Global Mapping of Cyber Attacks

Ghita Mezzour, L. Richard Carley, Kathleen M. Carley

2014
CMU-ISR-14-111

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This work is supported in part by the Defense Threat Reduction Agency (DTRA) under grant HDTRA11010102, and the Army Research Office (ARO) under grants ARO W911NF1310154 and ARO W911NF0910273, and the center for Computational Analysis of Social and Organizational Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of DTRA, ARO or the U.S. government. The authors would like to thank Symantec for granting us access to the WINE IPS telemetry data. The cyber-attack data used in the paper is accessible through the WINE infrastructure using reference WINE 2012 004. The other country measures used in the paper are available from the authors upon request.

Keywords: intrusion detection, anti-virus, socio-technical factors, cyber security, empirical study

Abstract

Identifying factors behind countries' weakness to cyber-attacks is an important step towards addressing these weaknesses at the root level. For example, identifying factors why some countries become cyber-crime safe heavens can inform policy actions about how to reduce the attractiveness of these countries to cyber-criminals. Currently, however, identifying these factors is mostly based on expert opinions and speculations.

In this work, we perform an empirical study to statistically test the validity of these opinions and speculations. In our analysis, we use Symantec's World Intelligence Network Environment (WINE) Intrusion Prevention System (IPS) telemetry data which contain attack reports from more than 10 million customer computers worldwide. We use regression analysis to test for the relevance of multiple factors including monetary and computing resources, cyber-security research and institutions, and corruption.

Our analysis confirms some hypotheses and disproves others. We find that many countries in Eastern Europe extensively host attacking computers because of a combination of good computing infrastructure and high corruption rate. We also find that web attacks and fake applications are most prevalent in rich countries because attacks on these countries are more lucrative. Finally, we find that computers in Africa launch the lowest rates of cyber-attacks. This is surprising given the bad cyber reputation of some African countries such as Nigeria. Our research has many policy implications.

1 Introduction

Identifying factors that cause countries to become cyber-crime heavens or to become the main target of cyber-attacks. Such information can provide a sound basis for policy actions to address the problem at the root level. Prior work made observational comments about factors that impact cyber attacks in different countries. However, these factors are not the focus of this prior work. For example, Caballero et al. [7] suggest that fake anti-viruses target countries in Europe and North America because these countries are richer. Moreover, many international cyber security collaborations take the form of cyber security training [32] responding to an underlying assumption that local cyber security expertise shortage is a major problem.

In this paper, we take a first step towards empirically addressing 3 related research questions: (1) How does the prevalence of cyber attacks vary across countries, and what factors explain such variation?, (2) How does the number of attacks launched by computers vary across countries, and what factors explain such variation? and (3) How do attacks launched by computers in one country spread internationally, and which factors explain such spread? In our analysis, we use Symantec's World Intelligence Network Environment (WINE) Intrusion Prevention System (IPS) telemetry data set. WINE is a platform for repeatable experimental research through which researchers can access data used at Symantec Research Labs. The IPS is an end-host system that detects and blocks malicious network activity. The data contain attack reports collected from more than 10 million Symantec customer computers worldwide over the time period November 2009 - September 2011. An attack report contains the IP address of the victim computer, the IP address of the attacker computer, as well as information about the nature of the attack detected.

As the IPS exclusively examines network activity, the main attack types in the data are exploits, web attacks and fake applications (mostly fake anti-viruses). We find that web attacks and fake applications are most prevalent in developed countries. Our analysis confirms that attackers target these countries because of the large computing and monetary resources in these countries. On the other hand, exploits are most prevalent in countries with emergent economies. Computing and monetary resources also have a positive impact on exploit prevalence. Another interesting finding is that computers in Eastern Europe launch the highest quantities of attacks on average. Eastern Europe is preferred for hosting attacking computers because of a combination of good computing infrastructure and high corruption rate. Fast computers with high Internet bandwidth are preferred because these computers can aggressively serve attacks. Moreover, the high corruption levels facilitate cyber criminal activity such as registering malicious web sites and keeping attacking computers up despite complaints.

Unfortunately, if these countries continue to excessively host attacking computers, IP addresses from these countries may become blocked in bulk and users in these countries may see themselves virtually blocked from parts of the Internet. It is important to improve cyber security practices in these countries for the benefit of users in these countries and worldwide. As these countries would also benefit from addressing the problem, this paper advocates a soft power solution. In such a solution, countries collaborate because they perceive such collaboration as attractive, rather than because they are coerced into collaborating.

The remainder of the paper is organized as follows. We provide background in Section 2 and discuss related work in Section 3. We describe our data in Section 4 and threats to validity in Section 5. We discuss countries' exposure to attacks in Section 6, countries' hosting of attacking computers in Section 7 and the international cyber attack networks in Section 8. We discuss future work in Section 9 and conclude in Section 10.

2 Background

2.1 Cyber Attacks

The main attack types in the IPS telemetry data are exploits, web attacks and fake applications. We review these attacks in this section.

Exploits Exploits are malicious programs that take advantage of software vulnerabilities in the operating system, Java or other programs[1]. Some of the worst exploits enable an attacker to run arbitrary code on the victim machine without the user being aware of the attack.

Web attacks Web attacks are exploits on web browsers or web browser plugins. A victim encounters a web attack upon visiting a malicious website that launches the web attack. The victim may directly visit the malicious website, or may be directed to the malicious website after visiting a hacked webpage that contains iFrames or malicious java-script. The redirection is typically transparent to the user. Web attacks are typically used to deliver malware within the context of "drive-by-downloads".

A Pay-Per-Install (PPI) business model [7, 16] to deliver malware has emerged around web attacks and drive-by-downloads. In this model, we find clients, PPI providers and affiliates. Clients have malware that they are interested in disseminating. For example, clients can be the people that write such malware. Clients pay PPI providers to distribute their malware to victim computers, and pay providers by the number of victim computers on which the malware is installed. The rate ranges from \$100-\$180 per 1000 computers in the United States and the United Kingdom to \$7-\$8 in less demanded regions such as some Asian countries [16]. PPI providers are responsible for managing malicious web sites and directing web traffic to these websites. In some cases, PPI providers outsource some of these tasks to affiliates.

Fake applications Fake applications are applications that pretend to have a useful utility, but offer no utility or are malicious. The most common fake applications in the IPS telemetry data are fake anti-viruses. Fake anti-viruses falsely claim to find malware on the victim's computer and ask the victim to pay a premium to remove the malware. Some victims fall for the trick and pay the premium. Fake anti-viruses may also install additional malware on the victim's computer. Fake anti-viruses reach users mainly via two channels. Users download fake anti viruses manually thinking they provide free anti-virus protection. Alternatively, fake anti-viruses are distributed as part of drive-by-download attacks.

2.2 Factors Impacting the Number of Cyber Attacks Encountered per Computer

We present factors that may impact the number of cyber attacks encountered by computers in different countries.

Web visits The more webpages a user visits, the more likely is the user to encounter an attack. For example, the user may visit a malicious webpage that launches a web attack. Alternatively, the user may see an advertisement for a fake application and be tempted to download the fake application. Finally, the user may download a malicious video or PDF file that contains an exploit. It is worth noting that not all attacks require users to visit webpages. For example, many exploits can reach any machine connected to the Internet.

[1] Following Symantec's naming conventions, we refer to exploits on web browsers or web browser plugins as web attacks, and discuss them separately

Computing resources Attackers may prefer to attack fast computers with high Internet bandwidth in order to use these computers to launch other attacks. For example, such computers can send more spam messages and more Denial of Service packets. Therefore, computers in countries with large computing resources may encounter more attacks.

Monetary resources The majority of cyber attacks nowadays have a monetary goal. Attackers may target richer countries in order to make larger profits. For example, stealing credit card information of people in rich countries is more profitable. Similarly, people in rich countries are more likely to be able to afford paying a premium for a fake anti-virus.

Cyber security research Expertise gained in cyber security research may transfer to cyber security practitioners and end-users. Such expertise might reduce exposure to attacks. For example, cyber security expert users are less likely to open suspicious emails and click on malicious links. Similarly, cyber security expert IT administrators are more likely to patch systems and correctly configure firewalls.

Cyber security institutions and policy Computers in countries that have cyber security institutions e.g. CERT and policy may encounter less attacks. For example, through training and awareness programs, such institutions may help improve cyber security practices and reduce exposure to attacks.

International relations International relations may affect the number of cyber attacks encountered. For example, a country involved in inter-state conflicts may be the target of cyber attacks as was the case of Stuxnet [37]. Similarly, countries are usually less likely to attack their allies, and thus countries with many allies may experience less cyber attacks. It is, however, worth noting that cyber attacks among military allies have been reported [42].

2.3 Factors Impacting the Number of Cyber Attacks Launched per Computer

Computing resources Attackers may prefer to use fast computers with high Internet bandwidth as bots or malicious web servers. As a result, we expect computing resources to have a positive impact on the number of cyber attacks launched.

Number of cyber attacks encountered As a computer encounters more attacks, the computer is more likely to become infected and start launching attacks. Therefore, the number of attacks launched is likely to increase with the number of attacks encountered.

Cyber security research Cyber security expertise may decrease the number of attacks launched by computers in a country. Cyber security expertise of users and IT administrators may reduce the likelihood that computers become infected and start launching attacks. On the other hand, cyber security expertise may also increase the number of attacks launched. Expert hackers can more easily maintain botnets and malicious servers used to launch attacks. It is worth noting that attackers can perform such tasks remotely, and thus cyber security expertise in a country may not necessarily increase the number of attacks launched by computers in *that* country.

Cyber security institutions and policy The presence of cyber security institutions and policy in a country may help improve cyber security practices in a country, which may reduce the number of attacks launched.

Corruption Corruption facilitates criminal activity through the complicity of ISPs and law officials. For example, registering malicious websites and keeping malicious computers up despite complaints is easier when there is high corruption.

International relations Attackers may be discouraged from hosting malicious computers to attack computers in country V, in a country A that collaborates with V on cyber security issues. Such collaboration may be based on formal agreements such as extradition treaties or informal agreements [25]. Informal agreements may be easier among military allies, and harder among military enemies.

