Assessing the Global Cyber and Biological Threat

Ghita Mezzour

CMU-ISR-15-102

April 2015

Electrical and Computer Engineering Department
and Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Dr. L. Richard Carley, co-chair
Dr. Kathleen M. Carley, co-chair
Dr. Nicolas Christin
Dr. Mathew Elder, Symantec Research Labs

*A dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Philosophy.*

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.

Keywords: Cyber security, Science of security, Computers and Society, Public Policy Issues, Abuse and Crime Involving Computers, Global Cyber Security, Cyber Weapons, Cyber Warfare, Cyber Capabilities, Biological Weapons, Bioweapons, Bio Capabilities, Socio-cultural model, Friedkin model, Empirical analysis

Abstract

In today's inter-connected world, threats from anywhere in the world can have serious global repercussions. In particular, two types of threats have a global impact: 1) cyber crime and 2) cyber and biological weapons. If a country's environment is conducive to cyber criminal activities, cyber criminals will use that country as a basis to attack end-users around the world. Cyber weapons and biological weapons can now allow a small actor to inflict major damage on a major military power. If cyber and biological weapons are used in combination, the damage can be amplified significantly.

Given that the cyber and biological threat is global, it is important to identify countries that pose the greatest threat and design action plans to reduce the threat from these countries. However, prior work on cyber crime lacks empirical substantiation for reasons why some countries' environments are conducive to cyber crime. Prior work on cyber and biological weapon capabilities mainly consists of case studies which only focus on select countries and thus are not generalizeable. To sum up, assessing the global cyber and biological threat currently lacks a systematic empirical approach.

In this thesis, I take an empirical and systematic approach towards assessing the global cyber and biological threat. The first part of the thesis focuses on cyber crime. I examine international variation in cyber crime infrastructure hosting and cyber crime exposure. I also empirically test hypotheses about factors behind such variation. In that work, I use Symantec's telemetry data, collected from 10 million Symantec customer computers worldwide and accessed through the Symantec's Worldwide Intelligence Network Environment (WINE). I find that addressing corruption in Eastern Europe or computer piracy in Sub-Saharan Africa has the potential to reduce the global cyber crime.

The second part of the thesis focuses on cyber and biological weapon capabilities. I develop two computational methodologies: one to assess countries' biological capabilities and one to assess countries' cyber capabilities. The methodologies examine all countries in the world and can be used by non-experts that only have access to publicly available data. I validate the biological weapon assessment methodology by comparing the methodology's assessment to historical data. This work has the potential to proactively reduce the global cyber and biological weapon threat.

Acknowledgments

First, I would like to thank my advisors Kathleen Carley and Rick Carley for giving me the opportunity to work with them and conduct research with real world impact. My advisors have always been very supportive and encouraging. At times, they believed in me more than I believed in myself. I have developed considerable respect and esteem for them over the years.

I would also would like to thank my committee members Nicolas Christin and Matthew Elder for their constructive and invaluable feedback and discussions. Thank you also to Tudor Dumitras for helping me access and use the Symantec WINE data. Long discussions with Tudor also helped have a better understanding of the potential and limitations of the WINE data.

Many thanks to CASOS members throughout the years for providing a welcoming and supportive environment. I have always felt at home at CASOS and have developed great friendships. Thank you Neal Altman, Hope Armstrong, Matthew Benigni, Phil Curley, Jana Diesner, Will Frankenstein, Laurie Jones, Kenny Joseph, Frank Kunkel, Hemank Lamba, Peter Landwehr, Mike Lanham, Momin Malik, Mike Martin, Geoff Morgan, Juergen Pfeffer, Jeff Reminga, Jon Storrick, Sarah Tedrow-Azizi, and Wei Wei. Will, in particular, has been a great officemate and collaborator.

I am also deeply grateful to my family for their love and support. My relationship with my family is as strong as ever despite years of living apart. My parents have always put their children's happiness and success before theirs. My siblings Kamal, Farah, and Mariam are the best friends I have ever had. Nabila, Yahya, Ali, and Noor have joined the family after I left my home country. Although I have never had the opportunity to live in the same country as you, I love you profoundly. I am always amazed by how fun and intelligent the little ones are. My cousins were my favorite playmates as a child and remain very close. Last but not least, my aunts and uncles have always been fun and supportive.

During my years living abroad, I have known great friends who shared my joys and helped me throughout my struggles. Marta Habermann and Lucille Thornton have opened their houses to me and provided maternal affection. Bara Ammoura, Yasmin Bakrim, Sara Belahsen, Hanan Hibshi, Maryem Mabrouk, Ashwini Rao, and Divya Sharma have made Pittsburgh an enjoyable place. Selma Aboussaoud, Aziza Benaboud, and Loubna El Arabi were my surrogate family when living in Lausanne.

Finally, thank you to Samantha Goldstein and Connie Herold for their help working out the administrative procedure of the joint Ph.D. degree. Thank you also to everybody who has contributed to my long educational journey.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	**Interpretation**
atks	attacks
att	attacker
AV	anti-virus
Asia-PC	Asia-Pacific
btw	betweenness
BW	biological weapons
E. Eur	Eastern Europe
enc.	encountered
GDP PC	GDP per capita
ICT	information and communication technology
IPS	intrusion prevention system
Lat. Am.	Latin America
lau.	launched
MrQAP	Multiple regression quadratic assignment procedure
TP	true positive
TN	true negative
vic	victim
W. Eur+	Western Europe and others
WINE	World Intelligence Network Environment

Chapter 1 Introduction

In today's inter-connected world, threats have serious global repercussions. There are, however, two global threats that have not been globally and systematically studied: 1) cyber crime and 2) cyber and biological weapons. Cyber crime from around the world costs society billions of dollars [4]. Countries that offer a favorable environment to cyber criminal activities are used as a basis for launching cyber attacks on end-users around the world. Cyber and biological weapons are asymmetric weapons that could allow small actors to incur massive damage on a major military power. If cyber and biological weapons are used in combination, the impact can be disastrous. For example, a biological attack combined with a cyber attack that cripples hospitals' IT systems can cause mass casualties.

Given that the cyber and biological threat is global, it is important to identify the most concerning countries and design action plans to reduce the threat from these countries. In terms of cyber crime, the most concerning countries are those that foster an environment that is conducive to cyber crime. For example, the environment in these countries may be favorable to setting up and maintaining cyber criminal infrastructure, such as malicious servers. Alternatively, computers in these countries may encounter high rates of malware and become bots as a result. Empirically identifying factors that cause countries' environments to become attractive to cyber crime provides a sound basis to policies to reduce the attractiveness of these countries to cyber crime. This, in turn, could significantly reduce the global cyber crime.

In terms of cyber and biological weapons, the most concerning countries are those that develop such weapons. Many countries have announced intent to build cyber warfare capabilities and established military centers to build such capabilities [87]. Cyber weapons can have a dreadful impact. For example, Russia allegedly launched a massive Denial of Service attack on Estonia that crippled Estonia's information technology infrastructure for three weeks in 2007 [22]. Assessing countries' cyber warfare capabilities can inform policies about the cyber security investment required to keep up with other countries. Such assessment can also help towards attributing future highly sophisticated attacks by limiting the number of actors that are capable of launching such attacks. Biological weapons are weapons of mass destruction prohibited by international law. Despite the outlaw status of these weapons, many countries are believed to have acquired them. In order to reduce the risk of further biological weapon proliferation and the risk of a combined cyber biological attack, it is important to assess countries' biological weapon capabilities. If the international community has an early signal about a country developing biological weapons, the international community could use pressure and negotiations to stop the development of these weapons.

In the context of cyber crime, prior work provides plausible explanations of why some countries' environments are very conducive to cyber crime, but does not empirically test the accuracy of these explanations. For example, Provos et al. [116] observe that China hosts 60% of drive-by-download servers and state that this due to website administrators' poor security practices such as running out-dated and unpatched versions of web server software. Provos et al., however, do not empirically test the accuracy of their explanation. The literature on assessing cyber and biological weapon capabilities mainly contains case studies about capabilities of select countries. A case study provides an in-depth examination of a country's capabilities. However, case studies require substantial effort and expertise and thus usually only focus on the “obvious” countries such as Russia and China. As a result, relying solely on case studies may allow some countries to develop capabilities off-the-radar.

In this thesis, I take a first step towards addressing 2 research questions: 1) What factors cause countries’ environments to become conducive to cyber crime, and 2) How to systematically identify countries that could pose a cyber weapon or a biological weapon threat? In order to address the first question, I empirically test alternative hypotheses about factors behind international variation in cyber crime exposure and hosting. I use Symantec telemetry data Telemetry data which consist of threat reports from 10 million Symantec computers around the world.

I statistically test for the effect of various country-level social and technological factors. In order to address the second research question, I develop two computational methodologies: one that assesses countries’ cyber weapon capabilities and one that assesses countries’ biological weapon capabilities. Both methodologies examine all countries in the world and can be used by non-experts. The two methodologies leverage the fact that the strength of countries’ weapon capabilities depend on countries’ motivations for these weapons and countries’ latent abilities to acquire these weapons. Latent abilities are existing expertise and infrastructure that a country can build on to develop weapons.

It is worth noting that cyber security research rarely distinguishes between cyber crime and cyber weapons. In this thesis, I make such distinction for pedagogical reasons. My analysis of cyber crime focuses on cyber attacks on end-users that can be detected by commercial security products. On the other hand, my work on cyber weapons focuses on cyber attack capabilities by governments. The reader should keep in mind that my definition of cyber crime and cyber weapons is not perfect. Governments may use their cyber capabilities for crime, and not just war. Similarly, non-state actors may also develop cyber weapons.

