwide reports on the state of cybersecurity during the reporting period and trends over time. Some CERTs also ask institutions about their defense and security technology, as well as request self-criticisms by institutions of their security readiness for different types of attacks, and policies, and standards, etc.

The aggregated survey method has some interesting methodological artifacts that are worth noting. They are best illustrated by two examples: first, if a single virus hits 1000 institutions (and they all report), then that virus is counted 1000 times; second, if 1000 different viruses hit a single enterprise, an "incident" reporting method will lead to 1000 hits, while a "respondent" method will report only one hit, as a "respondent" method simply asks whether the respondent has experienced that specific problem (i.e., a virus attack) in the reporting period.

While each CERT is usually consistent between reporting periods[3], data consistency between CERTs is limited. CERTs have no standardized typology of data: their surveys ask different questions and create different categories of attacks and vulnerabilities. CERTs lack a consistent data presentation method: some present data in absolute numbers of reports, and others in percentages only. Term definition across CERTs is also sometimes inconsistent or unclear.

---

http://www.apcert.org/about/structure/members.html.
[3] Though definitions and collections techniques can change over time.

Although comparison and international aggregation can be difficult, there are a number of types of data that are commonly reported, in some form.

2.4 US-CERT

The United States national CERT (US-CERT)[4] is affiliated with the Department of Homeland Security, and is a distinctly different entity from CERT/CC at CMU. These two largest US CERTs share information and in case of a large-scale attack, will often coordinate extensively in leading a response. US-CERT is charged with providing response support and defense against cyber attacks for the Federal Civil Executive Branch and information sharing and collaboration with the state and local government, industry and international partners.

US-CERT provides the most comprehensive and detailed definition of terms, as explained: "A *computer incident* within US-CERT is, as defined by NIST Special Publication 800-61 (Grance et al. 2004), a violation or imminent threat of violation of computer security policies, acceptable use policies, or standard computer security practices." There are six categories regarding computer incidents used by US-CERT.

- CAT 1 – *Unauthorized Access*: An individual gains logical or physical access without permission to a federal agency network, system, application, data, or other resource.
- CAT 2 – *Denial of Service* (*DoS*): An explicit attempt by attackers to prevent legitimate users of a service from using that service.
- CAT 3 – *Malicious Code*: Successful installation of malicious software (e.g., virus, worm, spyware, bot, Trojan horse, or other code-based malicious entity that infects or affects an operating system or application).
- CAT 4 – *Improper Usage*: Violation of acceptable usage policies as established by the organization.
- CAT 5 – *Scans, Probes, or Attempted Access*: any activity that seeks to access or identify a federal agency computer, open ports, protocols, services, or any combination for later exploit. This activity does not directly result in a compromise or denial of service.
- CAT 6 – *Investigation*: Unconfirmed incidents of potential malicious or anomalous activity deemed by the reporting entity to warrant further review.

The above definitions of the categories of computer incidents are not shared universally by other CERTs, but certainly provide a relatively authoritative guide to what statistical data represents.

As an example, Figure 1 presents two charts regarding the computer incidents reported by US-CERT during the fourth quarter of 2008 (2008 Q4) (US-CERT 2008). Figure 1(a) breaks down the reported incidents by official US-CERT category. As can be seen, the greatest threat is "Scans, Probes & Attempted Access". Thus, the vast majority of threats reported to US-CERT are related to attempts to deceive the user (including phishing and malicious website), rather than direct attacks against the defenses of the computer or the network. Figure 1(b) is a breakdown of the top five incidents and events versus all others. The top incident type reported to US-CERT