2.4 Factors Impacting the Inter-Country Cyber Attack Network

We present factors that impact the inter-country cyber attack network which represents the number of attacks that a computer in attacker country A launches on a computer in victim country V.

Country attributes Countries' attributes discussed in Section 2.2 likely have an impact on countries' indegree in the cyber attack network similar to the impact discussed in that section. Similarly, countries' attributes discussed in Section 2.3 likely have an impact on countries' outdegree similar to the impact discussed in that section.

Geographical proximity Some attacks use propagation strategies that favor geographically close computers. Thus, we expect to see more attacks among neighboring countries. For example, some forms of random scanning favor local computers. Attacks that spread through email and social media are another example since email and social media contacts are more likely to be at a close geographical distance.

2.5 MrQAP Regression

MrQAP regression [23] is a regression technique suitable for network data. Network data violates the independence assumption required for Ordinary Least Squares (OLS) regression. MrQAP regression on networks produces the same regression coefficients than OLS regression on the vector representation of these networks. The vector representation of a network is obtained by concatenating the rows from that network. However, contrary to OLS, MrQAP produces accurate p-values that account for intra-column and intra-row dependence in network data. In order to find these p-values, the MrQAP regression leverages the Quadratic Assignment Procedure (QAP) test, which is a non-parametric test based on random permutations of rows and columns.

3 Related Work

Most prior empirical cyber security work is interested in characterizing the mode of operation of attack campaigns. Unfortunately, this line of research is mostly uninterested in empirically testing hypotheses about factors that impact the number of cyber attack in different countries. For example, we find studies on spam [35], denial-of-service attacks [31], pay-per-install [7] and exploit-as-a-service [16]. We also find empirical work that aims at exposing malicious or negligent Internet Service Providers (ISPs) [40, 22, 21] that offer bullet-proof hosting to cyber criminal activities. However, such work overlooks testing hypotheses about factors that cause malicious ISPs to emerge in some geographical regions more than others.

Figure 1: Attack report generation

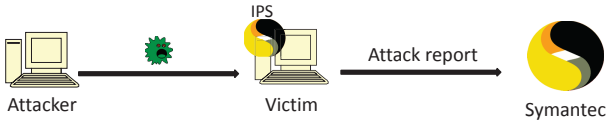


Table 1: Attack report example

Field	Value
Attack name	Web Attack: Adobe Flash CVE-2011-2140
computer ID	AB:12:35:DC:02:EA
IP address victim	172.268.12.156
IP address attacker	157.23.56.589

Finally, researchers [47, 34, 12] have reviewed international institutions involved in cyber-security. Examples of such organizations are the Community Emergency Response Teams (CERTs), the United Nations, the Organization for Economic Co-operation and Development (OECD). Such institutions respond to occurring cyber attacks, raise awareness about best cyber security practices, coordinate cyber security training and help set cyber security policies.

4 Data

4.1 Cyber Attack Data Sets

World Intelligence Network Environment (WINE) telemetry Intrusion Prevention System (IPS) data Symantec's WINE IPS telemetry data consist of attack reports sent by more than 10 million Symantec customer computers worldwide during the period November 2009-September 2011. The IPS is an end-host system that monitors the host's network activity. Upon detecting a malicious activity, the IPS blocks that activity and sends an attack report to Symantec as illustrated in Figure 1. An attack report contains the name of the attack detected, the IP address and unique identifier of the victim computer [2] and the IP address of the attacker computer as illustrated in Table 1.

It is important to note that the number of attack reports a computer sends depends on the number of attacks the computer encounters, but does *not* depend on the user's diligence about updating the attack signatures. Symantec uses automatic signature updates. All Symantec computers that are online obtain signature updates at approximately the same time. If a computer is offline when new signatures are released, the computer obtains these signatures as soon as the computer is online.

Attack catalog The attack catalog contains structured descriptions of attacks reported in the IPS WINE telemetry data set. We extracted this catalog from Symantec's online attack descriptions [41] in our prior work [28]. The catalog contains the attack name, the attack family name, the type, the pre/post infection attribute and the attack infrastructure type. The attack name is the name used by Symantec to uniquely identify the attack. The attack family name is a generalization of the attack name that we associate with the attack name. Type is the attack type. Examples of attack types in the catalog are web attacks and

[2] In this paper, we only consider attack reports where the Symantec computer sending the attack report is the victim computer. This is the case for more than 96% of attack reports. We disregard attack reports where the Symantec computer sending the attack report is the attacker computer. These attack reports are a minority, and are not representative

Table 2: Examples of attack catalog entries

Entry	Field	Value
1	Attack name	Attack: MS SQL Server 2000 Resolution Server 2000 Resolution Service CVE-2002-0649
	Attack family name	Attack: MS SQL Server 2000 Resolution Server 2000 Resolution Service CVE-2002-0649
	Type	exploit
	Pre/post infection	pre-infection
	Attack infrastructure	exploiting computer
2	Attack name	Web Attack: Blackhole Toolkit Website
	Attack family name	Blackhole
	Type	Web attack
	Pre/post infection	pre-infection
	Attack infrastructure	Malicious web page
3	Attack name	Fake App Attack: FakeAV Executable Download
	Attack family name	FakeAV
	Type	Fake application
	Pre/post infection	pre-infection
	Attack infrastructure	Malicious web page
4	Attack name	HTTP W32 Waledac Activity 3
	Attack family name	Waledac
	Type	worm
	Pre/post infection	post-infection
	Attack infrastructure	unknown

exploits. The pre-post infection attribute indicates whether this is an attack attempt that has been blocked (pre-infection), or that the computer is already infected (post-infection). All web attacks and exploits are pre-infection attacks, whereas we find both pre-infection and post-infection attacks among the other attack types. The attack infrastructure type is the type of malicious computer that launches the attack. Examples of attack infrastructure types are malicious web page, hacked web page and exploiting computer. We extract the type, pre-post infection attribute and attack infrastructure type based on keywords in the attack name and online description.

Table 2 provides examples of entries in the attack catalog for various attack types. The first entry is about an exploit. The attack family name is simply equal to the attack name since we are unable to infer a specific attack family from the attack name or online description. This is a pre-infection attack in the sense that the exploit is blocked and prevented from affecting the victim computer. The attack infrastructure type is an exploiting computer which indicates that the attacker computer is a computer that launches the exploit on the victim computer, but that we have no further information about such computer. Such entry is representative of the vast majority of exploit entries in the catalog. The second entry is about a web attack. The infrastructure type is a malicious web page. We consider malicious web pages to be web pages that deliver attacks. This is in contrary to hacked web pages that contain iFrames or malicious javascript and direct to malicious web pages. For the vast majority of web attack entries in the catalog, the attack infrastructure is equal to malicious web page. The third entry is about a fake application. The attack family name is FakeAV, which simply indicates that the attack is about a fake anti-virus. We are unable to infer a specific attack family name from the attack name. The fourth entry is about a worm. Here, the attack family name is Waleda, the worm family. This is a post-infection entry in the sense that the victim computer is already infected by the worm. The attack infrastructure type is unknown since we are unable to identify the type of activity performed by the worm from the attack name and online description. Other attack types in the catalog include adware/spyware, backdoors and trojans.

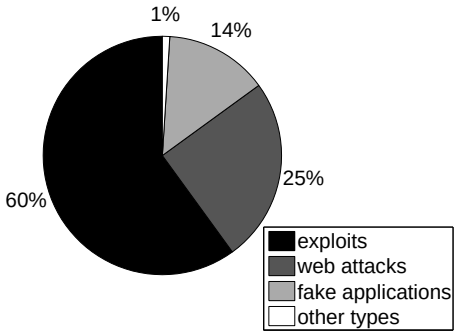


Figure 2: Distribution of attack types

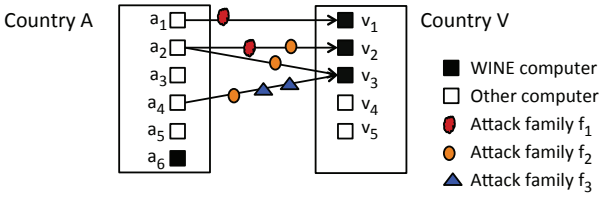


Figure 3: Toy example.

Figure 2 presents the distribution of attack types in the IPS telemetry data set. The main types are exploits, web attacks and fake applications. Other types such as worms, adware/spyware and trojans constitute 1% of attack instances. Exploits, web attacks and fake applications are the most prominent types because the IPS only examines a host's network activity. Threats that have no network activity are not detected by the IPS.

4.2 Country Cyber Attack Data

In this section, we explain how we compute the number of attacks encountered and attacks launched per computer in each country. We also explain how we compute the inter-country cyber attack network.

We start by explaining how we define an *attack instance*. We consider an attack instance (a, v, f) to be an attack by an attacker computer a on a victim computer v using attack family f. In other words, we consider all attack reports about an attacker computer a, a victim computer v and an attack family f to be a single attack instance. We choose to define an attack instance this way instead of simply considering each attack report to be an attack instance because an infected computer may send multiple attack reports about the same infection over time. In the toy example in Figure 3, there are 6 attack instances (a_1, v_1, f_1), (a_2, v_2, f_1), (a_2, v_2, f_2), (a_2, v_3, f_2), (a_4, v_2, f_2) and (a_4, v_3, f_3). It is worth noting that our definition of attack instance may underestimate the actual number of attacks, as is the case when we count (a_4, v_3, f_3) only once. We choose such definition in order to have a consistent measure across all attack types.

We define a *WINE computer* as a Symantec customer computer whose attack reports are included in the WINE telemetry data. The nearly 10 million WINE computers are randomly chosen from all Symantec customer computers worldwide. We determine the country where a computer is using IP geolocation [26] and only keep data from countries with at least 30 WINE computers. That is, we ignore data from North Korea, Nauru, Guinea-Bissau, Tuvalu, Eritrea, Cuba and Kiribati. In the remaining countries, the ratio of WINE computers out of the total number of computers in the country has mean 1.30% and standard

deviation 0.20 10^{-3}. We obtain the total number of computers in a country based on estimates of the number of computers by 100 people [43] in each country and estimates of the population size in these countries [44].