Table 1.1 provides an overview of the thesis chapters. Chapters 2 and 3 focus on cyber crime. In Chapter 2, I examine countries’ exposure to malware (trojans, viruses, and worms) using Symantec’s Anti-Virus (AV) telemetry data and in Chapter 3, I examine countries’ exposure to and hosting of network-based attacks (exploits, web attacks, and fake applications such as fake anti-viruses) using Symantec’s Intrusion Prevention System (IPS) telemetry data. The telemetry data are accessible through the Symantec Worldwide Intelligence Network Environment (WINE) platform, which provides sampled, anonymized access to data collected from users that have opted in to report telemetry data from Symantec security products [50]. The AV and IPS are two end-host security systems that often run side by side, but do not interact. An AV threat report contains the attack name and the victim computer’s IP address and unique identifier. An IPS threat report contains the same information as an AV threat report in addition to the attacking computer’s IP address. As measure of a country’s attack exposure, I use the average number of attacks encountered by a single Symantec victim computer in the country and as a measure of a country’s attack hosting, I use the average number of attacks launched by a single computer in the country. I statistically test alternative hypotheses about factors behind international variation in attack exposure and hosting. Examples of factors I include in my analysis are computing and monetary resources, web browsing behavior, computer piracy, cyber security research and institutions, corruption, and international relations

My analysis indicates that malware (trojans, viruses, and worms) is most prevalent in Sub-Saharan Africa because of high computer piracy rates in this region. End-users in these region mainly use pirated software and music purchased from street merchants. This pirated software and music is very likely to contain malware. On the other hand, web attacks and fake applications are most prevalent in North America and Western Europe and exploits are most prevalent in countries with emergent economies. My statistical analysis reveals that cyber criminals target these countries because of the abundance of computing and monetary resources in these countries. Finally, I find that many Eastern European countries host disproportionate quantities of attacks. My analysis indicates that the environment in these Eastern European countries is favorable to attack hosting because that environment offers a combination of widespread corruption and a reasonable computing infrastructure. Widespread corruption facilities attack hosting because corrupt law officials and Internet Service Providers tend to turn a blind eye on that malicious hosting.

Chapters 5 and 4 describe my methodologies to assess countries' biological weapon and cyber weapon capabilities respectively. The methodologies are similar in the sense they both consist of a socio-cultural model to assess countries' motivations and indicators to assess countries' latent abilities. More specifically, in order to assess countries' motivations for biological (cyber) weapons, I adapt a well established socio-cultural model [55] to capture factors that motivate countries to develop these weapons and set the parameters of the adapted model using publicly available data. Factors that motivate countries to develop biological weapons are discussed in the literature [10, 64, 76, 149]. An example of such factor is deterrence of biological weapon enemies. Unfortunately, the literature on cyber weapons does not contain discussion of motivational factors. I identify these factors by testing alternative hypotheses against historical proliferation data. I assess countries' latent biological weapon capabilities by examining countries' dual-use biological trade, conventional arms purchase from a biological weapon country, and material power. I assess countries' latent cyber weapon capabilities by examining countries' cyber security research, cyber security institutions, and IT preparedness.

I validate the biological weapon assessment methodology by comparing the methodology's assessment results against historical biological proliferation data. The biological weapon methodology has high accuracy despite the high secrecy surrounding biological weapons. Validating the cyber weapon methodology in a similar manner is the subject of future work. Finally, Chapter 6 compares the two methodologies and identifies countries that could pose both a biological weapon and a cyber weapon threat.

The thesis chapters are intended to be self-containing causing some inevitable redundancy across chapters.

Table 1.1: Overview of thesis chapters

Chapter	**Contribution**	**Approach**	**Data**
Ch 2	Identification of factors behind international variation in exposure to malware (viruses, worms, and trojans)	Statistical hypothesis testing	Symantec anti-virus (AV) telemetry data and various country-level social and technical indicators
Ch 3	Identification of factors behind international variation in exposure and hosting of network-based attacks (web attacks, exploits, and fake applications)	Statistical hypothesis testing	Symantec Intrusion Prevention System (IPS) telemetry data and country-level social and technical indicators
Ch 5	Computational methodology to assess countries' biological weapon capabilities	1) Develop a socio-cultural model to assess countries' motivations for biological weapons 2) Identify indicators of countries' latent biological weapon abilities	international relations, country-level indicators of latent biological weapon capabilities, and biological weapon proliferation timeline
Ch 4	Computational methodology to assess countries' cyber warfare capabilities	1) Develop a socio-cultural model to assess countries' motivations for cyber weapons 2) Identify indicators of countries' latent cyber weapon capabilities	international relations, country-level indicators of latent cyber weapon capabilities, cyber warfare proliferation timeline
Ch 6	Comparison of methodologies to assess biological weapon and cyber warfare capabilities	Comparison of methodologies in Chapters 5 and 4	Data from Chapters 5 and 4
Ch 7	Conclusion	N/A	N/A

Chapter 2 Global Mapping of Malware

Research questions: How does malware exposure vary internationally? What factors explain such variation?

2.1 Introduction

Computers in different countries encounter very different quantities of malware. Empirically identifying factors behind such phenomenon can provide a sound basis to policies to reduce malware encounters in countries that encounter disproportionate quantities of malware. Reducing malware encounters in these countries is likely to benefit end-users worldwide given that we live in an interconnected world.

Many studies notice international differences in malware encounters [19, 20, 97, 158], but mostly hypothesize about reasons behind such differences without empirically testing the validity of such hypotheses. For example, Caballero et al. [19] hypothesize that fake anti-viruses are most prevalent in Western European and North American countries because attackers are interested in taking advantage of the large computing and monetary resources in these countries. Other prior work studies the correlation between users' demographics and attack encounters [20, 82, 91, 111, 127, 158], but does not statistically explain international differences.

In this chapter [100], I examine international variation in the number of trojan, worm, and virus encounters and statistically test hypotheses about factors behind such variation. Such analysis allows policy actions aiming at reducing international malware encounters to rely on scientific empirical evidence instead of simply relying on expert opinions.

I extract the number of malware encounters in different countries from the Symantec Anti-Virus (AV) telemetry data. The AV telemetry data consist of threat reports collected from more than 10 million Symantec customer computers worldwide over the period November 2009 - September 2011. The AV telemetry data are accessible through the Symantec Worldwide Intelligence Network Environment (WINE) platform, which provides sampled, anonymized access to data collected from users that have opted in to report telemetry data from Symantec security products [50]. I use regression analysis to test for the relevance of various factors such as computing resources, web browsing behavior, cyber security expertise, computer piracy, and international relations.

I find that Sub-Saharan African countries are the most affected countries by trojans, worms, and viruses. Many Asian countries are also very affected by malware. The regression analysis reveals that widespread computer piracy especially when combined with poverty is the main reason behind high malware encounters in these countries. In poor countries where piracy is common, users obtain pirated software from P2P networks and merchants that sell pirated CDs publicly with near-impunity. Such pirated software is very likely to contain malware. In rich countries where piracy is common, piracy also takes the form of sharing legitimately purchased software copies among friends, family, and co-workers. Such form of piracy is less likely to propagate malware. Surprisingly, the regression analysis reveals that web browsing behavior, cyber security expertise, and international relations have no significant effect. My study has important policy implications.

Mainly, reducing malware encounters in the most affected regions requires addressing computer piracy. Other policy actions such as providing cyber security training or establishing cyber security institutions are unlikely to be effective if the piracy problem is not addressed.

The remainder of the paper is organized as follows. Section 2.2 discusses related work, Section 2.3 provides background, Section 2.4 presents the data, and Section 2.5 discusses threats to validity. Section 2.6 contains a descriptive analysis of the variation in the number of malware encounters across countries and Section 2.7 contains an explanatory analysis of that variation. Finally, Section 2.8 presents future work and Section 2.9 concludes.

2.2 Related Work

Prior work examines international variation in the number of attacks provides plausible explanations for this variation, but does not empirically test the validity of these explanations. For example, Caballero et al. [19] developed a malware measurement infrastructure and deployed it across 15 countries. The malware measurement infrastructure was used to interact with 4 Pay-Per-Install providers (one type of malware distribution services). Caballero et al. found that fake anti-viruses tend to target Western European and Northern American countries, and conjecture that this is because these countries have abundant monetary resources. Similarly, many security vendors e.g. Microsoft [102], McAfee [97], and Akamai [2] use their corporate data to create reports that depict international variation in the number of attacks. These reports, however, do not statistically test hypotheses about factors behind such international variation.

Moreover, several studies [20, 82, 91, 111, 127, 158] examine the relationship between users' demographics and/or behavior, and malware exposure. That type of user-level analysis is important, but does not provide insight into how various country-level technical, social, economic, and policy factors affect countries' exposure to malware. Understanding the effect of these country-level factors provides an opportunity for large scale reduction in the number of malware encounters. Some of these studies [82, 91] do not examine the effect of country-level characteristics because of the limited international span of their user base. For example, the study by Levesque et al. [82] uses data from 50 subjects in the Montreal area whose laptops were set up to gather data about malware infections and users' behavior. Maier et al. [91] use network traffic data of DSL users from an American university, an European urban area, and an Indian rural area. Covering 3 countries does not allow for the type of global analysis I perform in this chapter. Other studies use social networking platforms [111] or Mechanical Turk [127] to recruit users, but do not report information about these users' countries. Finally, some studies [20, 158] have a wide international coverage, but have a different focus than empirically identifying factors behind international variation in the number of malware encounters. For example, Canali et al. [20] use Symantec data with wide international coverage to explore the possibility of predicting users' risk of encountering malicious web pages based on users' web browsing behavior. Yen et al. [158] use data about malware encounters in a large multi-national to predict the likelihood that hosts encounter malware based on the characteristics of these hosts' users.

2.3 Background

In this section, I discuss factors that may impact the number of malware encounters in different countries. I choose such factors based on observations and hypotheses discussed in prior work.

Computing and monetary resources The majority of attacks nowadays have a monetary goal. Therefore, I expect the availability of large computing and monetary resources to cause an increase in malware encounters. For example, Caballero et al. [19] notice that fake anti-viruses are most prevalent in Western Europe and North America, and hypothesize that this is due to the large resources in these countries.

Cyber security expertise Prior work contains seemingly contradictory evidence about the effect of cyber security expertise. Onarlioglu et al. [111] perform a controlled experiment where experts and non-experts encounter the same attack scenarios. Onarlioglu et al. find that experts are more likely than non-experts to avoid sophisticated attack scenarios. On the other hand, Levesque et al. [82] and Yen et al. [158] find that experts encounter more malware than non-experts when these experts and non-experts perform their daily activities on their own computers. The findings by Levesque et al. and Yen et al. are probably due to the fact that experts use their computers for longer periods of time and to perform more sophisticated tasks.

Web browsing When browsing the web, users may encounter a web attack that triggers a drive-by-download of malware [116]. Canali et al. [20] find that the number of web pages visited as well as the diversity of these pages (as measured by the number of top level domains visited) are correlated with the number of web attack encountered.

Computer piracy Cyber criminals use pirated software to spread malware [73]. End-users obtain pirated software for the functionality, but often become infected as a result [73]. Therefore, we expect computer piracy to cause an increase in malware encounters.

International relations Countries that have many international hostilities may be the target of sophisticated malware such as Flame and Stuxnet [106]. Similarly, countries with many military allies may be less likely to encounter malware since countries usually refrain from attacking their allies. Finally, countries that sign extradition treaties with more countries may encounter less malware because these countries may be able to prosecute cyber criminals from a wide range of countries.