---

was phishing, accounting for 71.8% of all incidents reported. Comparing the two charts, it could be concluded that phishing makes up the vast majority of the "Scans, Probes & Attempted Access" attacks (at about 77% of all incidents).
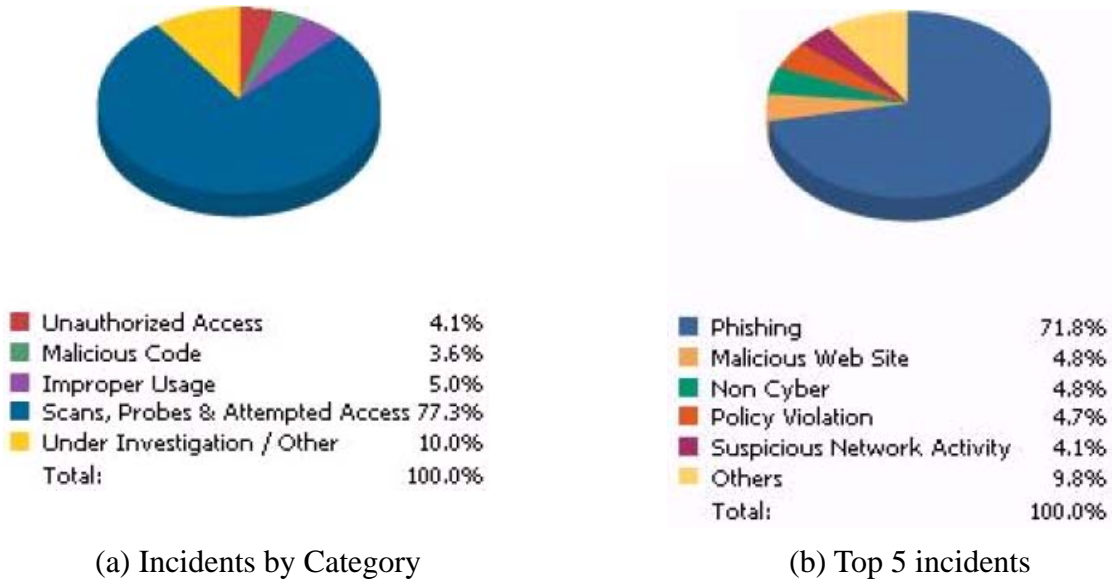

Unauthorized Access	4.1%
Malicious Code	3.6%
Improper Usage	5.0%
Scans, Probes & Attempted Access	77.3%
Under Investigation / Other	10.0%
Total:	100.0%

Phishing	71.8%
Malicious Web Site	4.8%
Non Cyber	4.8%
Policy Violation	4.7%
Suspicious Network Activity	4.1%
Others	9.8%
Total:	100.0%

(a) Incidents by Category (b) Top 5 incidents

Figure 1. Chart Examples of US-CERT Incidents in 2008 Q4

3. ECIR Data Dashboard Prototype

3.1 Purpose

The ECIR Data Dashboard is developed to provide historical trend data as well as current statistics and news to policymakers, academics, IT professionals and other stakeholders. By consulting the Dashboard, the user can compare trends in national-level cybersecurity threats and vulnerabilities among several countries and/or regions. Also, the user can use the Dashboard to compare these trends against other relevant national-level statistics to find patterns and correlations. To this extent, the Dashboard provides data in three categories:

- *Demographic Data*: Basic data about a country's population, economy, education level, and other attributes that may affect the development of the country's Internet services or IT security sectors. (Sources: World Development Indicators Database)
- *IT Data*: Data outlining the state of the country's IT infrastructure, usage, and security, including Internet bandwidth, users, services, etc. (Sources: ITU, World Development Indicators, CIA World Factbook)
- *Cybersecurity Data*: Data provided largely by national CERTs that reflect chronological trends of threat/vulnerability statistics.

The Dashboard allows the user to select any number of countries (and/or regions) with which to compare the data. The X-axis measurement is time (currently year), any of the data mentioned above can be selected for the Y-axis, allowing the user to compare correlations in multiple strands of data, each of which is for one country or region. Also, the Dashboard allows the user to divide one strand of data into another. This allows the user to compare the data in new ways. For example: dividing population into any measurement creates a "per capita"

6

measurement. Likewise, the user can compare the viruses reported per number of Internet users. Additionally, the user can select to graph the data on a linear or logarithmic scale, which can be very helpful in studying trends when comparing data, say from a large country with that of a small country (the small country data might appear as a straight line at zero on a linear scale). The Dashboard thus provides the user with a great amount of flexibility and power in finding exactly what data to compare, how to compare it, and how to illustrate it, so that multinational cybersecurity data can be deeply and robustly investigated.

3.2 Development

The Dashboard technology has been developed in three primary parts: Web user interface, database system, and recent news headlines, each of which is introduced as follows.