Attacks encountered per computer This is the average number of attack instances encountered by a WINE computer in each country. In the toy example in Figure 3, the 3 WINE computers in V encounter 6 attack instances. Therefore, the attacks encountered per computer in V is equal to $6/3 = 2$.

Attacks launched per computer This is the average number of attacks launched by a single computer in each country. For a given country, we count the number of attack instances where the attacker computer is in that country. We then divide by the total number of computers in that country. We divide by the total number of computers, and not just the number of attacker computers in the WINE data because the WINE data record attacks by *any* computer on WINE computers. In the toy example in Figure 3, the 6 computers in A launch a total of 6 attack instances. Therefore, the attacks launched per computer in A is $6/6 = 1$.

Cyber attack network This is a country by country network that represents the average number of attack instances from a single computer in country A on a single WINE computer in country V. We count the number of attack instances where the attacker computer is in A and the victim computer is in V. Then, we divide that number by the product of the total number of computers in A and the number of WINE computers in V. In the toy example in Figure 3, we obtain the edge weight from A to V by dividing 6, the number of attack instances by 18, the product of the total number of computers in A and the number of WINE computers in V. The edge weight from A to V is thus $1/3$.

In addition to computing the attacks encountered per computer, the attacks launched per computer and the cyber attack network when taking into account all attack instances, we also compute these measures for particular attack types. For example, we compute the exploits encountered per computer by only taking into account attack instances that have attack type "exploit". One exception is that when computing measures and networks for web attacks and fake applications, we only keep attack instances from the corresponding attack type that have "malicious web page" as attack infrastructure. We disregard attack instances that have "hacked web page" as attack infrastructure because these attack instances are a minority, and the factors that affect the likelihood that hacked web pages appear are different from factors that affect the likelihood that malicious web pages are hosted.

4.3 Explanatory Variables

In this section, we present data we use to measure explanatory factors discussed in Sections 2.2, 2.3 and 2.4. When applicable, we present more than one way to measure certain factors.

Web visits We collect from Alexa [3] the list of the top 500 websites in each country, obtaining a total of 28,660 websites. For each website i, Alexa provides the global ranking r_i of that website and the percentage of visitors p_{ij} to that website from each country j. We infer the total number of visits to a website by assuming that the number of visits to websites follows a power law distribution with exponent $\alpha = -2.07$ [1], which implies that the number of visits vs_i to i is proportional to $r_i^{-\alpha}$ [1]. Thus, we find that the number of web visits from users in j is proportional to $\sum_i p_{ij} vs_i$.

Cyber security research We estimate cyber security research strength in a country by counting the number of cyber security research papers published by that country during the period 2002-2011. We include papers starting from 2002 as expertise gained in research requires time to transfer to industry and the general public. We collect from SCOPUS [38] all papers published during the time period 2002-2011 in conferences and journals that contain "security" in their title and that belong to the engineering or computer science areas. We obtain 28,400 papers. For each country, we count the number of papers that have an author affiliated in that country.

Cyber security institutions and policy We construct a binary variable about which countries have cyber security institutions and policy. We obtain the list of such countries by combining lists from multiple sources [24, 11, 18]. It is worth noting that the strength of institutions and policies may be relevant beyond their simple existence. However, systematically measuring such strength in all countries is difficult.

Monetary resources We measure how rich people are in different countries using the GDP per capita [44, 10].

Computing resources Relevant computing resources mainly consist of Internet bandwidth and computer speed. We use the Internet bandwidth per Internet user indicator from the International Telecommunication Union (ITU) branch of the United Nations [19]. Measuring average computer speed in a country is difficult. As a proxy, we use the Information and Communication Technology (ICT) development index from the ITU [19]. This index is computed by combining 11 indicators: international Internet bandwidth (bits/s) per Internet user, percentage of households with a computer, percentage of households with Internet access, percentage of individuals using the Internet, fixed (wired)-broadband Internet subscriptions per 100 inhabitants, active mobile-broadband subscriptions per 100 inhabitants, fixed-telephone lines per 100 inhabitants, mobile-cellular telephone subscriptions per 100 inhabitants, adult literacy rate, secondary gross enrollment ratio and tertiary gross enrollment ratio.

Corruption The World Economic Forum collects a bribes indicator [15] by sending a questionnaire to a large number of business executives about how common is it for firms to make undocumented payments or bribes connected with imports and exports, public utilities, annual tax payments, awarding of public contracts and licenses, or obtaining favorable judicial decisions. Countries where bribes are common score *low* on the bribes indicator.

Attacks encountered We use the attacks encountered per computer measure presented in Section 4.2.

International relations International relations are more naturally represented as networks. We include international relations in regressions of attacks encountered and attacks launched by using betweenness centrality in these networks.

We collect the list of international military and non-military conflicts during the period 1992-2010 [9, 14]. We compute a binary country by country network $H = [h_{ij}]$ that captures the existence of a hostility between two countries i and j. In other words, $h_{ij} = 1$ indicates a conflict between i and j, and $h_{ij} = 0$ indicates otherwise. We collect the list of military alliances [13]. The list covers defense pacts, neutrality and non-aggression pacts, and entente pacts. Based on that list, we compute a binary alliance network $A = [a_{ij}]$ that captures the existence of an alliance between i and j. Finally, we use the list of extradition

Table 3: Geographical regions [33]

Abbreviation	Region name	# countries
Africa	Africa	51
Asia-Pc	Asia-Pacific	50
E. Eur.	Eastern Europe	23
Lat. Am.	Latin America & Caribbean	30
W. Eur.+	Western European & others (Canada, Israel, N. Zealand, Turkey and the US)	30

treaties from the United Nations Crime and Justice Information network [46] in order to construct a binary country by country extradition network $E = [e_{ij}]$ [46].

Regional membership We use the geographical subdivision of countries from the United Nations regional group membership [33] summarized in Table 3. The table only considers countries that have more than 30 WINE computers. We compute a country by country geographical membership network $G = [g_{ij}]$, where $g_{ij} = 1$ if countries i and j are in the same geographical region, and $g_{ij} = 0$ if i and j are in different geographical regions.

We use data from year 2010 for the number of computers by 100 people, population size, GDP per capita, ICT development index and bribes indicator. When data for such year is missing, we use data from 2009 or 2008 when available. Otherwise, we use the average indicator value among countries with similar income level [44] (high income, upper middle income, lower middle income and low income). For the ICT development index, we collect the 11 indicators [44] used to compute such index, we use the above strategy to fill in missing values for each of these indicators, and then compute the index according to the methodology described by the ITU [19]. For the population size, using data from a previous year is sufficient to find data for all countries in our analysis.

5 Threats to Validity

The IPS WINE telemetry data set is collected from 10 million customer computers worldwide, and is very rich. However, such data also have limitations. First, the data only records attacks detected by Symantec IPS. Although Symantec uses heuristics to detect unknown attacks, detecting such attacks is never perfect. Unknown attacks include zero-day attacks and repackaged attacks. The implication of such limitation is that our findings do not apply to the most sophisticated cyber attacks. Moreover, the data are only about attacks on Symantec customers. There is no reason to believe that findings from customers of other anti-virus vendors would be different. However, findings from users that use no anti-virus protection would probably be different. Unfortunately, correcting for the above biases is difficult given that little is known about sophisticated attacks [6, 45] and about the difference between the quantity and type of attacks that anti-virus and non anti-virus users encounter internationally. Despite the limitations of the WINE telemetry data, to the best of our knowledge it is currently the best data available to academics to study the problem at hand. We are unaware of a platform other than WINE through which external researchers can access a vendor's telemetry data. Other vendors typically only provide aggregate attack counts [2, 27, 29] and limited documentation of their methodology and of how to scale such counts.

We also rely on Symantec's labeling of attacks. Researchers [4, 8, 30] have pointed out some inconsistencies into how different anti-virus vendors labels attacks. However, we believe that such inconsistencies reflect the lack of unified labeling guidelines across vendors rather than the fact that attack labeling from any

particular vendor is wrong. Moreover, researchers often use anti-virus labels as ground-truth for evaluating new approaches [5, 17, 36].

In our study, we rely on IP geolocation to identify the country where a computer is. IP geolocation at the country level is very accurate. One risk, however, is IP spoofing. A victim computer has no reason to spoof its IP address when sending attack reports to Symantec. Thus, IP spoofing has no affect on the attacks encountered per computer measure. On the other hand, an attacker computer may spoof its IP address. IP spoofing is relatively easy to carry out when the attacker computer does not need to set up a TCP connection with the victim computer in order to deliver the attack. This is the case when using UDP or in the context of some TCP Denial of Service (DoS) attacks. On the other hand, IP spoofing is more difficult and less common when a TCP connection is required. A remote attacker computer that spoofs its IP address will not see the TCP sequence numbers and will be unable to establish a connection with the victim computer. It is worth noting that TCP is much more commonly than UDP in the Internet [20], and that DoS attacks constitute less than 1% of attacks in our data. A related issue is that we only have data about the attacker computer that has a direct network communication with the victim computer. We have no data about other malicious computers that are part of the attack infrastructure, but that never directly communicate with victim computers.

Finally, many explanatory factors are difficult to measure precisely. Examples of such factors are corruption rate and computing resources. Such issue is typical in social science research. In order to alleviate this issue, we use indicators from respected organizations that use well documented methodologies.

6 Attacks Encountered

6.1 Descriptive Analysis

Figure 4 contains boxplots that depict the distribution of the number of attacks encountered by computer in different regions. We use geographical subdivision in Table 3. From the figure, we see that the number of exploits encountered per computer varies considerably across countries, and this even when considering countries in different regions separately. We also see that E. Eur ranks highest on the number of exploits encountered and that Africa ranks lowest on that measure. The difference across the 3 other regions is not significant. For web attacks, the variation within geographical regions is relatively small. W. Eur.+ is the most exposed region to web attacks, followed by Eastern Europe, Latin America and finally Africa. The pattern for fake apps is similar to the pattern for web attacks. In other words, the variation within geographical regions is small. Moreover, W. Eur.+ ranks highest followed by E. Eur., Lat. Am., Asia-Pc., and Africa. When considering all types of attacks, we see that E. Eur. and W. Eur.+ encounter the highest number of attacks followed by Asia-Pc. and Lat. Am.. Finally, Africa encounters the smallest number of attacks.