2.4 Data

2.4.1 Malware Encounters

In this section, I explain how I use the AV telemetry data [50] to compute the average number of trojans, worms, and viruses encountered by computer in each country. The AV telemetry data consist of threat reports from more than 10 million Symantec customer computers worldwide. The AV is an end-host system that detects malicious files on computers. Upon detecting a malicious file, the AV quarantines the file and sends a threat report to Symantec. Such threat report contains the threat name, and the reporting machine's IP address and unique identifier. A threat report

Table 2.1: Example of a threat report

Field	Value
Threat name	W32.Aimdes.A@mm
IP address	259.23.78.45
machine ID	104951814

Table 2.2: Example of a threat catalog entry

Field	Value
Threat name	W32.Aimdes.A@mm
Threat family name	Aimdes
Type	worm

example is given in Table 2.1. I use the IP address to identify the computer's country [96]. IP geolocation is very accurate at the country level. I use the unique identifier to distinguish between victim computers.

I complement the AV telemetry data by a threat catalog that contains structured descriptions of threats reported in the AV telemetry data. The threat catalog is extracted from Symantec's online threat descriptions [141]. I provided detailed information about how I extracted the catalog in Section A.1. A threat catalog entry contains the threat name, the threat family name, and the threat type. An example of such entry is given in Table 2.2. The threat name is the unique name that Symantec assigns to the threat, the threat family name is a generalization of the threat name, and the threat type can be trojan, worms, virus, or adware/spyware, etc. The main types in the AV telemetry data are trojans, worms, and viruses as depicted in Figure 2.1.

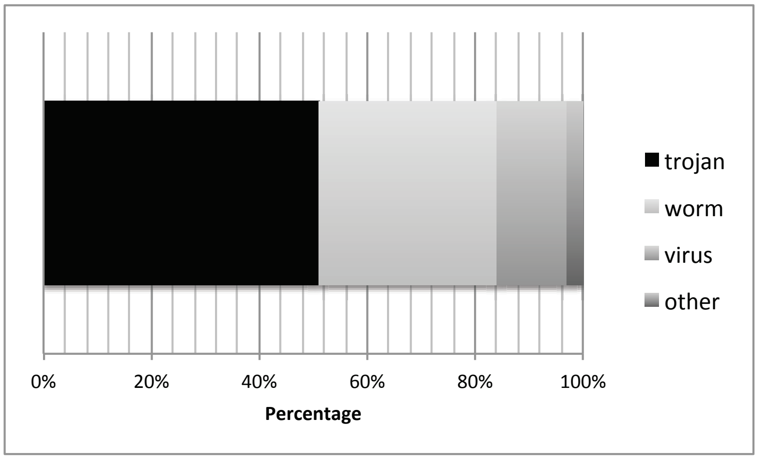
The Symantec customer computers whose threat reports are in the data are randomly sampled among all Symantec customer computers worldwide using a sampling strategy that ensures that the WINE data are representative of all data collected by Symantec [114]. This work covers all 184 countries that have at least 30 Symantec customer computers included in the data. Countries excluded because they have less than 30 of these computers are North Korea, Nauru, Guinea-Bissau, Tuvalu, Eritrea, Cuba, and Kiribati.

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 a user's diligence about updating attack signatures. Symantec uses automatic signature updates, which implies that all online Symantec computers obtain signature updates at approximately the same time, while offline computers obtain these signatures as soon as they become online.

I define the *number of trojans encountered per computer*, the *number of worms encountered per computer*, and the *number of viruses encountered per computer* in a country as the average number of unique threat families of type trojan, worm, and virus respectively that Symantec customer computers in that country encounter. Similarly, I define the *number of all types encountered per computer* in a country as the average number of unique threat families that Symantec customer computers in that country encounter.

I use the average number of malware encountered by Symantec customer computers in a country instead of the total number of malware encountered by these computers in order to allow for a meaningful comparison between countries with different numbers of Symantec customer computers. Moreover, I count the unique number of threat families instead of the number of threat reports because a computer may send multiple threat reports over time about the same infection.

Figure 2.1: Percentage of different threat types in the AV telemetry data



I now consider a toy example in order to illustrate how I compute these measures. Assume that in a certain country there are only two Symantec customer computers A and B. A sends 1 threat report about W32.Aimdes.A@amm (of threat family Aimdes and type worm) and 1 threat report about trojan.Tellafriend (of threat family Tellafriend and type trojan.), B sends 1 threat report about trojan.Tellafriend and 5 threat reports about threat Ada (of threat family Ada and type virus.) The number of trojans encountered per computer in the country is 1, the number of worms encountered per computer is 1/2, the number of viruses encountered per computer is 1/2, and the number of all types encountered per computer is 2.

2.4.2 Explanatory Variables

In this section, I explain how I measure explanatory factors discussed in Section 2.3.

Computing and monetary resources In order to estimate the strength of computing resources in a country, I use the ICT development index and the Internet bandwidth measure from the International Telecommunication Union branch of the United Nation [67]. I estimate the wealth of people in a country using the GDP per capita [26, 146].

Cyber security expertise Measuring the average security expertise of a country's users is difficult. As a proxy, I use the strength of cyber security research and the existence of cyber security institutions such as CERTs in the country.

I estimate the strength of cyber security research in a country by counting the number of cyber security research papers that the country wrote during the period 2002-2011. I take into account research papers published over 10 years because expertise gained in research takes time to transfer to the general public. I collect from SCOPUS [123] all 28,400 research papers that contain "security" in their title or abstract and that belong to the computer science or engineering areas. I consider that a country wrote a paper if at least one of the authors has an affiliation in that country.

I obtain the list of countries that have cyber security institutions by combining lists from multiple sources [28, 66, 86]. I construct a binary variable that captures whether a country has such institutions.

Web browsing I use the average number of unique web pages visited (hits) and the average number of top level domains visited (tops) in a country from Canali et al. [20]. Canali et al.

compute these estimates based on data collected from a subset of Symantec customers that agree to share their web browsing histories with Symantec.

Computer piracy I use the piracy rate from the Business Software Alliance (BSA) [18]. The piracy rate in a country is the ratio of the number of unlicensed software units installed to the total number of software units installed. The BSA collects such data by surveying users in different countries about their practices.

International relations Based on the list of military and non-military hostilities [24, 48] during the period 1992-2010, I construct a binary country-by-country hostility network $H = [h_{ij}]$ where $h_{ij} = 1$ indicates the existence of a hostility between i and j, and $h_{ij} = 0$ indicates otherwise. I then compute *hostility betweenness* as countries' betweenness centrality in that network. Betweenness centrality [54] is a standard measure of the importance of a node in a network. More specifically, betweenness centrality measures the extent to which a node falls in shortest paths among other nodes.

I proceed similarly for international alliances and extraditions. I construct a binary alliance network and a binary extradition network based on the list of international military alliances [40] and the list of international extradition treaties [151] respectively. I then compute *alliance betweenness* and *extradition betweenness* as countries' betweenness centrality in the alliance network and extradition network respectively.

2.5 Threats to Validity

Many limitations of this paper mainly stem from limitations of the data I use. First, Symantec AV data inform about malware encounters of Symantec AV home users. Malware encounters of home users of other AV products have no reasons to be different. However, malware encounters of corporate users and users that use no AV protection may be different. Unfortunately, I am unable to correct for this bias because I am unaware of any study that compares malware encounters of different types of users at a global scale. Moreover, the data cover the time period 2009-2011 and the world has moved towards smart phones since then. It is worth noting, however, that the move towards smart phones is unlikely to affect the malware situation is Sub-Saharan Africa where smart phone penetration is still very low. Furthermore, the Symantec AV telemetry data only cover malware detected by Symantec AV. Sophisticated malware such as targeted attacks and zero-day attacks are missing in the data. It is worth noting, however, that targeted attacks and zero-day attacks are a small minority compared with other attacks [12, 147].

One risk associated with the use of regression analysis is endogeneity which occurs when the error term is correlated with one or more regressors (independent variables). Two phenomena could cause endogeneity in the regression in this chapter: omitted variable bias and simultaneous causality. Omitted variable bias occurs when one variable that is determinant of the dependent variable and is correlated with a regressor is omitted from the regression. In this work, my strategy to prevent omitted variable bias is to include in the regression all variables that the literature says could be relevant. Simultaneous causality occurs when there is a causal link from the dependent variable to a regressor in addition to the causal link from the regressor to the dependent variable. In this work, there may be a simultaneous causality link between malware exposure and cyber security research and institutions. In ohter words, countries that have encountered large quantities

Table 2.3: Top 10 countries on the number of malware encounters per computer

Country	# trojans	Country	# worms	Country	# viruses	Country	# all types
Burundi	1.56	Solomon Islds	8.76	Chad	1.99	Solomon Islds	17.51
Solomon Islds	1.43	Bhutan	5.94	Somalia	1.78	Swaziland	13.14
Bhutan	1.24	Cent. Afr. R.	5.55	Burundi	1.64	Bhutan	13.09
Ethiopia	1.20	Swaziland	5.38	Cent. Afr. R.	1.58	Chad	11.78
Cambodia	1.19	Chad	5.36	Swaziland	1.57	Cent. Afr. Rep.	11.68
Yemen	1.19	Samoa	5.05	Solomon Islds	1.36	Burundi	11.10
Cent. Afr. R.	1.18	Somalia	4.68	Gambia	1.29	Somalia	10.84
Swaziland	1.18	Burundi	4.46	Togo	1.28	Samoa	9.88
Chad	1.17	Sierra Leone	4.05	Cameroon	1.23	Sierra Leone	9.10
Syria	1.13	Lesotho	3.77	Sierra Leone	1.23	Togo	9.08

of malware in the past may have encouraged cyber security research and institutions to address the problem. Because of this potential simultaneous causality, the coefficients on cyber security research and institutions may not be reliable. As future work, it would be interesting to address this simultaneous causality issue using instrumental variables or fixed point models.

It is also worth mentioning that the Symantec labeling of attacks that I use in this work is imperfect. Inconsistencies among AV vendors in labeling malware have been reported [6, 21, 104]. However, I believe that the lack of a unified malware labeling taxonomy is the main reason behind such inconsistencies. The fact that different AV vendors label malware differently does not necessarily indicate that the labeling from any AV vendor is wrong. Many research papers [8, 65, 118] use AV labels as a ground-truth when evaluating the accuracy of new approaches.