3.2.1 Web User Interface

The user interface is a Web application designed to query the database and creates charts of information on-the-fly. The user interface is implemented using the JavaServer Pages (JSP) Technology and Java programming. It provides a number of fields from which the user can select the countries/regions of interest, the X-axis variable (i.e., start year and end year for the observation), and the Y-axis variable (i.e., measurement data to observe) as well as the graphing type (linear or logarithmic). After the `Show Chart` button sends the request, the Dashboard application reads the requested data from the back-end database and draws the charts by automatically scaling the axis to reflect a "best fit" view of the data. The drawing feature is implemented by the JFreeChart (Gilbert and Morgner 2009), a set of Java APIs.
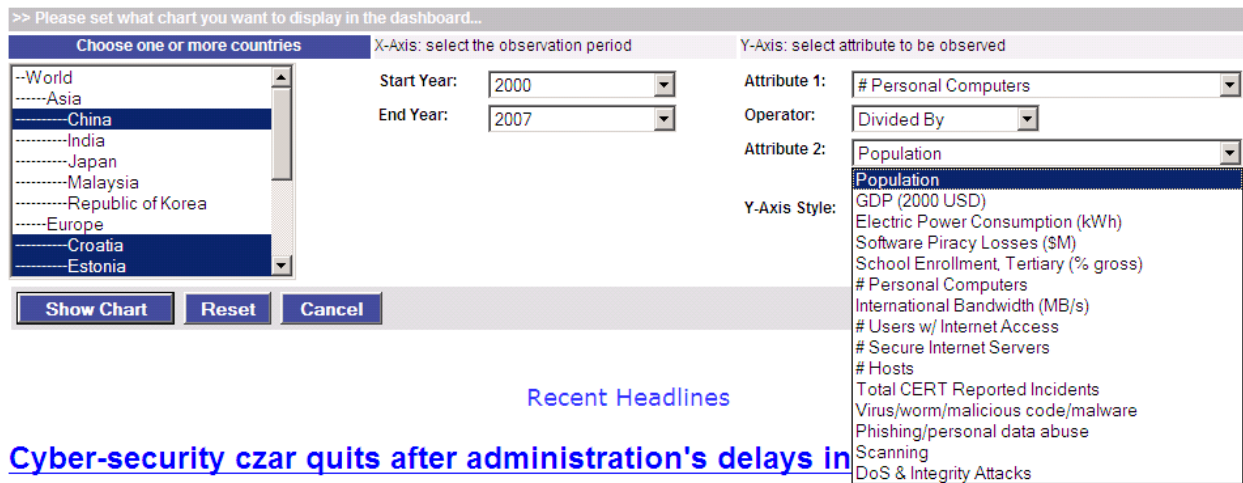

Figure 2. Example Request to Generate Chart of # Personal Computers per Capita

Figure 2 is a screenshot of the Dashboard configuration. As shown in Figure 2, a number of countries are listed in the left side. In the selection list, the countries are grouped into corresponding hierarchical regions. From the list, the user can select several countries (and/or regions) of interest. By selecting the start year and end year, the user can set the observation period. The Dashboard currently incorporates data from the chronological range from 2000 to

7

2008. In the right side, the user can select one or two attributes (i.e., measurement data). In case of two attributes, the user should also select an operator by which the data of interest can be calculated from them, such as the Division operator by which Attribute 1 is divided by Attribute 2 to be observed. The user can also set the Y-axis to a linear or logarithmic scale, which is particularly helpful when comparing data stands that different considerably in values, such as comparing large and small countries, as illustrated later. For example, Figure 2 is a request to display the number of PC per capita of three countries (China, Croatia, and Estonia) from 2000 and 2004.

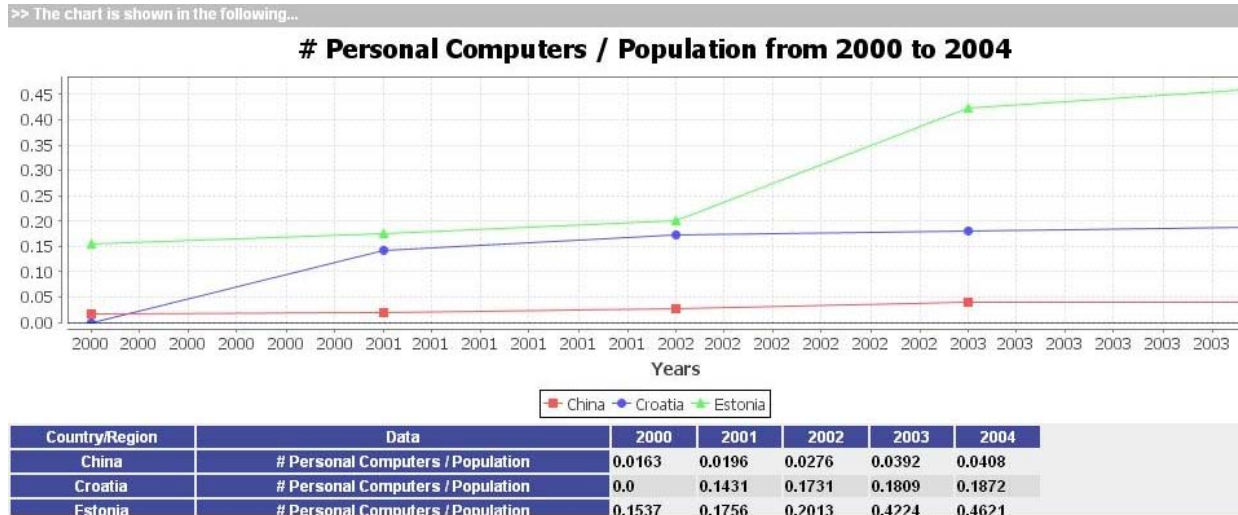| Country/Region	Data	2000	2001	2002	2003	2004
China	# Personal Computers / Population	0.0163	0.0196	0.0276	0.0392	0.0408
Croatia	# Personal Computers / Population	0.0	0.1431	0.1731	0.1809	0.1872
Estonia	# Personal Computers / Population	0.1537	0.1756	0.2013	0.4224	0.4621

Figure 3. Generated Chart of # Personal Computers per Capita

Figure 3 is the resulting screenshot from the Dashboard. For convenience and added precision, the actual data from the database is displayed in the table below the chart.