Table 4 presents the countries that score highest on the attacks encountered per country measure. From the table, we see that the highest ranking countries on the number of exploits encountered are countries with emergent economies, whereas the highest ranking countries on the number of web attacks encountered are developed countries. Finally, the list of the highest ranking countries when considering all attack types is similar to the list of highest ranking countries on the number of exploits encountered because exploits constitute the majority of attacks in our data (viz. Figure 2).

Figure 5 presents a visualization of the number of attacks encountered for all countries. From Figure 5a, we see a lack of a clear geographical pattern of the number of exploits encountered. The only exception are African countries that encounter a smaller quantity of attacks. From Figures 5b and 5c, we see that

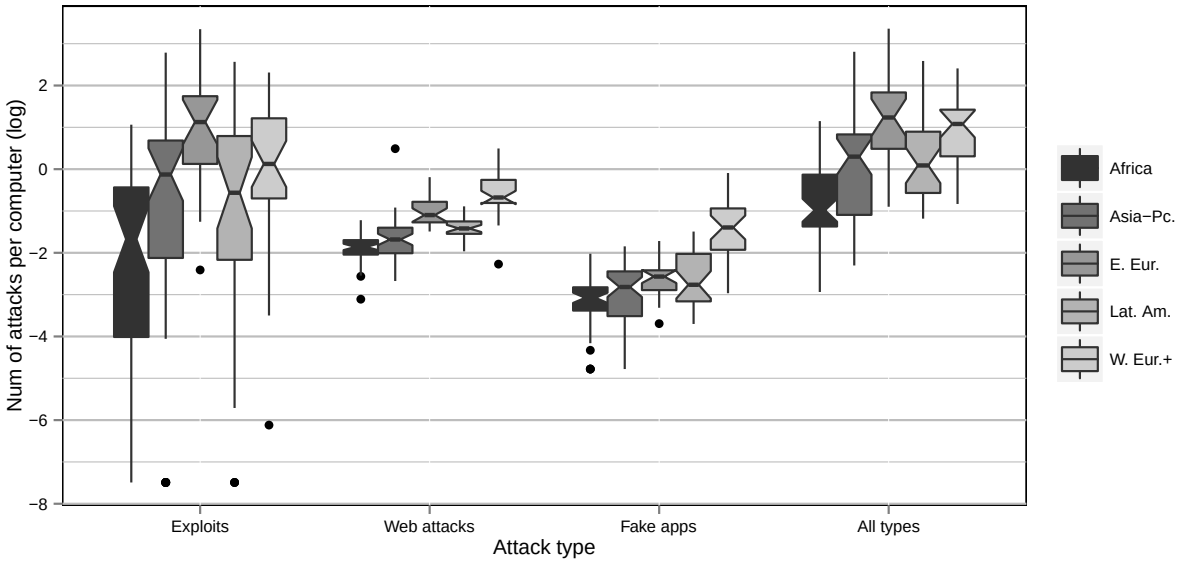


Figure 4: Attacks Encountered per computer. Regional distribution (log scale)

Table 4: Attacks encountered per computer. Top countries (log scale)

Exploits		Web attacks		Fake apps		All types	
Country	Value	Country	Value	Country	Value	Country	Value
Moldova	28.42	Germany	1.64	United States	0.92	Moldova	28.7
India	16.22	S. Korea	1.64	United Kingdom	0.83	India	16.56
Taiwan	15.75	United States	1.29	Canada	0.76	Taiwan	15.91
Nicaragua	13.02	United Kingdom	1.25	Australia	0.68	Nicaragua	13.3
Latvia	12.58	Netherlands	1.06	Ireland	0.59	Latvia	13.05
Italy	10.09	Canada	0.99	New Zealand	0.56	Italy	11.13
Israel	9.54	Australia	0.99	Norway	0.46	Israel	10.1
Uruguay	8.23	Russia	0.83	Switzerland	0.4	Uruguay	8.41
Bosnia & Herz.	6.86	Belgium	0.81	Belgium	0.38	Bosnia & Herz.	7.45
Georgia	6.54	Italy	0.79	Italy	0.79	Georgia	7.07

Table 5: Attacks encountered per computer. Summary statistics of variables used in the regression.

Abbreviation	Variable	Obs	Mean	S.D.	Min	Max
Exploits enc.	Exploits encountered (log)	184	-1.05	2.63	-7.49	3.35
Web atks enc.	Web attacks encountered (log)	184	-1.43	0.63	-3.11	0.49
Fake apps enc.	Fake apps encountered (log)	184	-2.68	0.90	-4.78	-0.09
All types enc.	All types encountered (log)	184	0.10	1.25	-2.94	3.36
Bandwidth	Bandwidth	184	30	57.66	0.10	547
ICT	ICT index	184	3.82	2.05	0.85	8.45
GDP PC	GDP per capita (log)	184	8.49	1.53	5.29	11.55
Web viz.	Web visits	184	2.91	8.72	0	92.47
Research	Cyber security research	184	175.5	830.41	0	7911
Institutions	Institutions & policy	184	0.36	0.48	0	1
Alliance btw	Alliance betweenness	184	0	0.01	0	0.05
Hostility btw	Hostility betweenness	184	0	0.002	0	0.02
Extradition btw	Extradition Betweenness	184	0	0.04	0	0.48

most countries W. Eur.+ and some countries in E. Eur. encounter excessive quantities of web attacks and fake apps. The pattern for the total attacks encountered in Figure 5d is very similar to the pattern of exploits encountered since exploits constitute the majority of attack instances in the IPS telemetry data (viz. Figure 2).

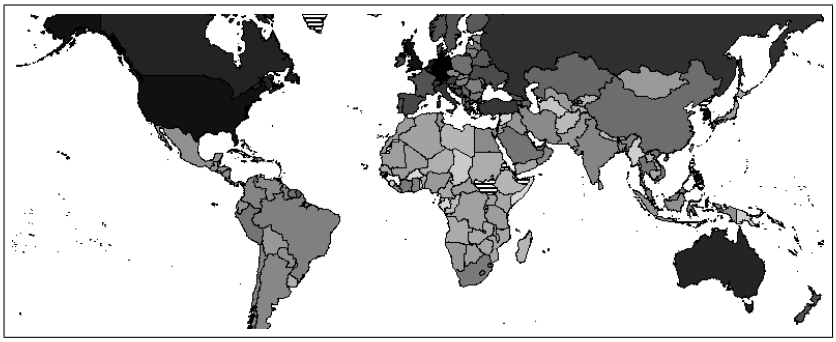
6.2 Explanatory Analysis

Table 5 presents summary statistics of variables used in the explanatory analysis of the number of attacks encountered per computer, and Table 6 presents the correlation between these variables. From Table 6, we see a high correlation between the number of web attacks and the number of fake application encountered, which may indicate a similarity in the factors that cause users to encounter these two attack types. On the other hand, we see a small correlation between web attacks and exploits, and between fake applications and exploits. We also see a very high correlation between ICT and GDP PC, which is expected given that rich countries typically have large computing resources. Moreover, ICT and GDP PC are moderately correlated with the number of exploits encountered, and highly correlated with the number of web attacks and fake applications encountered. In other words, countries with large computing and monetary resources encounter moderately more exploits, and much more web attacks and fake applications. The amount of cyber-security research and the existence of cyber-security institutions are positively correlated with the number of attacks encountered. This is surprising given that cyber security research and institutions are supposed to provide protection that reduces the number of attacks encountered. Cyber security research and institutions are also moderately correlated with ICT and GDP PC, which is expected given that developed countries tend to be more active in research and more concerned about cyber security issues. Finally, the correlation between international relations and the number of attacks encountered is small.

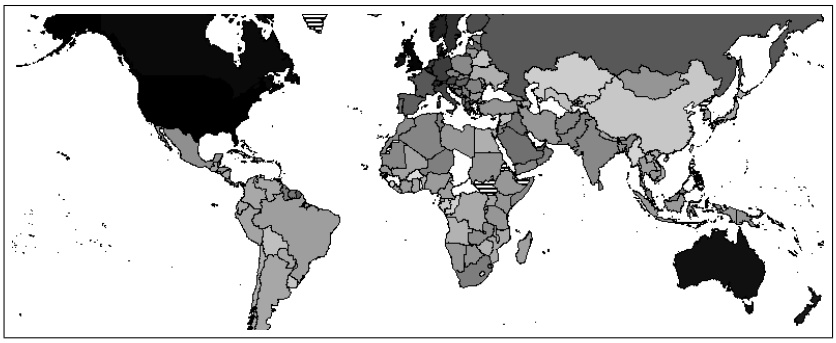
Table 7 presents the results of the regression analysis on the attacks encountered by computer measure. From the table, we see that computing and monetary resources have the most important impact. This is due to the fact that most cyber crime nowadays has a monetary goal. Large computing resources are attractive because these resources can be used for launching other attacks such as spam or Denial of Service attacks. Large monetary resources are attractive because hacking bank accounts of rich users generates large profits and because rich users are more likely to pay a premium for fake applications than poor users.

From the table, we also see that the coefficient on cyber security institutions and policy is positive. This is different from our prediction as cyber security institutions and policy are supposed to improve cyber security practices. One possible explanation for this finding is that countries that have cyber security institutions and

(a) Exploits



(b) Web attacks



(c) fake applications

(d) Total attacks

Figure 5: Attacks encountered per computer. Visualization (log scale).