Finally, I rely on IP geolocation to identify countries where computers are. Threats to validity that usually come with using IP geolocation are IP spoofing and the use of anonymization software. I do not believe that IP spoofing is an issue for our data given Symantec customer computers have no reason to spoof their address when sending threat reports to Symantec. Similarly, the use of anonymization software such as Tor is unlikely to cause a significant bias because only a small minority of the general population uses such software. Pirated software users also typically do not use anonymization software. In places such as Sub-Saharan Africa, pirated software is sold publicly in the streets and there is no penalty for using such software that would cause users to hide their identity using anonymization software [15].

2.6 Descriptive Analysis

Table 2.3 contains the list of countries that rank highest on the number of malware encounters. It can be seen that these countries are primarily poor countries in Sub-Saharan Africa. This is surprising given that Caballero et al. [19] find that fake applications and web attacks are most prevalent in Western Europe and North America. This finding probably indicates that factors that drive encounters of trojans, worms, and viruses are very different from factors that drive encounters of web attacks and fake applications.

Figure 2.2 presents a map visualization of the number of malware encounters in all countries. It can be seen that Sub-Saharan Africa encounters the largest quantities of trojans, followed by Asia and North Africa. Other regions encounter relatively small quantities of trojans. Moreover, it can be seen that Sub-Saharan Africa is also the most affected region by worms and viruses, followed by South-East Asia and South Asia. Other regions encounter relatively small quantities of worms and

Table 2.4: Summary statistics of variables used in the explanatory analysis

Statistic	N	Mean	St. Dev.	Min	Max
Trojans encountered per computer	184	0.55	0.25	0.22	1.56
Worms encountered per computer	184	1.42	1.34	0.045	8.76
Viruses encountered per computer	184	0.34	0.39	0.031	1.99
All types encountered per computer	184	4.06	2.64	0.94	17.51
Bandwidth	184	30.95	57.66	0.10	547.10
ICT	184	3.82	2.05	0.85	8.45
GDP per capita (log)	184	8.49	1.53	5.29	11.55
Web hits	184	1,020.52	586.86	105	5,363
Web tops	184	12.89	4.11	1.00	32.33
Piracy	184	0.65	0.21	0.20	0.93
Security research	184	175.50	830.41	0	7,911
Security institutions	184	0.36	0.48	0	1
Alliance betweenness	184	0.001	0.007	0	0.053
Hostility betweenness	184	0.0004	0.002	0	0.017
Extradition betweenness	184	0.004	0.036	0	0.483

viruses. Finally, when taking into consideration all malware types, the resulting pattern is similar to the pattern of trojans. This is due to the fact that trojans constitute the majority of malware types in the data as depicted in Figure 2.1.

2.7 Explanatory Analysis

The goal of this section is to identify factors that explain international variation in malware encounters. In order to identify these factors, I use regression analysis which aims at establishing causation, and not just correlation. Table 2.4 contains summary statistics of the variables I use in this section. Each variable is of length 184, the number of all countries in the world that have at least 30 Symantec customer computers whose threat reports are included in the AV telemetry data as I explain in Section 2.4.1. Trojans encountered per computer, worms encountered per computer, viruses encountered per computer, and all types encountered per computer are the dependent variables I will use in the regression analysis. These variables represent the average number of trojans, worms, viruses, and all malware types respectively encountered by a single Symantec customer computer [1] as explained in Section 2.4.1. Other variables are the explanatory variables I intend to use in the regression analysis and that were explained in Section 2.4.2.

Table 2.5 represents the correlation table between the explanatory variables. Each element in the table represents the correlation between two explanatory variables. For example, the correlation

[1] The careful reader will notice that all types encountered per computer is *not* the sum of trojans, worms, viruses and other types encountered per computer. This is because these measures are scaled by the number of Symantec customer computers in each country. If we were to consider the unscaled measures, we would find that all types encountered by *all* computers in the world is the sum of trojans, worms, viruses, and other types encountered by *all* computers in the world.

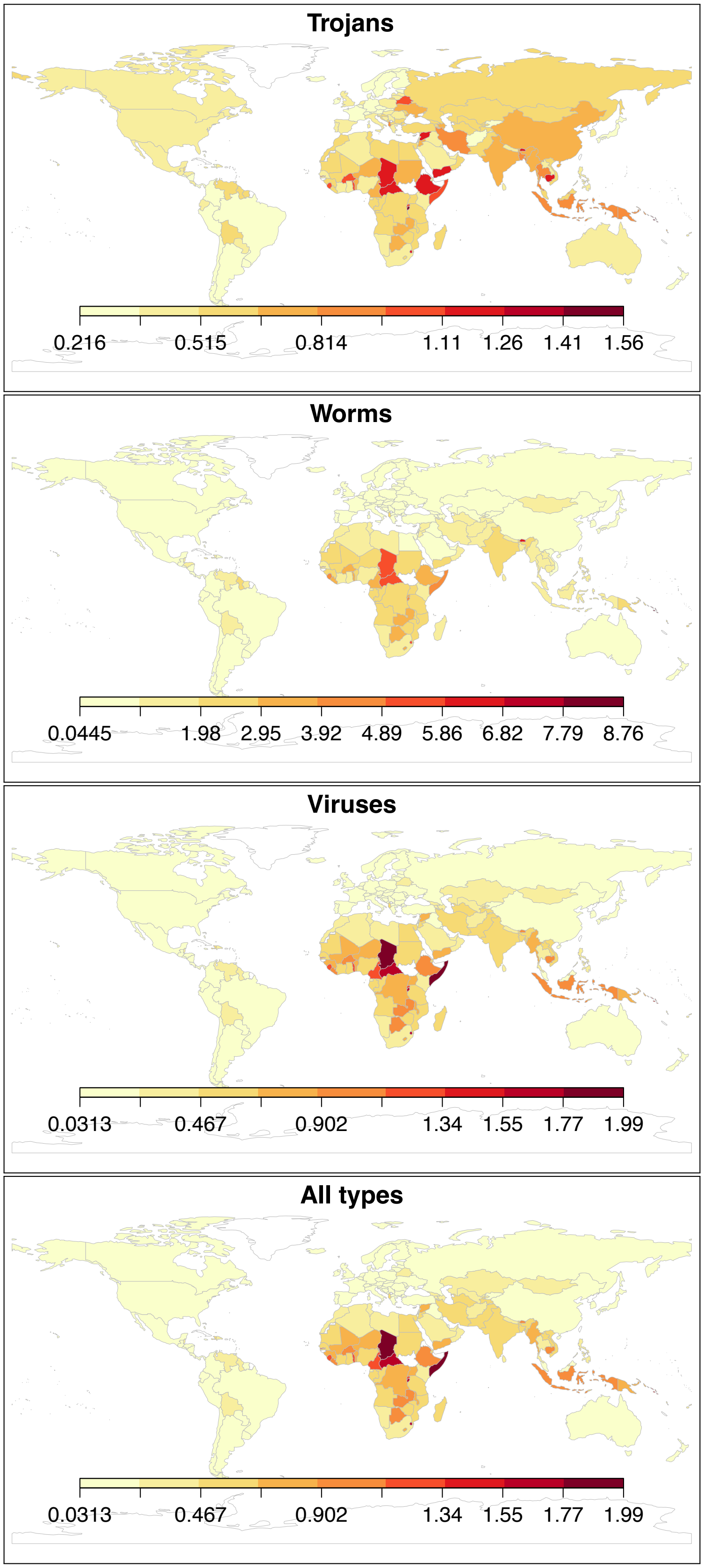


Figure 2.2: Map visualization of the number of threats encountered per computer

Table 2.5: Correlation table between explanatory variables

	Bandwidth	ICT	GDP PC	web hits	web tops	piracy	research	institutions	alliance	hostility
ICT	0.60***									
GDP PC	0.52***	0.92***								
web hits	0.13	0.25***	0.24**							
web tops	-0.04	0.02	0.00	0.65***						
piracy	-0.54***	-0.85***	-0.85***	-0.23**	0.04					
research	0.09	0.23**	0.20**	0.10	-0.10	-0.24**				
institutions	0.34***	0.56***	0.47***	0.14	-0.03	-0.49***	0.27***			
alliance	0.10	0.21**	0.18*	0.06	-0.06	-0.21**	0.46***	0.20**		
hostility	0.03	0.02	0.00	0.00	-0.01	-0.03	0.38***	0.07	0.30***	
extradition	0.03	0.13	0.12	0.06	-0.05	-0.19*	0.70***	0.12	0.58***	0.55***

*p<0.1; **p<0.05; ***p<0.01

between ICT and bandwidth is 0.60. In the correlation table, I only report values below the diagonal because correlation is symmetric. For example, the correlation between GDP PC and ICT is the same as the correlation between ICT and GDP PC. From Table 2.5, it can be seen that GDP PC is highly correlated with ICT, that is countries with high income per individual tend to have strong computing infrastructure. Because of this very high correlation value (0.92), I only use one of these two variables in the regression analysis. From the table, it can also be seen that computer piracy is more prevalent in countries with low income. This reflects the fact that low-income individuals are unable to afford to legitimate software and music, and resort to pirated products instead.

Table 2.6 presents the results of the regression analysis. From the table, it can be seen that the only factors that have statistically significant coefficients are piracy and piracy x ICT. This indicates that piracy and piracy x ICT are the only factors that are confirmed to affect international variation in malware exposure according to my analysis. In other words, computer piracy especially when combined with poverty is the main factor behind high malware encounters. The regression coefficients in the table are standarized, which allows comparing the effect of different factors. For example, the 0.493 coefficient on piracy in the regression about exposure to trojans can be interpreted as: increasing the piracy rate by 0.21 (the standard deviation of the piracy rate according to Table 2.4) results in an increase in the number of trojans encountered by 0.493 * 0.25 (where 0.25 is the standard deviation of the number of trojans encountered according to Table 2.4).

In poor countries where piracy is common, people obtain pirated software and music through P2P networks and merchants that sell these products publicly in the streets with near-impunity [15]. These merchants obtain pirated products and license keys from dubious Internet websites. As a result, pirated software in these poor countries is very likely to contain malware. On the other hand, in relatively rich countries where piracy is common such as Singapore, piracy also takes the form of sharing legitimately purchased products among friends, family, and co-workers because the collectivism culture encourages sharing and helping others [5, 128]. This form of piracy is less likely to spread malware.

From Table 2.6, it can be seen that web browsing behavior has surprisingly no significant effect on international variation in the number of malware encounters. Web browsing has no significant impact because, as explained in the previous paragraph, computers encounter malware primarily through pirated computer products, and not through web attacks and drive-by-downloads. Cyber security expertise, too, has surprisingly no significant effect. Many people are aware that pirated products are less safe [18], but prefer such pirated products because of economic, social, and legal reasons [18]. Finally, it is surprising that international relations have no significant effect. Despite the wide media coverage of state-sponsored malware, the volume of such malware is very small compared with criminally motivated malware [147]. Moreover, state-sponsored malware tends to

Table 2.6: Regression Analysis of the number of malware encounters in different countries. Regression coefficients are standardized.