3.2.2 Database System

The back-end database of the Dashboard is the Palo MOLAP database (Raue et al. 2009). MOLAP stands for "Multidimensional On-Line Analytical Processing" which is an approach to quickly answer multidimensional analytical queries (Thomsen 2002). The Palo database uses a multidimensional data model (Pedersen and Jensen 1999), allowing multidimensional structures to organize data and express the relationships between the data. These structures are broken into cubes; the cubes are able to store and access data within the confines of each cube. Each cell within a multidimensional structure contains aggregated data related to elements along each of its dimensions. The output of a MOLAP query is displayed in a matrix format in which the dimensions form the rows and columns, and the relevant measurements form the data values. By using a MOLAP database, the Dashboard can quickly answer queries of any aggregated data, such as regional data. Palo consists of a mature MOLAP database server and an Excel add-in. Furthermore, JPalo[5] provides a set of Java API to manipulate the Palo database. These features make it an excellent choice as the back-end database of the Data Dashboard project.

In the current stage, there exists one cube with three dimensions in the Palo MOLAP

---

[5] http://www.jpalo.com/

database. The three dimensions are "Countries", "Years" and "Attributes". When the countries (and/or regions), years for the observation, and one or two attributes are determined, the corresponding measurement data can be accessed.

3.2.3 Recent News Headlines

To help the user to be informed of recent developments, the Dashboard also lists a number of high-relevance recent news headlines by using the MIT-developed Cameleon software [Madnick et al. 2000] which is a Web extraction engine designed to automatically extract data of interest from semi-structured documents (e.g., Web pages). In this example, the recent news articles listed by the Dashboard using the search terms "`cyber security OR computer spam OR cyber`" in Google News. The Dashboard displays the up-to-date news story snippets at the bottom of the user interface page, with hyperlinks that allow the user to open the full story in a new window or tab on their browser. For example, Figure 4 shows the recent news headlines listed by the Dashboard on July 17, 2009.

Figure 4. Dashboard Recent News Headlines (on July 17, 2009)

3.3 Current Status of Dashboard Prototype

The current status as of August 7, 2009, includes a working prototype of the Dashboard. Table 2 presents the current variables available in the Dashboard prototype. The countries available in the current Dashboard are: United States, China, India, Germany, Japan, Republic of Korea, Brazil, Estonia, Latvia, Croatia, Malaysia, and Australia. The current cybersecurity data availability of each category, by country, is presented in Table 3. In Table 3, "Prop." represents a source hosting proportional data; "Abs." represents absolute numerical data; "None" represents no data. Most data threads are not available for all the years of the Dashboard (2000 – 2008); many CERTs have only published quantitative data in the past few years; many have not yet released 2008 data. These represent some of the data challenges in performing multi-country cybersecurity analysis as discussed in more detail in a later section of this paper.

9

Table 2. Variables in the Data Dashboard

Demographic Data	IT Data	Cybersecurity Data
Population (#)	Internet User (#)	Total incidents (#)
Gross Domestic Product (USD)	International Bandwidth (MBps)	Phishing (#)
Software Piracy Losses (USD)	Personal Computers (#)	Trojan/worm/malware (#)
Energy Consumption (KWh/yr)	Hosts (#)	DoS (#)
Total Education Enrollment (%)	Secure Servers (#)	Spam (#)

The current database has some gaps in cross-time or cross-national CERT coverage. In the next phase, more extensive types of data and better sources of data are being sought. Also, both the number of countries and the categories of data will be significantly expanded in future versions of the Dashboard prototype. In spite of these shortcomings, there is sufficient CERT data available to perform some interesting initial analyses.

Table 3. Availability of CERT-based Cybersecurity Data by Country

Category	USA	China	India	Korea	Malaysia	Brazil	Germany	Japan	Estonia	Croatia	Latvia
Malicious Code	Prop.	Abs.	Abs.	Abs.	Abs.	Abs.	None	None	Prop.	None	None
Phishing	Prop.	Abs.	Abs.	None	None	None	None	Abs.	Prop.	None	Prop.
Scanning	Prop.	Abs.	None	None	None	Abs.	None	Abs.	None	None	None
Spam	None	Abs.	Abs.	None	Abs.	Abs.	None	None	Prop.	None	None
DoS	None	Abs.	None	None	Abs.	Abs.	None	None	None	None	Prop.