Table 6: Attacks encountered per computer. Correlation Table of variables used in the regression

	Exploits enc.	Web atcks enc.	Fake apps enc.	All types enc.	Bandwith	ICT	GDP PC	Web viz.	Research	Institutions	Alliance btw	Hostility btw
Exploits enc.												
Web atks enc.	0.46***											
Fake apps enc.	0.28***	0.71***										
All types enc.	0.88***	0.62***	0.43***									
Bandwidth	0.16*	0.39***	0.44***	0.26***								
ICT	0.36***	0.70***	0.65***	0.51***	0.60***							
GDP PC	0.28***	0.61***	0.64***	0.42***	0.52***	0.92***						
Web viz.	0.02	0.15*	0.16*	0.04	0.02	0.16*	0.17*					
Research	0.11	0.33***	0.20**	0.16*	0.09	0.23**	0.20**	0.64***				
Institutions	0.44***	0.51***	0.35***	0.50***	0.34***	0.56***	0.47***	0.15*	0.27***			
Alliance btw	0.10	0.32***	0.31***	0.15*	0.10	0.21**	0.18*	0.48***	0.46***	0.20**		
Hostility btw	-0.02	0.10	0.11	-0.01	0.03	0.02	0.00	0.40***	0.38***	0.07	0.30***	
Extradition btw	0.04	0.22**	0.22**	0.08	0.03	0.13	0.12	0.78***	0.70***	0.12	0.58***	0.55***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

policy are also countries with large resources, and are thus more attractive attack targets. Some of these resources are not captured by the variables already in the model. Another possible explanation is reverse causality. Countries that encounter large quantities of attacks establish cyber security institutions and policy in order to address the problem. In future work, we intend to investigate this further using instrument variables.

International relations have a non-significant effect. State-sponsored attacks are still relatively rare [45] and thus do not make a significant difference in the regression. Moreover, many state-sponsored attacks are very sophisticated and are mostly undetected by the Symantec IPS.

7 Attacks Launched

7.1 Descriptive Analysis

Figure 6 depicts the distribution of the number of attacks launched by computer in different geographical regions. It is worth reminding that this measure is about attacks launched by computers, and not people in these countries. From the figure, we see that the variation in the number of exploits launched is smaller than the variation in the number of web attacks and fake applications launched. We also see that E. Eur. ranks highest on the number of exploits launched, and that Africa ranks lowest on that measure. For web attacks and fake applications, both E. Eur. and W. Eur.+ rank highest. Lat. Am. ranks lower in general, but some countries in that region are outliers with very large number of web attacks and fake applications launched. Finally, when taking into consideration all attack types, we see that E. Eur. ranks highest followed by both Lat. Am. and W. Eur.+. We find then Asia-Pc followed by Africa. It is surprising that Africa ranks very low on the number of attacks launched per computer given the scam reputation that some African countries e.g. Nigeria [39] have.

Table 8 presents the countries that score highest on the attacks launched per computer measure. From the table, we see that many Eastern European countries such as Moldova, Bosnia & Herz and Ukraine are among the countries that score highest in the 4 categories. Besides Eastern European countries Belize, Dominica and Trinidad & Tobago are among the highest scoring countries on the number of web attacks launched per computer, and that Dominica, Trinidad & Tobago, Luxembourg, Panama and Belize are among the highest scoring countries on the number of fake applications launched per computer.

Figure 7 presents the number of attacks launched per computer for all countries. From the figure, we see that most Eastern European score very high on this measure. On the other hand, although Belize and Dominica are among the top countries on the number of web attacks and fake applications launched (viz Table 8), most Latin American countries score low on these measures. Finally, African countries score very

Table 7: Attacks encountered per computer. Regression analysis. Regression coefficients are standarized.

	Exploits	Web atks	Fake apps	All types
Computing & Monetary resources				
bandwidth	-0.096	-0.051	0.080	-0.081
	(0.081)	(0.059)	(0.079)	(0.061)
ICT	0.22^{*}	0.61^{***}	0.59^{***}	0.38^{***}
	(0.094)	(0.087)	(0.093)	(0.090)
Web viz.	-0.11	-0.19	-0.12	-0.16
	(0.095)	(0.11)	(0.11)	(0.10)
Cyber security research and institutions				
Research	0.016	0.16	-0.056	0.029
	(0.026)	(0.086)	(0.056)	(0.040)
Institutions	0.35^{***}	0.13^{*}	-0.028	0.31^{***}
	(0.074)	(0.060)	(0.060)	(0.080)
International relations				
Alliance btw	0.019	0.15	0.15	0.037
	(0.020)	(0.082)	(0.086)	(0.032)
Hostility btw	-0.046	0.025	0.031	-0.055
	(0.066)	(0.032)	(0.067)	(0.046)
Extradition btw	0.059	0.050	0.17	0.11
	(0.066)	(0.10)	(0.10)	(0.062)
F-Statistics testing coefficients (p-value)				
Resources	0.06	<0.001	<0.001	<0.001
Research & institutions	<0.001	0.017	0.71	<0.001
International relations	0.60	0.13	0.52	0.16
N	184	184	184	184
R^2	0.22	0.56	0.47	0.34

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Attacks launched per computer. Top countries (log scale)

Exploits		Web attacks		Fake applications		All types	
Country	Value	Country	Value	Country	Value	Country	Value
Belarus	4.85	Belize	8.41	Dominica	13.82	Dominica	21.66
Moldova	2.88	Dominica	3.76	Trinidad & Tobago	5.27	Belize	9.08
Georgia	2.68	Moldova	2.97	Latvia	2.31	Trinidad & Tobago	7.29
Bulgaria	2.52	Ukraine	1.99	Bosnia & Herz.	1.01	Moldova	6.75
Bosnia & Herz.	1.96	Latvia	1.57	Moldova	0.81	Latvia	5.63
Ukraine	1.96	Trinidad & Tobago	1.33	Luxembourg	0.76	Belarus	4.91
Latvia	1.56	Lithuania	0.92	Panama	0.67	Ukraine	4.06
Congo	1.49	Bosnia & Herz.	0.75	Belize	0.55	Bosnia & Herz.	3.77
Hungary	1.42	Romania	0.74	Romania	0.50	Georgia	2.72
Romania	1.39	Russia	0.52	Ukraine	0.35	Romania	2.67

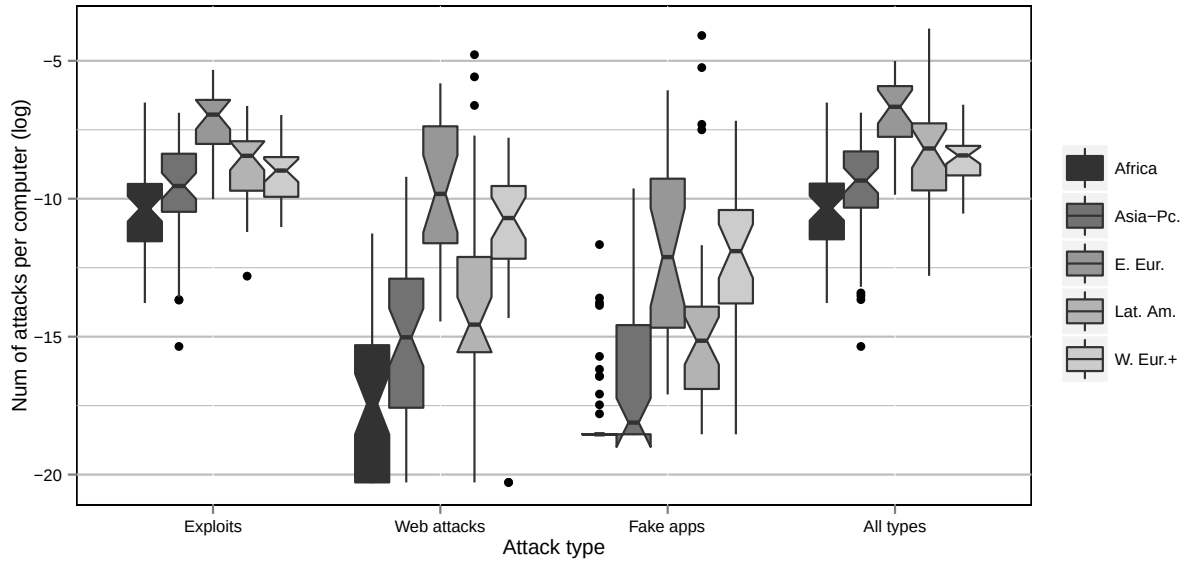


Figure 6: Attacks launched per computer. Regional distribution (log scale)

low on the number of web attacks and fake applications launched per computer.

7.2 Explanatory Analysis

Table 5 presents summary statistics of the variables used in the explanatory analysis of the number of attacks launched per computer, and Table 6 presents the correlation between these variables. From Table 10, we see a moderate correlation between the number of exploits and web attacks launched, and between the number of exploits and fake applications launched. We also see a high correlation between the number of web attacks and fake applications launched. This probably results from the fact that the attacking computers for both web attacks and fake applications are malicious web servers, but that the attacking computers for exploits can be different as explained in the attack catalog paragraph in Section 4.1. Moreover, the bribes index is highly correlated with the ICT index, indicating that countries with large computing resources tend to have low levels of corruption[3]. Furthermore, the number of attacks encountered is highly correlated with the number of exploits launched and moderately correlated with the number of web attacks and fake applications launched. Finally, international relations have a very small or no correlation with the number of attacks launched per computer.

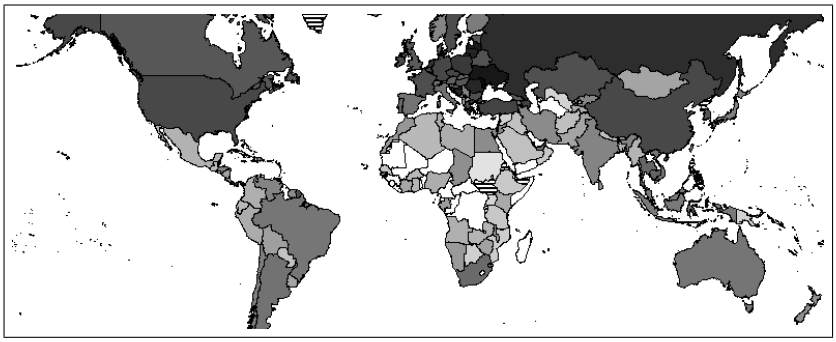
Table 11 presents the results of the regression analysis on the attacks launched per computer measure. From the table, we see that countries that have a combination of good computing resources and high levels of corruption are countries that launch most attacks. This finding is consistent across all attack types.