	Trojans	Worms	Viruses	All types
Computing and monetary resources				
Bandwidth	-0.010	0.031	0.013	0.022
	(0.081)	(0.069)	(0.060)	(0.070)
ICT	-0.068	-0.387	0.302	-0.313
	(0.285)	(0.242)	(0.211)	(0.247)
Web browsing				
Web hits	0.026	0.036	0.021	0.032
	(0.087)	(0.074)	(0.064)	(0.075)
Web tops	0.090	0.018	0.024	0.047
	(0.085)	(0.072)	(0.063)	(0.074)
Piracy				
Piracy	0.493**	0.149	0.813***	0.282
	(0.222)	(0.189)	(0.164)	(0.193)
Piracy x ICT	-0.244*	-0.299**	-0.684***	-0.285**
	(0.145)	(0.123)	(0.107)	(0.125)
Security research and institutions				
Security research	0.082	-0.059	-0.085	-0.043
	(0.090)	(0.077)	(0.067)	(0.078)
Security institutions	0.079	-0.070	0.056	-0.010
	(0.076)	(0.065)	(0.056)	(0.066)
International relations				
Alliance betweenness	0.016	0.011	0.009	0.011
	(0.077)	(0.065)	(0.057)	(0.066)
Hostility betweenness	0.023	0.010	0.054	0.016
	(0.074)	(0.063)	(0.055)	(0.064)
Extradition betweenness	-0.023	0.013	0.015	0.014
	(0.104)	(0.088)	(0.077)	(0.090)
Observations	184	184	184	184
R^2	0.352	0.534	0.647	0.514

Standard errors in parentheses
*p<0.1; **p<0.05; ***p<0.01

be very sophisticated and is often undetected by anti-viruses.

2.8 Future Work

As future work, it would be interesting to examine measures beyond the average number of malware encounters per computer. For example, it would be interesting to examine the median and the ratio of Symantec customer computers that encounter malware out of all Symantec customer computers in a country. Another future work direction is to use data from other AV vendors such as Microsoft and McAfee. Moreover, in the future, I intend to use a stepwise approach such as the Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) approach in the regression analysis. Finally, case studies about piracy and malware encounters in Sub-Saharan Africa would shed greater light into problematic practices in these countries and appropriate policy actions to address these practices.

2.9 Conclusion and Discussion

In this chapter, I empirically test hypotheses about factors behind international variation in the number of malware encounters. Such hypothesis testing provides an empirical scientific basis to the community's understanding of global malware encounters and to policy actions to reduce such encounters. To the best of my knowledge, this is the first research piece that statistically tests hypotheses about reasons behind international differences in the number of trojans, worms, and viruses encountered.

I extract the number of malware encounters in each country from the Symantec AV telemetry data. That data consist of threat reports from more than 10 million Symantec customer computers worldwide. I use regression analysis to test for the relevance of computing and monetary resources, web browsing behavior, computer piracy, security expertise, and international relations.

I find that trojans, worms, and viruses are most prevalent in Sub-Saharan Africa. The regression analysis reveals that this is mainly due to widespread computer piracy combined with poverty in this region. In poor countries where piracy is widespread, users obtain pirated computers products from P2P networks and merchants that sell pirated CDs on the streets with near impunity. These merchants obtain these pirated products from dubious parts of the Internet. On the other hand, in rich countries where piracy is widespread, it is more common for a person to buy legitimate software or music and share it with friends, family, and coworkers.

The regression analysis reveals many surprising findings. For example, given that malware have primarily a monetary goal, I hypothesized that cyber criminals would target rich countries because the monetary benefit of compromising rich people's computers is higher. My hypothesis is consistent with what Caballero et al. [19] observed about fake anti-viruses. In this work, I found that, surprisingly, poor countries in Sub-Saharan Africa are most exposed to malware. One possible explanation is that cyber criminals target computers in Sub-Saharan Africa because of the low cost of attacking computers in this region. Another possible explanation is that cyber-criminals do not target this region on purpose. It is just that one malware distribution method, mainly computer piracy, happens to be particularly popular in Sub-Saharan Africa.

My work has many policy implications. Reducing malware encounters in Sub-Saharan Africa requires reducing computer piracy in this region. Other policy actions such as providing cyber security training are unlikely to be effective. Software and music industries are the entities that have the highest incentive to combat computer piracy in order to increase revenues. Governments in this region are aware that legitimate software is more stable and safer than pirated software. However, these governments are reluctant to fight computer piracy despite lobbying by the software industry because such fight would prevent their populations from keeping up with international knowledge and technology. For example, the GDP per capita is 700 USD in Central African Republic [26], while Mircosoft Windows costs 119 USD [101]. It is unreasonable to believe that poor people would pay more than 15% of their yearly income in order to acquire a legitimate copy of Microsoft Windows. I join Gobal [128] in suggesting that the software and music industries should adjust their international prices to countries' income level. This could be a win-win solution because the software industry would be able to collect some revenue instead of losing almost all potential revenue to piracy and these countries' populations can have access to safe and robust computer products.

Meanwhile, given that malware can be "contagious", it is important to reduce the risk that malware from these countries spreads elsewhere. First, global organizations with offices in the most at-risk countries should have very strong, enforced, and enforceable policies for monitoring for, and correcting for malware. Second, those traveling to the most at-risk countries should be particularly wary of linking their own machines to machines in these countries and transferring material from the machines in these countries without having sufficient malware detectors and barriers in place. Third, soft-power solutions that raise awareness about the issue and provide anti-viruses to these countries could reduce the overall global risk.

Chapter 3 Global Mapping of Network-Based Attacks

Research questions: How do network-based attack exposure and hosting vary internationally? What factors explain that variation?

3.1 Introduction

In some countries, computers encounter disproportionate quantities of attacks, while in other countries, computers host disproportionate quantities of these attacks [19, 97, 158]. Such phenomenon has many plausible explanations such as the shortage of cyber security expertise, the abundance of resources or the computing culture in these countries. Testing the accuracy of these explanations against real data can inform policy makers about how to reduce attack exposure and hosting in the most affected countries. This will, in turn, have a worldwide benefit given that we live in an interconnected world. Testing alternative hypotheses or explanations is considered a very important and novel contribution in the social and natural sciences [115] and is strongly encouraged under the concept of science of security [83, 129].

Prior work [2, 19, 97, 102] notices international variation in attack exposure and hosting, and provides plausible explanations for such variation. However, prior work does not empirically test the validity of these explanations. For example, Caballero et al. [19] notice that fake anti-viruses are more prevalent in Europe and North America and conjecture that fake anti-viruses target these regions because users there tend to be wealthy. Other prior work [20, 82, 91, 111, 127, 158] examine the relationship between attack exposure and users' demographics and behavior, but do not statistically explain international variation.

In this chapter, I statistically test alternative hypotheses about factors behind international variation in attack exposure and hosting. More specifically, I address 3 related research questions: (1) What factors explain international variation in attack exposure?, and (2) What factors explain international variation in attack hosting?, and (3) What factors explain how attacks propagate across countries?

I use Symantec's World Intelligence Network Environment (WINE) Intrusion Prevention System (IPS) telemetry data. WINE is a platform for repeatable experimental research through which external researchers can access data used at Symantec Research Labs [50]. The IPS is an end-host system that detects and blocks malicious network activity. The IPS telemetry data contain attack reports from more than 10 million Symantec customer computers worldwide collected during November 2009 - September 2011. The main attack types in the data are exploits, web attacks, and fake applications (mainly fake anti-viruses) because the IPS exclusively examines network traffic. I statistically test for the effect of various technical, social, economic, and political factors on the international variation in cyber attacks.

I use the average number of attacks that a single computer encounters as a measure of a country's exposure to attacks. I find that exploits are most prevalent in countries with emergent economies such as India and Taiwan, while web attacks and fake applications are most prevalent in Western Europe and North America. My analysis confirms that these countries encounter more attacks because cyber criminals are interested in taking advantage of the abundant computing and monetary resources in these countries. Surprisingly, countries' cyber security expertise (that I measure through the strength of cyber security research and the existence of cyber security institutions) does not reduce the countries' cyber attack exposure. Moreover, international relations have no significant effect on attack exposure.

Furthermore, I use the average number of attacks from a single computer in the country as a measure of the extent to which a country hosts attacks. I find that many Eastern European countries and a few Latin American rank highest on this measure, while African countries rank lowest. The regression analysis reveals that a combination of reasonable computing resources and high levels of corruption is a very favorable condition to hosting criminal computing infrastructure. Reasonable computing resources ensure that malicious computers can aggressively deliver attacks, while corruption facilitates conducting cyber-criminal activities through the complicity of law officials and ISPs. Surprisingly, again, cyber security expertise in a country does not reduce attack hosting in a country. Similarly, international relations have no significant effect on attack hosting.

In order to investigate how cyber-attacks propagate internationally, I examine a cyber-attack international network where an edge weight from a country A to country V represents the average number of attacks from a single computer in A on a single computer in V. I find that exploits tend to propagate to geographically nearby countries. On the other hand, malicious web servers that serve web attacks and fake applications tend to be in Eastern Europe and Latin America, while victims tend to be in Western Europe and North America.

My results imply that effectively addressing cyber security requires addressing social issues, and not just designing more secure systems or providing cyber security training. Countries that excessively host attacks have sufficient cyber security expertise to reduce the likelihood that their systems are used to host attacks. However, widespread corruption in these countries causes law officials and ISPs to turn a blind eye to cyber crime hosting hosting. Effectively reducing attack hosting requires cracking down on the most corrupt ISPs and law officials.

The remainder of the chapter is organized as follows. I provide background in Section 3.3, describe my data in Section 3.4 and explain threats to validity in Section 3.5. I examine attack exposure in Section 3.6, attack hosting in Section 3.7, and international attack networks in Section 3.8. I discuss future work in Section 3.9 and conclude in Section 3.10.

3.2 Related Work

Many studies e.g. [2, 41, 97, 102, 116] observe international differences in the number of attacks encountered and hosted, but do not statistically test hypotheses about factors behind such differences.

Moreover, prior work [20, 82, 91, 111, 127, 158] examines the relationship between users' demographics and/or behavior, and malware exposure. However, this prior work performs analyses at the user level instead of the country level, and thus does not provide the same opportunity for large scale reduction in attack exposure and hosting that this work provides.