3.4 Interesting Demonstrations and Observations

To show the utilization, we provide several interesting examples using the Data Dashboard prototype.

Figure 5 shows a screenshot of the total CERT reported incidents of three countries (China, Malaysia, and Brazil) from 2003 to 2008. It can be seen that the number of the total CERT reported incidents of Brazil is much greater than that of China and Malaysia in almost of all years. Thanks to the logarithmic Y-axis scale for the chart, the data strands of the three countries are shown distinctly. If only the linear Y-axis scale could be used, the data strands of China and Malaysia would be pushed to the bottom of the chart because of the huge differences.

Figure 6 is a screenshot of "Virus/worm/malicious code/malware" divided by "Total CERT Reported Incidents" of two countries (Malaysia and Brazil) from 2002 to 2008 with logarithmic Y-axis scale in the chart. In other words, Figure 6 shows the data strands of the percentage of a category of the total reported CERT incidents, in this case, "Virus/worm/malicious code/malware".
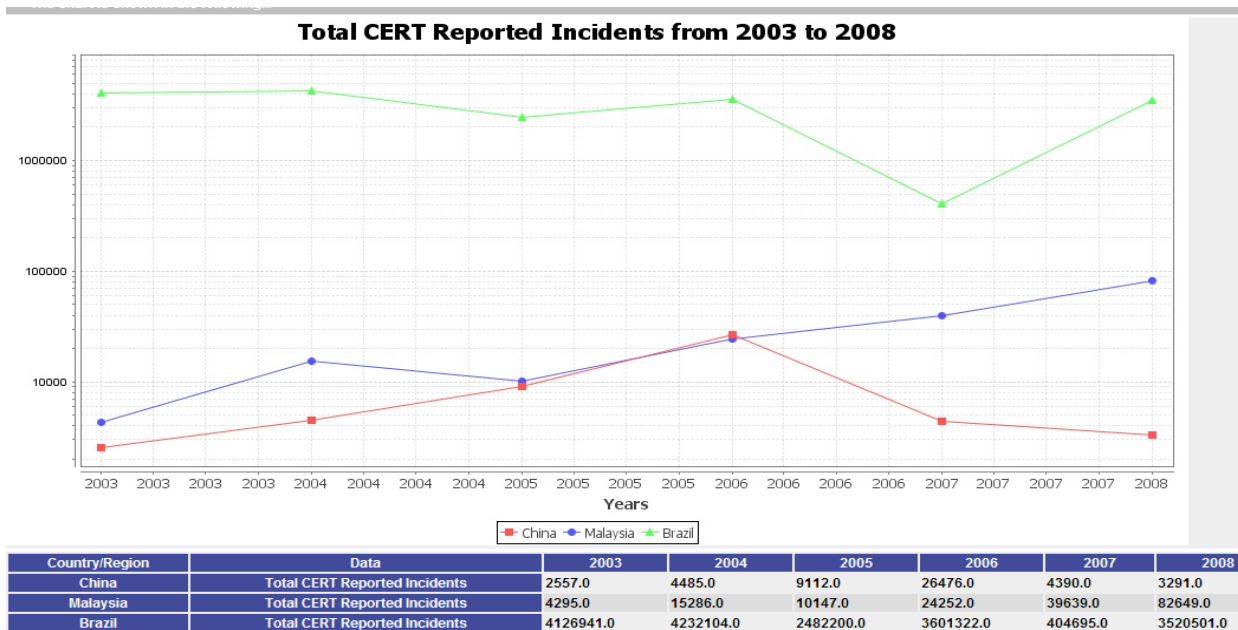
Figure 5. Total CERT Reported Incidents from 2003 to 2008 (Logarithmic)
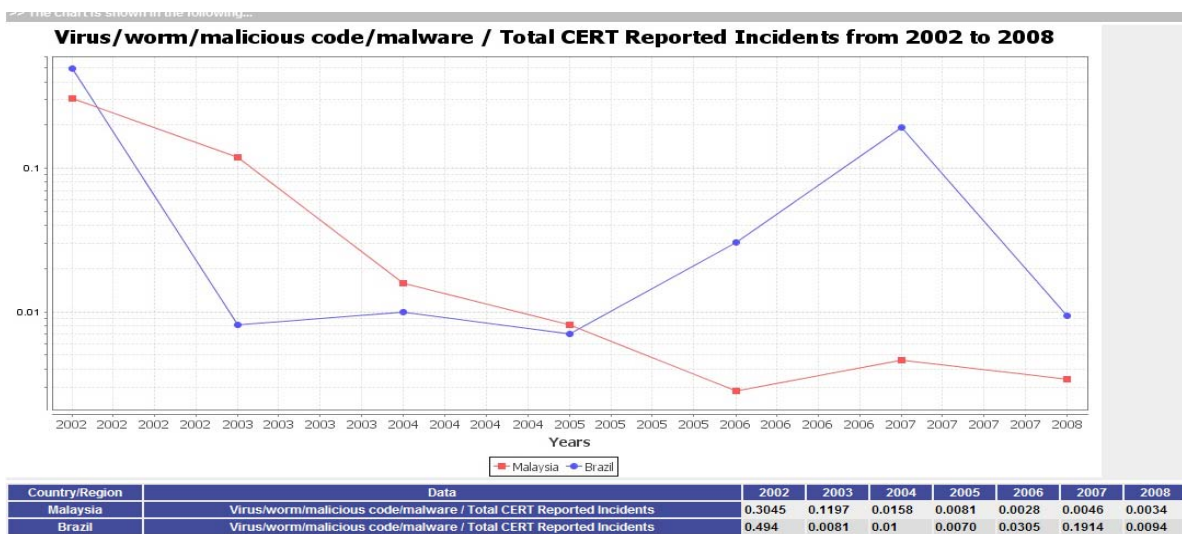