The coefficients of cyber security research, and institutions and policy are positive. It is, however, difficult to draw conclusions based on that observation. As discussed in Section 6.2, cyber security research, and institutions and policy may be correlated with resources not captured in the model, or may be boosted as a response to past cyber attacks.

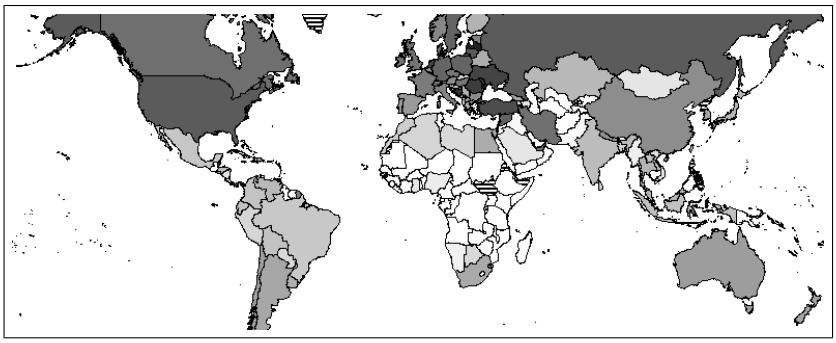
From the table, we see that encountering more attacks results in launching more attacks. As computers encounter more attacks, these computers are more likely to become compromised and start launching

[3] A small bribes index indicates high levels of corruption

(a) Exploits



(b) Web attacks



(c) Fake applications

(d) Total attacks

Figure 7: Attacks launched per computer. Visualization (log scale).

Table 9: Attacks launched per computer. Summary statistics of variables used in the regression.

Abbreviation	Variable	Obs	Mean	S.D.	Min	Max
Exploits lau.	Exploits launched (log)	184	-9.39	1.86	-15.35	-5.33
Web atks lau.	Web attacks launched (log)	184	-14.35	4.13	-20.28	-4.78
Fake apps lau.	Fake apps launched (log)	184	-15.25	3.61	-18.54	-4.08
All types lau.	Total attacks launched (log)	184	-9.12	2	-15.35	-3.83
Bandwidth	Bandwidth	184	30	57.66	0.10	547
ICT	ICT index	184	3.82	2.05	0.85	8.45
Bribes	Bribes	184	4.10	1.11	2.50	6.70
Research	Cyber security research	184	175.5	830.41	0	7911
All types enc.	All types encountered (log)	184	0.10	1.25	-2.94	3.36
Institutions	Institutions & policy	184	0.36	0.48	0	1
Alliance	Betweenness alliance	184	0	0.01	0	0.05
Hostility	Betweenness hostility	184	0	0.002	0	0.02
Extradition	Betweenness extradition	184	0	0.04	0	0.48

Table 10: Attacks launched per computer. Correlation Table of variables used in the regression

	Exploits lau.	Web atks. lau.	Fake apps lau.	All types lau.	Bandwidth	ICT	Bribes	Research	Institutions	All types enc.	Alliance	Hostility
Exploits lau.												
Web atks lau.	0.57***											
Fake apps lau.	0.49***	0.79***										
All types lau.	0.94***	0.71***	0.67***									
Bandwidth	0.16*	0.37***	0.41***	0.21**								
ICT	0.42***	0.61***	0.61***	0.48***	0.60***							
Bribes	0.20**	0.35***	0.39***	0.23**	0.57***	0.79***						
Research	0.04	0.22**	0.23**	0.09	0.09	0.23**	0.17*					
Institutions	0.38***	0.54***	0.48***	0.38***	0.34***	0.56***	0.39***	0.27***				
All types enc.	0.76***	0.60***	0.53***	0.72***	0.26***	0.51***	0.32***	0.16*	0.50***			
Alliance	0.06	0.18*	0.21**	0.09	0.10	0.21**	0.11	0.46***	0.20**	0.15*		
Hostility	-0.07	0.03	0.02	-0.05	0.03	0.02	-0.02	0.38***	0.07	-0.01	0.30***	
Extradition	-0.01	0.11	0.14	0.04	0.03	0.13	0.08	0.70***	0.12	0.08	0.58***	0.55***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Attacks launched per computer. Regression analysis. Regression coefficients are standarized.

	Exploits	Web atks	Fake apps	All types
Computing resources				
Bandwidth	-0.13	0.73	1.1	0.0083
	(0.36)	(0.41)	(0.55)	(0.36)
ICT	1.2***	1.07***	0.63*	1.3***
	(0.24)	(0.27)	(0.31)	(0.28)
Corruption				
Bribes	0.33*	0.098	0.014	0.25
	(0.13)	(0.15)	(0.14)	(0.13)
Computing resources x Corruption				
Bandwidth x Bribes	0.17	-0.57	-0.89	0.063
	(0.36)	(0.42)	(0.54)	(0.36)
ICT x Bribes	-1.48***	-0.93*	-0.37	-1.35***
	(0.30)	(0.37)	(0.39)	(0.35)
Cyber security research and institutions				
Research	-0.067	0.093**	0.081**	-0.028
	(0.036)	(0.029)	(0.028)	(0.041)
Institutions	-0.014	0.16*	0.092	-0.032
	(0.051)	(0.065)	(0.076)	(0.053)
All types enc.	0.69***	0.28***	0.23***	0.61***
	(0.060)	(0.054)	(0.063)	(0.058)
International relations				
Alliance	-0.031	0.0054	0.042	-0.037*
	(0.016)	(0.023)	(0.032)	(0.015)
Hostility	-0.029	-0.014	-0.059	-0.045
	(0.035)	(0.074)	(0.035)	(0.034)
Extradition	0.0048	-0.046	0.0073	0.032
	(0.033)	(0.043)	(0.030)	(0.033)
F-Statistics testing coefficients (p-value)				
Computing resources	<0.001	<0.001	<0.001	<0.001
Computing resources x Corruption	<0.001	0.0016	0.034	<0.001
Research & institutions	0.151	<0.001	0.001	0.612
International relations	0.21	0.107	0.19	0.069
N	184	184	184	184
R^2	0.65	0.57	0.49	0.61

Standard errors in parentheses
$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

attacks. Finally, international relations have mostly a non-significant effect on attacks launched.

8 Cyber Attack Networs

8.1 Descriptive Analysis

Visualizing the entire cyber attack networks would very complex. Instead, we include a regional aggregation of the networks in Tables 12 and a visualization of the strong edges in the networks in Figures 8.

In tables 12, the edge weight from region A to region B represents the average number of attacks from a computer in A on a WINE computer in B. From Table 12a, we see that exploits have a high tendency to propagate within geographical regions. When only considering inter-region propagation, we see that all regions encounter more exploits from E. Eur. than from any other region. Moreover, all regions launch more

exploits on E. Eur, W. Eur.+ and Lat. Am. than on Africa and Lat. Am. From Tables 12b and 12c, we see more intra-region than inter-region propagation of web attacks and fake applications. This is, however, less pronounced than for exploits. We also see that all regions encounter more web attacks from E. Eur. and W. Eur.+ than they encounter from other regions. We also see that all regions with the exception of Africa launch more attacks on W. Eur.+ than they launch on other regions. Finally, when taking into accounts all attack types in Table 12d, we see a pattern similar to the pattern of exploits. This is due to the fact that exploits constitute the majority of attacks in the data as depicted in Figure 2.

Figure 8 presents a visualization of the exploit and web attack networks. We do not include the fake application and the all attack type networks because the fake application network is similar to the web attack network and the characteristics of the all attack type network are a combination of the characteristics of the exploit, web attack and fake application networks. In order to produce meaningful visualizations, we only keep the strongest edges in each network. We ensure that the visualized networks have the same density 0.003. In Figure 8a, we see a large Eastern European cluster, a small African cluster and a very small Latin American cluster. This indicates that exploits tend to propagate to geographically nearby areas as already seen in Table 12a. From Figure 8b, we see that most links are from a country in Latin America or Eastern Europe to rich countries primarly in Western Europe.

8.2 Explanatory Analysis

Table 13 presents network level measures of the networks used in the explanatory analysis of the cyber attack networks. We only use attributes of victim countries and attacker countries that were found significant in Sections 6.2 and 7.2 respectively[4]. A component is a maximal set of nodes that are connected. In the table, we distinguish between components that contain at least 2 nodes and isolates (a single node disconnected from the rest of the network). Density is the ratio of edges present in the network to all possible edges in the network. The clustering coefficient measures the extent to which a node's neighbors are themselves neighbors. From Table 13, we see that the 4 cyber attack networks have different density, most probably due to the difference in the number of attack instances belonging to different types as depicted in Figure 2. The attribute networks (ICT att, bandwidth att, bribes att, ICT vic and ICT % diff) have all 1 component, and density and clustering coefficient equal to 1 because of the way these networks are computed. The regional network consists of 5 components corresponding to the 5 geographical regions (Africa, Asia-Pc, E. Eur, Lat. Am. and W. Eur+). Finally, the network level measures of the international networks reflect the characteristics of international relations. For example, the clustering coefficient of the hostility network is very small because countries that have a common enemy tend to be friends, not enemies.

Table 14 presents the correlation between networks we use in the MrQAP regression and Table 15 presents the results of the MrQAP regression on the cyber attack networks. From Table 15, we see that attributes of attackers and victims have a similar effect to the effect discussed in Sections 6.2 and 7.2. That is resources have a positive impact on the number of attacks encountered and launched. Similarly, a combination of good computing resources and high corruption increases the number of attacks launched. From the table, we also see that attacker and victim computers have a slightly higher chance of being in

[4]We compute the attribute network for the attacker by repeating that attribute column. For example, assume we have 3 countries with ICT indices $(3, 8, 6)$, the ICT attacker network is $\begin{bmatrix} 3 & 3 & 3 \\ 8 & 8 & 8 \\ 6 & 6 & 6 \end{bmatrix}$. A value in the matrix is the ICT index of the country on the row i.e. the attacker country. Similarly, we compute the attribute network for the victim by repeating the attribute row. The ICT victim network is then $\begin{bmatrix} 3 & 8 & 6 \\ 3 & 8 & 6 \\ 3 & 8 & 6 \end{bmatrix}$.