Finally, prior work [71, 72, 137] develops techniques to expose malicious or negligent Internet Service Providers (ISPs), but does not statistically analyze why malicious and negligent ISPs emerge in some regions of the world more than others. Malicious ISPs e.g. the Russian Business Network [14] are criminal organizations that offer bullet-proof hosting to a wide range of malicious activities such as malware, spam and child pornography, whereas negligent ISPs are legitimate organizations that neglect taking down malicious activities.

3.3 Background

3.3.1 Cyber Attacks

The main attack types in the IPS telemetry data are exploits, web attacks and fake applications. I briefly 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].

Web attacks. Web attacks are exploits on web browsers or web browser plugins that typically allow installing malware in what is called a drive-by-download [116]. 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 e.g. iFrames or malicious java-script.

A Pay-Per-Install (PPI) business model [19, 59] to deliver malware has emerged around web attacks and drive-by-downloads. In this model, there are find clients, PPI providers, and affiliates. Clients have malware that they are interested in disseminating. For example, clients can be the malware authors. 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. 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 typically ask the victim to pay a premium to remove the malware, which some victims do.

3.3.2 Factors Impacting Attacks Exposure

The number of attacks that a certain country's users encounter depend on multiple factors including users' behavior and attackers' desire to attack these users. In this section, I present hypotheses discussed in prior work about factors that affect attack exposure.

Web browsing. When browsing the web, a user may encounter a malicious web page that launches a web attack, or may see an advertisement for a fake application and be tempted to download that application. Alternatively, the user may see an advertisement for a fake application and be tempted to download that application. Canali et al. [20] find that exposure to web attacks is correlated with the number of web pages visited and the diversity of these web pages (as measured by the number of top level domains visited).

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

Computing and monetary resources. The majority of attacks nowadays have a monetary goal. Therefore, I expect I expect the abundance of resources to increase attack exposure [19]. For example, attackers may prefer to attack fast computers with high Internet bandwidth in order to use these computers to launch other attacks. Similarly, stealing credit card information of rich victims is likely to generate larger profits.

Currently, attackers have mechanisms to target victims in certain countries. For example, in the Pay-Per-Install (PPI) business model discussed in Section 3.3.1, the rate that clients pay to PPI providers varies depending on the desired location of victims. In the United States and the United Kingdom, the rate is \$100-\$180 per 1000 computers, whereas is less demanded regions such as some Asian countries, the rate is \$7-\$8 [59].

Cyber-security expertise. Conventional wisdom says that cyber-security experts should encounter less attacks, but empirical work is not very conclusive about this point. Onarlioglu et al. [111] compare the performance of experts and non-experts at avoiding a set of attack scenarios, and finds that experts outperform non-experts at avoiding sophisticated attack scenarios. On the other hand, Levesque et al. [82] and Yen et al. [158] examine attack encounters of users as these users perform their daily computing activities, and find that experts encounter *more* attacks than non-experts.

One possible explanation for the discrepancy between these results is that, experts encounter more attack scenarios in real-life because of their higher and more advanced computer usage.

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 [122]. Similarly, countries are usually less likely to attack their allies, and thus countries with many allies may experience less cyber attacks.

3.3.3 Factors Impacting Attack Hosting

In order to host attacks, attackers either use 1) compromised computers or 2) malicious computers set up exclusively to host attacks[41, 116]. These malicious computers could be, for example, hosted within rogue Internet Service Providers (ISPs) that offer bullet-proof hosting [137]. Such ISPs keep the malicious computers up despite complaints and actions to take such servers down [137]. In the remainder of this section, I present hypotheses about factors that cause attackers to prefer using computers in a certain country to host attacks.

Computing resources. Cyber criminals may prefer to host their attacks in countries that have a strong IT infrastructure. Hosting in these countries is usually more reliable and cheaper. As long as cyber criminals ensure that the attacks hosted in these countries are undetected, these cyber criminals can enjoy a reliable service at a low price, thus maximizing their profit margins.

Corruption. Corruption is known to facilitate criminal activity [9, 29, 80, 94] through the complicity of law officials. Thus, cyber criminals may be interested in hosting attacks in countries with widespread corruption. This is because, in these countries, cyber criminals do not need to worry about remaining undetected by hosting services and law officials.

Attack Exposure. Exposure to attacks increases the likelihood that computers become compromised and start hosting attacks. Therefore, I expect attack exposure to increase attack hosting within compromised computers.

Cyber security expertise. When used by legitimate actors in a country, cyber security expertise can help prevent computer compromise and disinfect compromised computers. In that sense, cyber security expertise may reduce attack hosting within compromised computers.

On the other hand, cyber criminals may use their expertise to compromise computers and set up malicious ones. However, since cyber criminals can perform these operations remotely, I do not necessarily expect expertise of cyber criminals living in a country to increase attack hosting in *that* country.

International relations. Attackers may be discouraged from hosting malicious computers in a country A to launch attacks on a country V if A and V collaborate on cyber-security issues. Such collaboration may be based on formal agreements such as extradition treaties or informal agreements [90]. Informal agreements may be easier among military allies, and harder among military enemies.

3.3.4 Factors Impacting International Attack Propagation

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

Country attributes. Countries' attributes discussed in Section 3.3.2 likely have an impact the number of attacks V encounter from other countries similar to the impact discussed in that section. Similarly, countries' attributes discussed in Section 3.3.3 likely have an impact on the number of attacks A launches on other countries similar to the impact discussed in that section.

Geographical proximity. I expect to see more attacks among neighboring countries because some attacks use propagation strategies that favor geographically close computers. As example of such attacks are those that use random scanning techniques that favor local computers.

3.3.5 MrQAP Regression

MrQAP regression [78] is a regression technique suitable for network data. Ordinary Least Squares (OLS) regression is unsuitable for network data because such data violates the independence assumption required for OLS. MrQAP regression on networks produces the same regression coefficients as OLS regression produces on the vector representation of these networks, where a vector representation of a network is a vector obtained by concatenating the rows of that network. However, contrary to OLS, MrQAP produces accurate p-values that account for intra-column and intra-row dependencies. MrQAP produces these p-values by leveraging the Quadratic Assignment Procedure (QAP) test, which is a non-parametric test based on random permutations of rows and columns.

3.4 Data

3.4.1 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

Figure 3.1: Attack report generation

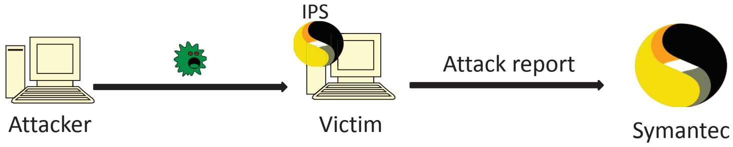


Table 3.1: Attack report example

Field	**Value**
Attack name	Web Attack: Blackhole Toolkit Website
computer ID	104951814
IP victim	172.268.12.156
IP attacker	157.23.56.589
Protocol	TCP

in Figure 3.1. As illustrated in Table 3.1, an attack report contains the name of the attack detected, the IP address and unique identifier of the victim computer [2], the IP address of the attacker computer and the network protocol of the malicious activity blocked. The unique identifier of the victim computer is the serial identifier associated with the installation of Symantec IPS on the machine.

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 a user's diligence about updating attack signatures. Symantec uses automatic signature updates, which implies that all online Symantec computers obtain signature updates at approximately the same time, while offline computers obtain these signatures as soon as they become online.

Attack catalog. The IPS attack catalog described in Section A.2 contains structured descriptions of attacks reported in the IPS WINE telemetry data, and was extracted from Symantec's online attack descriptions [142]. For a each attack, the catalog contains the attack name, the attack family names, the types, and the attack infrastructure type. The attack name is the name used by Symantec to uniquely identify the attack and the attack family name is a generalization of that attack name. Type is the type of the attack. The main attack types are exploits, web attacks, and fake applications. The attack infrastructure type is the type of malicious computers that launch that attack. The infrastructure type of web attacks and fake applications is mostly "malicious web page", but we also find "hacked web page". A malicious web server is a web server that delivers web attacks or malware. This is in contrary to hacked web pages that contain iFrames or malicious javascript and direct to these malicious web pages. The infrastructure type of exploits is "exploiting computer" which indicates that the attacker computer is a computer that launches the exploit, but no further information about such computer is available. Table 3.2 provides an example of an attack catalog entry.

Figure 3.2 presents the distribution of attack types in the IPS telemetry data. 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 exclusively examines hosts' network activity. The IPS is unable to detect threats that have no network activity.

[2] In this paper, I 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. I 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 3.2: Examples of an attack catalog entry

Field	**Value**
Attack name	Web Attack: Blackhole Toolkit Website
Attack family name	Blackhole
Type	Web attack
Attack infrastructure	Malicious web page

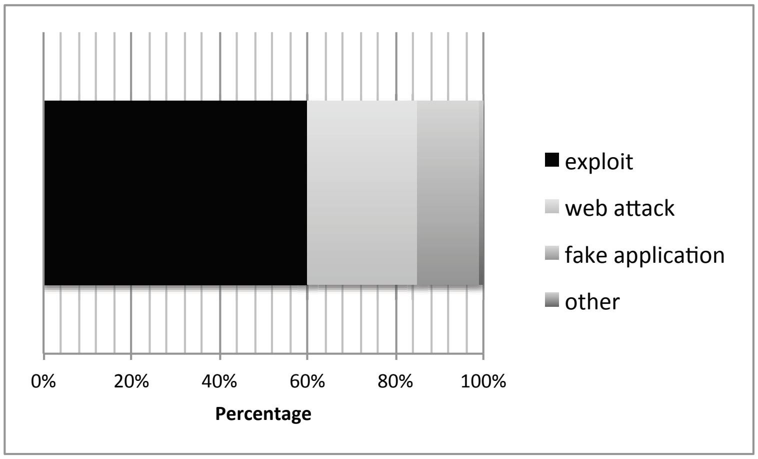


Figure 3.2: Distribution of attack types

3.4.2 Country Cyber Attack Data

In this section, I explain how we compute the number of attacks encountered and the number of attacks launched per computer.

I start by explaining how I define an *attack instance*. I 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, I consider all attack reports about an attacker computer a, a victim computer v and an attack family f to be a single attack instance. I use the unique computer IDs to distinguish between different victim computers and the attacker IP addresses to distinguish between different attacker computers. Such definition does not correctly account for when a attacks v multiple times using f, but such definition is preferable because it correctly accounts for when a victim sends a large number of reports over time about an infection or a quarantined threat. In the toy example in Figure 3.3, 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).