Figure 6. Percentage of Virus/worm/malicious code/malware from 2002 to 2008 (Logarithmic)

Figure 7 illustrates an interesting country comparison. It shows a screenshot of "Total CERT Reported Incidents" divided by "Population" (thus creating a per capita measurement) of two countries, Malaysia and Brazil, from 2003 to 2007 with a logarithmic Y-axis style in the chart. It is interesting that the per capita number of reported incidents started at very different levels (in 2003) – differing by about two orders of magnitude. But the rate has dropped sharply in Brazil while rising sharply in Malaysia such that they are at almost equal rates by 2007. This raises the intriguing question: "Why?" One set of hypotheses might be that (a) since Brazil had such a high rate in 2003, government and companies made extra efforts to reduce incidents and were successful, and (b) since there were so few incidents in Malaysia in 2003, no such efforts were

11

made and, as a result, incidents have been increasing. This just illustrates one possible set of hypotheses. A key goal of the ECIR research effort is to identify such interesting situations and to drill down deeper into the Dashboard data (as well as other sources of information) to develop hypotheses and theories to better understand cyber international relations – both within countries and between countries.

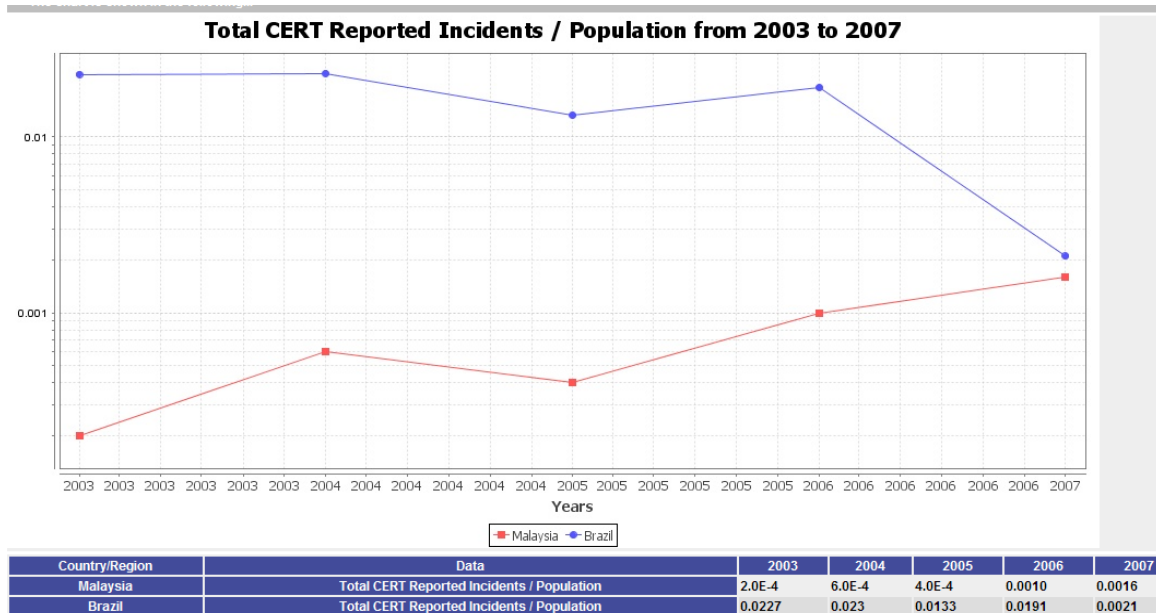| Country/Region	Data	2003	2004	2005	2006	2007
Malaysia	Total CERT Reported Incidents / Population	2.0E-4	6.0E-4	4.0E-4	0.0010	0.0016
Brazil	Total CERT Reported Incidents / Population	0.0227	0.023	0.0133	0.0191	0.0021