Table 12: Regional aggregation of cyber attack networks (log scale)

(a) Exploits

	Africa	Asia-Pc.	E. Eur.	Lat. Am.	W. Eur.+
Africa	-21.96	-26.32	-25.96	-26.89	-26.35
Asia-Pc.	-26.85	-22.94	-25.29	-26.25	-25.73
E. Eur.	-24.74	-23.57	-21.32	-24.17	-23.27
Lat. Am.	-26	-25.18	-25.05	-22.58	-24.88
W. Eur.+	-27.11	-25.99	-25.03	-26.58	-24.75

-28	-26.8	-25.7	-24.5	-23.3	-22.2

(b) Web attacks

	Africa	Asia-Pc.	E. Eur.	Lat. Am.	W. Eur.+
Africa	-27.21	-31.21	-31.4	-31.71	-30.39
Asia-Pc.	-29.43	-26.11	-28.65	-29.82	-26.36
E. Eur.	-25.46	-25.69	-23.77	-25.29	-23.49
Lat. Am.	-29.18	-29.73	-27.77	-26.26	-27.12
W. Eur.+	-26.69	-26.78	-26.13	-26.6	-25.21

-32	-30.5	-29	-27.5	-26	-24.5

(c) Fake apps

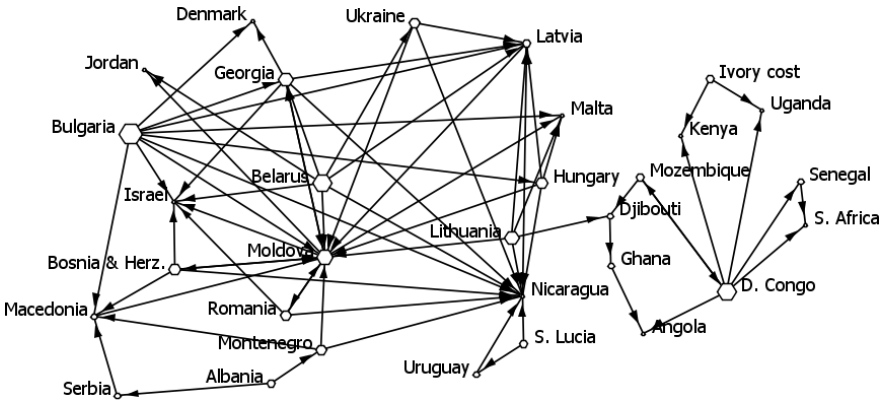
	Africa	Asia-Pc.	E. Eur.	Lat. Am.	W. Eur.+
Africa	-28.6	-33.03	-32.6	-31.86	-31.44
Asia-Pc.	-31.35	-29.47	-31.32	-31.37	-28.3
E. Eur.	-27.28	-27.41	-26.77	-27.11	-24.53
Lat. Am.	-30.45	-30.37	-29.87	-29.17	-26.61
W. Eur.+	-27.87	-27.93	-27.04	-27.89	-25.36

-34	-32.3	-30.7	-29	-27.3	-25.7

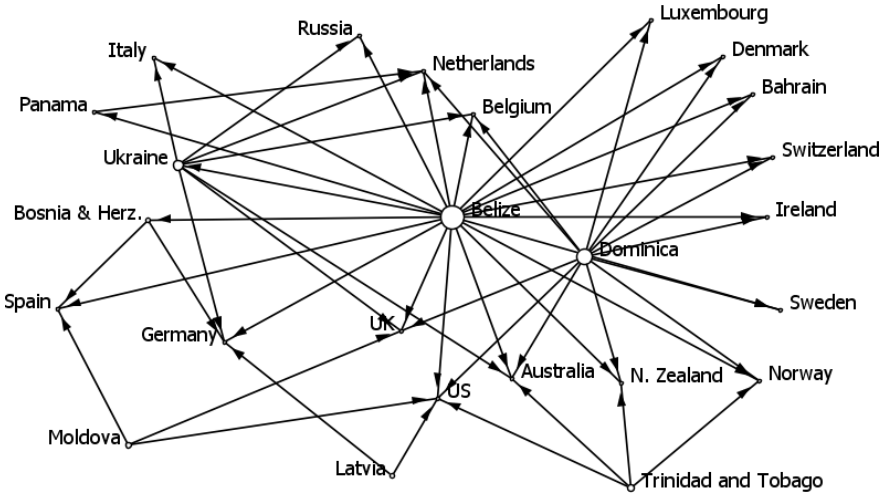
(d) All types

	Africa	Asia-Pc.	E. Eur.	Lat. Am.	W. Eur.+
Africa	-21.95	-26.3	-25.95	-26.87	-26.27
Asia-Pc.	-26.75	-22.88	-25.25	-26.21	-25.24
E. Eur.	-24.28	-23.44	-21.23	-23.84	-22.52
Lat. Am.	-25.94	-25.16	-24.97	-22.55	-24.62
W. Eur.+	-25.93	-25.45	-24.62	-25.72	-23.93

-27	-26	-25	-24	-23	-22



(a) Exploits



(b) Web attacks

Figure 8: Cyber attack networks. The figure only contains the strongest edges. Nodes are scaled by their outdegree

Table 13: Cyber attack networks. Network Level Measures of Networks used in the MrQAP regression

Abbreviation	Network	Size	Components (2+ nodes)	Isolates (1 node)	Density	Cluster. coef.
Exploits	Exploits (log)	184	1	0	0.39	0.77
Web atks	Web attacks (log)	184	1	0	0.21	0.75
Fake apps	Fake apps (log)	184	1	16	0.13	0.71
All types	All types (log)	184	1	0	0.43	0.79
ICT att	ICT attacker	184	1	0	1	1
Bandwidth att	Bandwidth attacker	184	1	0	1	1
Bribes att	Bribes attacker	184	1	0	1	1
ICT vic	ICT victim	184	1	0	1	1
ICT % diff	ICT % difference	184	1	0	1	1
Regional	Regional membership	184	5	0	0.21	1
Hostility	Hostility	184	9	133	0.003	0.011
Extradition	Extradition	184	1	40	0.045	0.67
Alliance	Alliance	184	6	61	0.074	0.54

Table 14: Cyber attack networks. Correlation Table of networks used in the MrQAP regression

	Exploits	Web atks	Fake apps	All types	ICT att	Bandwidth att	Bribes att	ICT vic	ICT diff	Regional	Hostiles	Extradition
Exploits												
Web atks	0.43***											
Fake apps	0.34***	0.62***										
All types	0.92***	0.61***	0.48***									
ICT att	0.24***	0.41***	0.36***	0.32***								
Bandwidth att	0.11***	0.27***	0.28***	0.18***	0.60***							
Bribes att	0.11***	0.23***	0.21***	0.16***	0.79***	0.57***						
ICT vic	0.19***	0.34***	0.29***	0.26***	0.92***	0.52***	0.78***					
ICT diff	0.03	0.18***	0.15***	0.08***	0.61***	0.29***	0.44***	0.61***				
Regional	0.07***	0.01	0.00	0.05***	-0.09	-0.05	-0.05	-0.08	0.10***			
Hostiles	0.05***	0.04***	0.03*	0.05***	0.00	0.01	0.00	0.00	0.00	0.04***		
Extradition	0.16***	0.23***	0.21***	0.17***	0.11***	0.10*	0.10***	0.11***	0.04***	0.25***	0.02	
Alliance	0.15***	0.14***	0.12***	0.15***	0.07***	0.02	0.03	0.07***	0.06***	0.32***	0.02*	0.28***

Table 15: MrQAP regression on cyber attack networks. Coefficients are standarized.

	Exploits	Web atks	Fake apps	All types
Attributes of attackers and victims				
ICT att	1.16***	1.00***	1.05***	0.60***
Bandwidth att	0.17***	0.13	0.10	0.092
Bribes att	0.057	0.079	0.23	0.11
ICT x Bribes att	-0.97***	-0.93***	-0.95	-0.47
ICT vic	-0.21***	-0.21***	-0.22	-0.20
Interaction between attributes of attackers and victims				
ICT att x ICT vic	0.19***	0.23***	0.51***	0.47***
Bribes att x ICT vic	0.21*	0.21*	-0.26*	-0.27
Bandwidth att x ICT vic	-0.11***	-0.11***	0.032***	0.071***
ICT % diff	-0.003	0.004	-0.105	-0.068
Geographical proximity				
Regional	0.055***	0.07***	-0.011	-0.041***
International relations				
Hostility	0.037***	0.038***	0.035***	0.021***
Extradition	0.062***	0.058*	0.13***	0.13***
Alliance	0.063***	0.054***	0.046*	0.051***
N	184	184	184	184
R^2	0.26	0.22	0.29	0.23

$^{*}\, p < 0.05, ^{**}\, p < 0.01, ^{***}\, p < 0.001$

the same geographical region, which is consistent with Tables 12. Finally, international relations have a very small effect that is not necessarily in the expected direction. Hostility has a positive effect as expected. However, extradition and alliance networks also have a positive effect contrary to what is expected.

9 Future Work

Symantec also has an Anti-Virus (AV) telemetry data that we intend to analyze in future work. The AV telemetry data contain attack reports sent by Symantec's AV. The AV and IPS are two end-host systems that run side-by-side, but operate differently and do not interact. More precisely, the IPS examines network activity, whereas the AV detects malicious files. The main types in the AV telemetry are trojans, viruses and worms. We do not analyze the AV telemetry in this paper because the IP address of the attacker computer is missing in the AV telemetry data. This means that we are unable to compute the attacks launched by computer measure and the cyber attack networks from the AV telemetry data. As future work, we intend to extract and analyze the number of attacks encountered by computer measure from the AV telemetry data.

In this paper, we have included explanatory factors that we believe to be the most pertinent. In future work, it would interesting to also include other factors such as the usage of different operating systems (e.g. Windows, Mac OS and Linux) in different countries. Finally, it would be interesting to conduct case studies about the cyber security situation in countries identified as originating the most attacks in order to gain an in-depth understanding of cyber security issues there.