I define a *WINE computer* as a Symantec customer computer whose attack reports are included

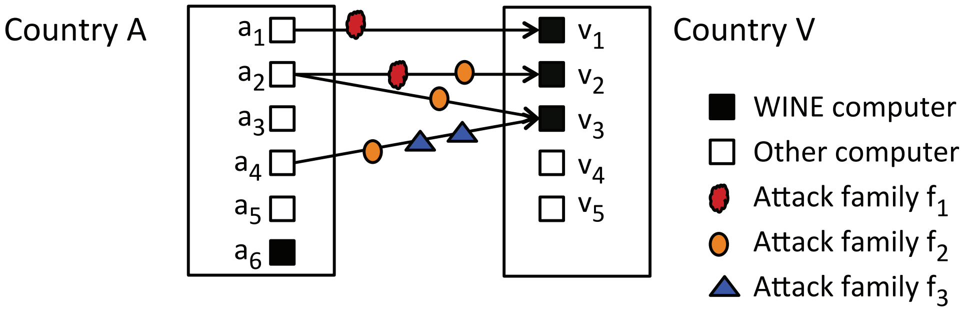


Figure 3.3: Toy example.

in the WINE telemetry data. The nearly 10 million WINE computers are randomly chosen from all Symantec customer computers worldwide. I determine the country where a computer is using IP geolocation [96]. IP geolocation at the country level is very accurate. One risk, however, is IP spoofing by attacking computers. IP spoofing is easy when TCP connections are unnecessary as is the case with UDP traffic and TCP SYN Denial of Service attacks. On the other hand, IP spoofing is relatively difficult when a TCP connection is required between the attacker and victim computers. A remote computer that spoofs its IP address, and that does not have access to the path between the spoofed IP address and the victim computer does not see the TCP sequence numbers. Therefore, the remote computer is unable to establish a TCP connection with the victim computer. In the IPS telemetry data, more than 98% of attack instances correspond to TCP traffic that is not Denial of Service traffic, which is consistent with the fact that TCP is much more common than UDP in the Internet [70]. Therefore, I expect IP spoofing to have a limited effect on this work.

In this work, I only keep data from countries with at least 30 WINE computers. That is, I 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 0.013% and standard deviation $0.23\ 10^{-3}$. I obtain estimates of the total number of computers in countries based on estimates of the number of computers by 100 people [145] and estimates of the population size [146] in these countries.

Attacks encountered per computer. This is the average number of attack instances encountered by a single WINE computer in each country. In the toy example in Figure 3.3, 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 country country. For each country, I divide the number of attack instances where the attacker computer is in that country by the total number of computers in that country. I 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 3.3, 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 A on a single WINE computer in V. In order to compute the edge weight from A to V, I count the number of attack instances where the attacker computer is in A and the victim computer is in V. Then, I divide that number by the product of the number of all computers in A and the number of WINE computers in V. In the toy example in Figure 3.3, the edge weight from A to V is 6 divided by 18, the product of the number computers in A and the number of WINE computers in V.

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, I also compute these measures for particular attack types. For example, I compute the exploits encountered per computer by only taking into account attack instances that have attack type “exploit”. Moreover, for web attacks and web attacks, we exclude a small number of attack instances where the attack type is “hacked web page”, and keep the majority of instances where the attack type is “malicious web page”. Different factors affect the likelihood that hacked and malicious pages appear, and our data about hacked web pages is insufficient to draw statistical conclusions.

3.4.3 Explanatory Variables

In this section, I present data I use to measure explanatory factors discussed in Sections 3.3.2 and 3.3.3. When applicable, I present more than one way to measure certain factors.

Web browsing. I use statistics about web browsing behavior in different countries that Canali et al. [20] extracted from data collected from a subset of Symantec customers who agree to share their web browsing histories with Symantec. More specifically, I use the average number of unique web pages visited (hits) and the average number of top level domains visits (tops) in a country.

Monetary resources. I measure people's wealth in different countries using the GDP per capita [26, 146].

Computing resources. Relevant computing resources mainly consist of Internet bandwidth and computer speed. I use the Internet bandwidth per Internet user indicator from the International Telecommunication Union (ITU) branch of the United Nations [67]. Measuring average computer speed in a country is difficult. As a proxy, I use the Information and Communication Technology (ICT) development index from the ITU [67].

Cyber-security expertise. It is difficult to directly measure cyber-security expertise of end-users, IT administrators and hackers in a country. As proxy, I use the strength of cyber-security scientific research, and the existence of cyber security policy and institutions. Expertise gained in cyber security research transfers over-time to cyber security practitioners and end-users. Moreover, cyber-security institutions like CERT work on increasing cyber-security expertise through awareness and training programs.

In order to measure the strength of cyber security research in a country, I use the number of cyber security research papers published by authors in the country. The number of scientific papers published is a standard bibliometric measure of expertise [105]. I collect from SCOPUS [123] all papers published during 2002-2011 in conferences and journals that contain "security" in their title and that belong to the engineering or computer science areas, obtaining a total of 28,400 papers. I start from 2002 as expertise gained in research requires time to transfer to industry and to the general public. I estimate cyber security research strength by counting the number of papers by each country. I consider that a paper is by a certain country if at least one of the paper authors has an affiliation in that country. The number of scientific papers published is a standard bibliometric measure of expertise [105]

I obtain the list of countries that have cyber-security institutions by combining data from multiple sources [28, 66, 86], and computing a binary variable that encodes whether countries have such institutions.

Corruption. The World Economic Forum collects a bribes indicator [156] by sending a questionnaire to a large number of business executives about how often firms 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.

Attack exposure. I use the attacks encountered per computer measure presented in Section 3.4.2.

International relations. International relations are more naturally represented as networks. I include international relations in the regression analysis of the number of attacks encountered and launched by using betweenness centrality in these networks.

I collect the list of international military and non-military conflicts during the period 1992-

2010 [24, 48]. I 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. Moreover, I collect the list of military alliances from the Correlates of War project [40], and use that list to compute a binary alliance network $A = [a_{ij}]$ that encodes the existence of military alliances between countries. Finally, I use the list of extradition treaties from the United Nations Crime and Justice Information network [151] in order to construct a binary extradition network $E = [e_{ij}]$.

I use data from year 2010 for the population size, the number of computers by 100 people, GDP per capita, ICT development index and bribes indicator. When data for such year is missing, I use data from 2009 or 2008 when available. Otherwise, I use the average indicator value among countries with similar income level [146] (high income, upper middle income, lower middle income and low income). For the population size, using data from a previous year is sufficient to find data for all countries in our analysis.

3.5 Threats to Validity

The IPS WINE telemetry data set are collected from 10 million customer computers worldwide and are very rich. However, such data also have limitations. First, the data cover attacks detected by the Symantec IPS, but do not cover sophisticated attacks that Symantec signatures and heuristics are unable to detect. Unfortunately, correcting for the above bias is difficult because there are no data about variation in the prevalence of sophisticated attacks [12, 147] across countries. Moreover, the data exclusively cover attacks that exhibit malicious network activity. Furthermore, the data are only about attacks on Symantec home customers. There is no reason to believe that findings from customers of other anti-virus vendors would be different, but findings from end-users that have no anti-virus vendors protection and from corporate end-users may be different because these end-users may have different computing behavior. Unfortunately, correcting for such bias is also difficult. Finally, the data are 5 years old. The cyber crime landscape may have changed since then especially with the rapid increase in smart phone penetration. I expect, however, mobile malware, to also target developed countries because these countries have more resources.

Another limitation is that I rely on Symantec's labeling of attacks. Researchers [6, 21, 104] have pointed out some inconsistencies between labels of different anti-virus vendors. However, I 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 [8, 65, 118].

It is worth noting that endogeneity is a potential risk for regression analysis because endogeneity can bias the regression coefficients. Endogeneity occurs when the error term in the regression is correlated with one or more regressors. Two phenomema can cause endogeneity: omitted variable bias and simultaneous causality. Omitted variable bias occurs when a variable that is determinant of the independent variable and that is correlated with one or more regressors is omitted from the regression. Simultaneous causality occurs when there is a causal link from the dependent variable to a regressor in addition to the causal link from the regressor to the dependent variable. In this work, there may be a simultaneous causality link between attack exposure and hosting and cyber security research and institutions. That is, countries may have invested in cyber security research and institutions as a response to high attack exposure or hosting. As a result, the coefficients on cyber security research and institutions may not be reliable. As future work, it would be interesting to address this simultaneous causality issue using instrumental variables or fixed point models.

My strategy to prevent omitted variable bias consists of including in the regression all variables that prior work said could be relevant ant that I could measure. However, one variable that could be relevant, but that I could not measure is the sophistication of the computing activities performed by users. This sophistication could increase attack exposure and is correlated with ICT, web browsing, and cyber security research and institutions. The fact that such sophistication is omitted from the regression likely inflates the regression coefficients on ICT, web browsing, and cyber security research and institutions.

Similarly, when counting attack instances, I rely on IP addresses to distinguish between attacker computers. In case many attacker computers are behind the same Network Address Translation (NAT) router, these attacker computers will appear to a victim computer outside the NAT router as a single computer. It is worth noting that for web attacks and fake applications, attacker computers are unlikely to be behind a NAT router. The attacker computers that deliver web attacks and fake applications are web servers that host malicious web pages as explained in the attack catalog paragraph in Section 3.4.1. It is unusual for multiple web servers to be behind a single NAT router because such web servers would have the same public IP address and the same port (80 for web traffic), and the NAT router would be unable to distinguish between traffic to the different web servers. On the other hand, for exploits, attacker computers are not necessarily web servers and it is possible that two attacker computers a_1 and a_2 are behind the same NAT router and have the same public IP address. In case both a_1 and a_2 attack the same victim computer v_1 using the same attack family f, the two attack instances (a_1, v, f) and (a_2, v, f) will be counted as a single attack instance because both a_1 and a_2 have the same public IP address. As future work, I intend to use port number analysis to de-alias NAT traffic. A threat report also contains the port number of the attacker computer, but I omitted discussing the port number in Section 3.4.1 in order to keep the discussion simple in that section.

A related issue is that I only account for malicious web sites that deliver web attacks and fake applications, and malicious computers that deliver exploits to victim computers. Other parts of the attack distribution infrastructure are not part of the study. It is worth noting, however, that Provos et al. [116] find that different parts of the attack distribution networks are highly localized within international boundaries in the context of drive-by-downloads (which are relevant for web attacks and to a less extent to fake applications). For example, 96$\%$ of hacked websites within China direct to malicious web sites within China [116].

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, I use, when possible, indicators from respected organizations that use well documented methodologies.