Figure 7: Total CERT Reported Incidents per Capita from 2003 to 2007 (Logarithmic)

It is worth noting that the independent national CERTs only provide the cybersecurity data regarding their own countries, but none of them provides extensive comparison data across different countries. This is an important contribution achieved using the ECIR Data Dashboard.

4. Challenges

A number of challenges and opportunities of using CERT data for discovery and improvement remain for the cybersecurity Dashboard project that has been uncovered in our efforts.

4.1 Data Sources and Data Availability

The data availability varies by category. In particular, it is sometimes difficult to find precise definitions of each category of cybersecurity data that are currently used by the CERTs. Although CERTs are often the primary sources of such cybersecurity data, many countries do not have national CERTs. Also, many national CERTs are new and do not provide much data, if any at all. The lack of data availability will continue to be a pressing challenge for exploring the cyber international relations. This will be addressed by working closer with the CERTs both to better understand their data and to encourage them to increase and improve their data efforts.

In addition to the CERTs, there are other national and international organizations that also collect relevant data – both public sectors (e.g., the FBI in the United States) and private organizations (e.g., Symantec, Arbor Networks, Microsoft, and McAfee). The Data Dashboard

project will seek cooperation and collaboration with these organizations.

4.2 Data Consistency and Reliability

Although many CERTs provide information on national threats and vulnerabilities, data consistency is a problem. Many CERTs have only started to record data within the last three or four years, leading to limited capacity for historical trend analysis. Furthermore, the lack of consistency between CERTs complicates the task of developing a robust framework to analyze cybersecurity data. CERTs often do not share categorization methods for threats and vulnerabilities, and although some general categories can be constructed successfully, they do not apply to all CERTs.

The surveys used by CERTs to collect data also vary greatly, rendering comparisons between countries difficult. For example, numerical comparisons can be misleading if the breadth of a survey is not explicitly clear; if both countries survey very different proportions of the population, the absolute numerical data will be incomparable. Additionally, even if survey respondents are relatively accurate, most respond on behalf of institutions on which disproportionate weights may be placed if response rates differ significantly. For example, it is unclear whether an incident at a large institution should be weighted the same as an incident at a smaller institution.

Lastly, most national CERTs are not mandated by federal governments. As a result, national CERTs tend to respond to incidents at their own discretion, and only handle a subset of the total number of national vulnerabilities and incidents. Thus, reported figures may not be indicative of the true volume of national domestic attacks. CERTs are also restricted from handling attacks on national defense and intelligence networks, so statistics concerning these types of incidents will likely remain unavailable for analysis.

These data consistency and reliability issues pose a challenge for the ECIR Dashboard project and addressing these issues will be a major focus of our future activities.

5. Related Works

Cybersecurity data sharing has emerged as an important issue in the recent years. While some organizations have regularly disclosed their cybersecurity reports, the reported data is usually difficult to be combined or compared because of various data inconsistencies. To facilitate the comparison of the cybersecurity data disclosed by different organizations, the Security and Networking Research Group at the Sandia National Lab and CERT/CC at CMU have developed a "common language" consisting of a set of "high level" terms and taxonomies that can be used to classify and understand cyber security incidents (Howard and Longstaff, 1998).

Despite the "common language", existing cybersecurity terms and categories are still diverse in both meaning and use. Since a large part of cyber threats and attacks are disseminated via the Web, a survey of Web security issues was offered which focused on particular areas of concern, such as server security, mobile code, data transfer and user privacy (Rubin and Geer, 1998). As can be seen, those Web security issues of particular interest were not classified or explored according to the "common language" developed in (Howard and Longstaff, 1998). For cyber intrusion detection, data mining techniques are usually applied to detect known instructions and

their variations as well as unknown instructions (Dokas et al. 2002). Yet further detection techniques need to be developed, so that cyber threats can be identified and detected in terms of a "common" taxonomy.