10 Conclusion

In this paper, we perform an empirical study of the quantity and type of cyber attacks that different countries encounter and launch. Our study is based on Symantec's IPS telemetry data set. We identify countries where computers encounter attacks most, and countries where computers launch attacks most. We also build an explanatory model of the quantity of cyber attacks that countries encounter and launch. Examples of factors included in the model are computing and monetary resources, cyber security research, corruption and international relations. Finally, we analyze the network of how attacks launched by computers in one country affect computers in other countries.

We find that countries with large computing and monetary resources encounter excessive quantities of attacks. This is because attacks on such countries have large potential monetary gains. We also find that many countries in Eastern Europe and a few countries in Central America are particularly attractive for hosting attacking computers. Such countries have a combination of good computing infrastructure and high levels of corruption. The high levels of corruption facilitate conducting criminal activities such as registering malicious websites and keeping malicious computers up despite complaints. We speculate that cyber criminals behind attacks from Eastern Europe are local hackers with high cyber security expertise and limited venues for lucrative legitimate work. On the other hand, cyber criminals behind attacks in Central America are probably from other regions, but choose this region for conducting cyber crime because of attractive conditions there.

Unfortunately, if these countries let such practices continue and grow, IP addresses from these countries may become blocked in bulk. This may cause end-users from these countries to become virtually blocked from parts of the Internet. The international community should work on addressing lax practices in these countries for the benefit of victim users worldwide, but also for the benefit of users in these countries. This is not necessarily easy as corruption is a social problem as old as humanity itself. The problem is best

addressed using soft-power solutions that raise awareness about the issue and highlight the danger of such lax practices to end-uses in these countries and worldwide.

References

[1] L. A. Adamic and B. A. Huberman. Zipf's law and the internet. *Glottometrics*, 3:143–150, 2002.

[2] Akamai. Akamai's state of the internet report, Q1 2014.

[3] Alexa. The top 500 sites in each country or territory. `www.alexa.com/topsites/countries`, 2013. Last accessed: October 2013.

[4] M. Bailey, J. Oberheide, J. Anderen, Z. M. Mao, F. Jahanian, and J. Nezario. Automated classification and analysis of internet malware. In *International Symposium on Research in Attacks, Instrusions and Defenses (RAID)*, September 2007.

[5] U. Bayer, P. M. Comparetti, C. Hlauschek, C. Kruegel, and E. Kirda. Scalable, behavior-based malware clustering. In *Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, February 2009.

[6] L. Bilge and T. Dumitraş. Before we knew it. An empirical study of zero-day attacks in the real world. In *Computer and Communication Security Conference (CCS)*, Raleigh, NC, October 2012.

[7] J. Caballero, C. Grier, C. Kreibich, and V. Paxson. Measuring pay-per-install: The commoditization of malware distribution. In *The 20th USENIX Security Symposium*, San Francisco, CA, August 2011.

[8] J. Canto, M. Dacier, E. Kirda, and C. Leita. Large scale malware collection: Lessons learned. In *IEEE SRDS Workshop on Sharing Field Data and Experiment Measurements on Resilience of Distributed Computed Systems*, October 2008.

[9] Center for International Development and Conflict Management. International crisis behavior project. `http://www.cidcm.umd.edu/icb/`. Last accessed: December 2011.

[10] Central Intelligence Agency. The World Factbook. `https://www.cia.gov/library/publications/the-world-factbook/`. Last accessed: March 2012.

[11] CERT. National computer security incident response teams. `http://www.cert.org/csirts/national/contact.html`, 2014. Last accessed: January 2014.

[12] N. Choucri, S. Madnick, and J. Ferwerda. Institutions for cyber security: International responses and global imperatives. *Information Technology for Development*, 19(5), 2013.

[13] Correlates of War Project. Alliances v3.03. `http://www.correlatesofwar.org/`. Last accessed: December 2011.

[14] Department of Peace and Conflict Research. Uppsala University. Ucdp dyadic dataset. `http://www.pcr.uu.se/research/ucdp/datasets/ucdp_dyadic_dataset/`. Last accessed: December 2011.

[15] W. E. Forum. The global competitiveness report. `http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2012-13.pdf`, 2012-2013.

[16] C. Grier, L. Ballard, J. Caballero, N. Chachra, C. J. Dietrich, K. Levchenko, P. Mavrommatis, D. McCoy, A. Nappa, A. Pitsillidis, N. Provos, M. Z. Rafique, M. A. Rajab, C. Rossow, K. Thomas, V. Paxson, S. Savage, and G. M. Voelker. Manufacturing compromise: The emergence of exploit-as-a-service. In *Proceedings of the Conference on Computer and communications Security (CCS)*, Raleigh, NC, October 2012.

[17] X. Hu, T. Chiueh, and K. G. Shin. Large-scale malware indexing using function-call graphs. In *Computer and Communication Security Conference (CCS)*, Chicago, IL, November 2009.

[18] International Cyber Center. George Mason University. Certicc home. `http://internationalcybercenter.org/certicc`, 2014. Last accessed: January 2014.

[19] International Telecommunication Union. Measuring the information society. `http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2012/MIS2012_without_Annex_4.pdf`, 2012.

[20] W. John and S. Tafvelin. Analysis of internet backbone traffic and header anomalies observed. In *ACM Internet Measurement Conference (IMC)*, San Diego, CA, August 2007.

[21] B. Johnson, J. Chuang, J. Grossklags, and N. Christin. Metrics for measuring ISP badness: The case of spam (short paper). In *Proceedings of the 16th International Conference on Financial Cryptography and Data Security*, Bonaire, February 2012.

[22] A. J. Kalafut, C. A. Shue, and M. Gupta. Malicious hubs: Detecting abnormally malicious autonomous systems. In *Proceedings of the 29th Conference on Information Communications (INFOCOM)*, San Diego, CA, March 2010.

[23] D. Krackhardt. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10(5):359–381, December 1988.

[24] J. A. Lewis and K. Timlin. Cybersecurity and cyberwarfare. preliminary assessment of national doctrine and organization. Technical report, Center for Strategic and International Studies, 2011.

[25] S. Madnick, X. Li, and N. Choucri. Experiences and challenges with using cert data to analyze international cyber security. In *Proceedings of the AIS SIGSEC Workshop on Information Security and Privacy (WISP)*, Phoenix, AZ, December 2009.

[26] Maxmind. Geolite free downloadable databases. geolite country. `http://dev.maxmind.com/geoip/legacy/geolite/`, November 2012.

[27] McAfree. Mcafee labs threats report. `http://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q1-2014.pdf`, June 2014.

[28] G. Mezzour, L. R. Carley, and K. M. Carley. Longitudinal analysis of a large corpus of cyber threat descriptions. *Journal of Computer Virology and Hacking Techniques*, in press.

[29] Micrsoft. Micrsoft security intelligence report. Worldwide threat assessment, July-December 2013.

[30] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver. Inside the slammer worm. *IEEE Security and Privacy*, 4(1):33–39, July 2003.

[31] D. Moore, G. M. Vœlker, and S. Savage. Inferring internet denial-of-service activity. In *10^{th} USENIX Security Symposium*, pages 115–139, Washington, DC, August 2001.

[32] North Atlantic Treaty Organization. NATO and cyber defense. `http://www.nato.int/cps/en/natolive/topics_78170.htm`.

[33] D. of General Assembly and C. Management. United Nations Regional Groups of Member States. `http://www.un.org/depts/DGACM/RegionalGroups.shtml`, 2014.

[34] M. Portnoy and S. Goodman, editors. *Global Initiatives to Secure Cyberspace. An Emerging Landscape*. Oxford University Press, Oxford, 2004.

[35] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *Proceedings of the ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM)*, pages 291–302, Pisa, Italy, September 2006.

[36] K. Rieck, T. Holz, C. Willems, P. Düssel, and P. Laskov. Learning and classification of malware behavior. In *Conference on Detection of Intrusions and Malware and Vulnerability (DIMVA)*, pages 108–125, Paris, France, July 2008.

[37] D. E. Sanger. Obama order sped up wave of cyberattacks against Iran. `http://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html?ref=stuxnet`, June 2012.

[38] SCOPUS. `www.scopus.com`. Last accessed: October 2012.

[39] F. Stajano and P. Wilson. Understanding scam victims: seven principles for systems security. *Communications of the ACM*, 54(3):70, Mar. 2011.

[40] B. Stone-Gross, C. Kruegel, K. Almeroth, K. Almeroth, A. Moser, and E. Kirda. Fire: Finding rogue networks. In *Annual Computer Security Applications Conference*, Honolulu, HI, December 2009.

[41] Symantec attack signatures. `http://www.symantec.com/security_response/attacksignatures/`. Last accessed: October 2012.

[42] The New York Times. Indignation over u.s. spying spreads in europe. `http://www.nytimes.com/2013/10/25/world/europe/indignation-over-us-spying-spreads-in-europe.html?_r=0`.

[43] The World Bank. The little data book on information and communication technology. `http://data.worldbank.org/products/data-books/little-data-book-on-info-communication-tech`, 2011.

[44] The World Bank. World development indicators (wdi) 2012. `http://data.worldbank.org/data-catalog/world-development-indicators/wdi-2012`, April 2012.

[45] O. Thonnard, L. Bilge, G. O'Gorman, S. Kiernan, and M. Lee. Industrial espionage and targeted attacks: Understanding the characteristics of an escalating threat. In *International Symposium on Research in Attacks, Instrusions and Defenses (RAID)*, September 2012.

[46] United Nations Crime and Justice Information Network. Bilateral agreements on extradition, judicial/legal assistance, control of narcotic drugs, and prisoner transfer by country. `http://www.uncjin.org/Laws/extradit/extindx.htm`.

[47] M. Vatis. International cyber-security cooperation. Informal bilateral models. In J. A. Lewis, editor, *Cyber Security: Turning National Solutions into International Cooperation*. CSIS Press, Center for Strategic and International Studies, Washington, DC, 2003.