On the other hand, there still exist a large number of organizations that haven't initiated cybersecurity data-sharing actions. Partially this is because those organizations haven't realized the importance of cybersecurity data sharing. Thus, it is necessary to identify economic incentives to encourage the desired data sharing (Dacey 2003). Focusing on private organizations (e.g., firms, sectors), the work of (Gal-Or and Ghose, 2003) identified a number of economic incentives of sharing cybersecurity data within industry-based Information Sharing and Analysis Centers (ISACs). In particular, an analytical framework was developed to investigate the competitive implications of sharing cybersecurity data and investments in cybersecurity technologies.

Although the related research described above represent important efforts in understanding and improving cybersecurity data, they do not directly overlap with the unique contributions of this paper which are to: (1) use country-specific CERT data to develop a global database to compare trends across countries, (2) describe software to support the necessary analysis and discuss some interesting results, and (3) identify strengths and weaknesses of CERT data.

6. Conclusion

With the increasing global interconnection of computer networks and sophistication of cyber attacks, the need for collective and cooperative cybersecurity measures have been drawing more attention, so that cybersecurity emergencies can be prevented or efficiently responded. This paper first introduces the various CERT-type organizations that produce quantitative and qualitative cybersecurity data. Although much information is freely shared and non-confidential, few organizations provide standardized or reliable metrics and where data exists, it is rarely aggregated. This paper introduces the ECIR Data Dashboard Project which is developed to gather the cybersecurity data that are publicly provided by national/international CERTs and provide stakeholders with a comprehensive set of data on nation-level cybersecurity, IT characteristics and demographics. We then present the current state of the data availability of the national CERTs and their limitations. In spite of the identified shortcomings of the CERT data, it is possible to learn and identify interesting trends using the Data Dashboard.

In the future, more data sources will be identified and incorporated to increase the effectiveness of the Dashboard project, so that further interesting international cybersecurity relations can be investigated. Also, existing and advanced techniques, such as the utilization of semantic integration technologies, will be applied to address and improve the data inconsistency problems among different national CERTs. Furthermore, it may be necessary to explore the economic or social incentives of sharing the nation-level cybersecurity data among the national and international organizations (e.g., CERTs) as well as the data of other organizations, both public and private, that gather relevant cybersecurity data.

Acknowledgements

References

Dacey, R. 2003. "Information security: Progress made, but challenges remain to protect federal systems and the nation's critical infrastructures," *U.S. General Accounting Office* (GAO), GAO-36-564T, pp. 1-75.

Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J., and Tan, P.N. 2002. "Data mining for network intrusion detection," in *Proceeding of NSF Workshop on Next Generation Data Mining*, pp. 21-30.

Gal-Or, E., and Ghose, A. 2005. "The economic incentives for sharing security information," *Information Systems Research* (16:2), pp. 186-208.

Gilbert, D., and Morgner, T. 2009. "JFreeChart," (available from http://www.jfree.org/jfreechart/).

Grance, T., Kent, K., and Kim, B. 2004. "NIST Special Publication 800-61, Computer [NI04] Security Incident Handling Guide," *National Institute of Standards and Technology* (available from http://csrc.nist.gov/publications/nistpubs/800-61/sp800-61.pdf).

Howard, J.D. and Longstaff, T.A. 1998. "A common language for computer security incidents," Sandia Report: SAND98-8667, *Sandia National Laboratories* (available from www.cert.org/research/taxonomy_988667.pdf).

Kimberland, K. 2003. "U.S. Department of Homeland Security Announces Partnership with Carnegie Mellon's CERT Coordination Center," *Carnegie Mellon University Press Release* (available from http://www.sei.cmu.edu/newsitems/uscert.cfm).

Madnick, S., Firat, A., and Siegel, M. "The Caméléon Web Wrapper Engine," in *Proceedings of the VLDB Workshop on Technologies for E-Services* [SWP #4128, CISL #00-03], Cairo, Egypt, 2000, pp. 269–283.

Orman, H. 2003. "The Morris worm: a fifteen-year perspective," *IEEE Security & Privacy*, pp. 35-43.

Pedersen, T.B., and Jensen, C.S. 1999. "Multidimensional data modeling for complex data," in *Proceedings of the 15th International Conference on Data Engineering* (ICDE 1999), pp. 336-345.

Raue, K., Barczaitis, H., Wolff, T. et al. 2009. "Palo Manual," Jedox AG, 2009.

Rubin, A.D. and Geer Jr, D.E. 1998. "A survey of Web security," *Computer* (31:9), pp. 34-41.

Software Engineering Institute. 2008. "Meet CERT," Carnegie Mellon University, July 21, 2008 (available from http://www.cert.org/meet_cert/).

Thomsen, E. 2002. "OLAP solutions: building multidimensional information systems," Wiley Press.

US-CERT. 2008. "US-CERT Quarterly Trends and Analysis Report," (available from http://www.us-cert.gov/press_room/trendsanalysisQ408.pdf).