3.6 Attack Exposure

3.6.1 Descriptive Analysis

Tables 3.3 contain the list of countries that score highest on the number of attacks encountered per computer. From the figure, we see that the highest ranking countries on exploits are mainly

Table 3.3: Attack exposure. Top 10 countries on the number of attacks encountered per computer (log)

Country	# Exploits
Moldova	28.42
India	16.22
Taiwan	15.75
Nicaragua	13.02
Latvia	12.58
Italy	10.09
Israel	9.54
Uruguay	8.23
Bosnia & H.	6.86
Georgia	6.54

Country	# web attacks
Germany	1.64
S. Korea	1.64
US	1.29
UK	1.25
Netherlands	1.06
Canada	0.99
Australia	0.99
Russia	0.83
Belgium	0.81
Italy	0.79

Country	# fake apps
US	0.92
UK	0.83
Canada	0.76
Australia	0.68
Ireland	0.59
New Zealand	0.56
Norway	0.46
Switzerland	0.4
Belgium	0.38
Italy	0.79

Country	# all types
Moldova	28.7
India	16.56
Taiwan	15.91
Nicaragua	13.3
Latvia	13.05
Italy	11.13
Israel	10.1
Uruguay	8.41
Bosnia & H.	7.45
Georgia	7.07

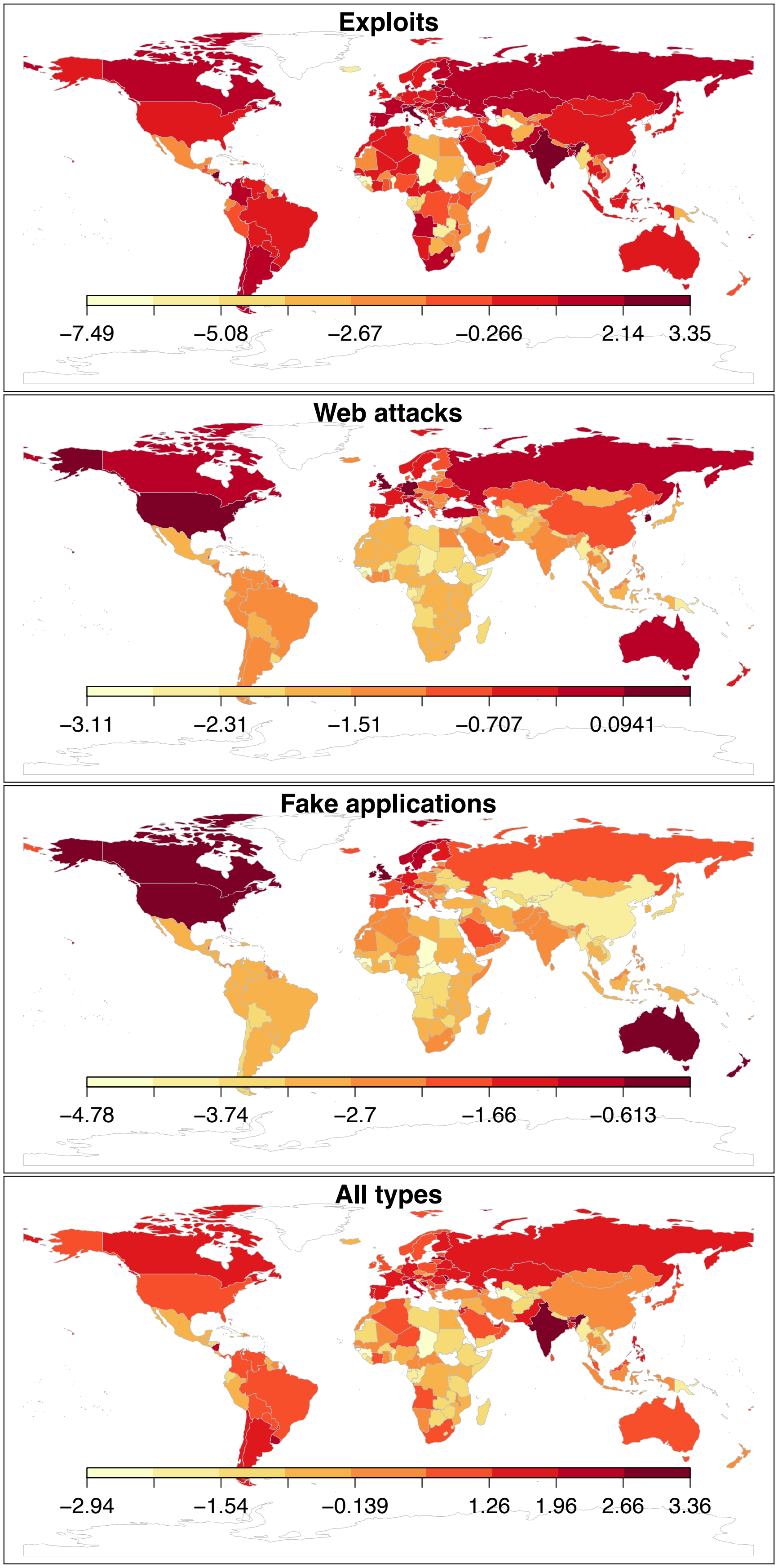
countries with emerging economies, whereas the highest ranking countries on web attacks and fake applications are mainly developed countries. Moreover, 6 countries appear among the top 10 countries on both web attacks and fake applications. These countries are the United States, the United Kingdom, Canada, Australia, Belgium and Italy. Finally, when taking into account all attack, I obtain a list that is very similar to the list of top countries for exploits because exploits constitute the majority of attacks in the data (see Figure 3.2).

Figure 3.4 contains a map visualization of the number of attacks encountered per computer. From the figure, it can be seen that no geographical region stands out as being much more exposed to exploits. On the other hand, African countries encounter less exploits than other countries in general. The figure also shows wide disparity in exposure to web attacks and fake applications between developed countries and developing countries. Finally, when taking into account all attack types, the pattern is similar to the pattern of exploits.

3.6.2 Explanatory Analysis

Table 3.4 presents summary statistics of variables used in the regression analysis of the number of attacks encountered per computer and Table 3.4 is the correlation table of explanatory variables I intend to use in the regression analysis. Table 3.4 reveals a very high correlation (0.92) between GDP PC and ICT which indicates that high income countries tend to have strong IT infrastructure. Because of this high correlation, I only keep ICT in the regression analysis.

Table 3.6 shows the regression analysis of the number of attacks encountered per computer. The regression coefficients are standardized, allowing for a comparison of the effect of different factors. According to the table, ICT has the highest impact with a coefficient ranging from 0.23



Exploits

Web attacks

Fake applications

All types

Figure 3.4: Attacks encountered per computer. Visualization (log scale).

Table 3.4: Attack exposure. Summary statistics of variables used in the regression.

Abbreviation	Variable	Mean	S.D.	Min	Max
Exploits enc.	Exploits enc. (log)	-1.05	2.63	-7.49	3.35
Web atks enc.	Web attacks enc. (log)	-1.43	0.63	-3.11	0.49
Fake apps enc.	Fake apps enc. (log)	-2.68	0.90	-4.78	-0.09
All types enc.	All types enc. (log)	0.10	1.25	-2.94	3.36
Bandwidth	Bandwidth	30	57.66	0.10	547
ICT	ICT index	3.82	2.05	0.85	8.45
GDP PC	GDP per capita (log)	8.49	1.53	5.29	11.55
Web hits	Web hits	1,020.52	586.86	105	5,363
Web tops	Web tops	12.89	4.11	1.00	32.33
Research	Security research	175.5	830.41	0	7911
Institutions	Security institutions	0.36	0.48	0	1
Alliance btw	Alliance btw	0	0.01	0	0.05
Hostility btw	Hostility btw	0	0.002	0	0.02
Extradition btw	Extradition btw	0	0.04	0	0.48

All variables are of length 184

for exploit exposure to 0.61 for web attack exposure. This indicates that an increase in ICT by 2.05 (the standard deviation of ICT according to Table 3.4) causes on average an increase in the logarithm of the number of exploits encountered by computer by 0.23 * 2.63 (where 2.63 is the standard deviation of the logarithm of the number of exploits encountered by computer according to Table 3.4) and an increase in the logarithm of the number of web attacks encountered by computer by 0.61 * 0.90 (where 0.90 is the standard deviation of the logarithm of the number of web attacks encountered by computer according to Table 3.4). From the table, it can also be seen that regression coefficients on cyber security research and institutions are also high, and surprisingly positive. This is surprising given that cyber security expertise is expected to reduce attack exposure. One possible explanation for this phenomenon is that expert users use their computers for longer time periods and for more sophisticated tasks, and are thus more likely to encounter attacks. For example, experts install more applications and are more likely to use the Tor network. Another possible explanation for this is that cyber security expertise by researchers and institutions does not necessarily transfer to home users. Furthermore, as expected, attack exposure increases with the diversity of web pages visited (as measured with web top, the number of top level domains visited). Finally, international relations do not have a significant impact on attack exposure. Even though state-sponsored attacks receive extensive media coverage, such attacks are still a minority compared with cyber-crime motivated attacks [147]. Moreover, state-sponsored attacks are typically very sophisticated and able to escape detection by security products.

3.7 Attack Hosting

3.7.1 Descriptive Analysis

In this section, I examine international variation in the number of attacks launched per computer, my measure of the extent to which cyber criminals host their attacking infrastructure in these countries. Tables 3.7 contain the list of countries that score highest on this measure. From the tables, it can be seen that many Eastern European countries host very large quantities of attacks. Other countries that host large quantities of attacks are Congo, Luxembourg, Belize, Dominica, and Trinidad and Tobago. The tables also show a large overlap in the lists of top countries across different attack types. Moldova, Bosnia & Herzegovina, Ukraine, Latvia and Romania appear in all 4 lists.

Table 3.5: Attack exposure. Correlation table of explanatory variables

	Bandwidth	ICT	GDP PC	Web hits	Web tops	Research	Institutions	Alliance btw	Hostility btw
Bandwidth									
ICT	0.60***								
GDP PC	0.52***	0.92***							
Web hits	0.13	0.25***	0.24**						
Web tops	-0.04	0.02	0.00	0.65***					
Research	0.09	0.23**	0.20**	0.10	-0.10				
Institutions	0.34***	0.56***	0.47***	0.14	-0.03	0.27***			
Alliance btw	0.10	0.21**	0.18*	0.06	-0.06	0.46***	0.20**		
Hostility btw	0.03	0.02	0.00	0.00	-0.01	0.38***	0.07	0.30***	
Extradition btw	0.03	0.13	0.12	0.06	-0.05	0.70***	0.12	0.58***	0.55***

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

Moreover, Belize, Dominica and Trinidad & Tobago appear in 3 lists.

It is surprising that China is not among the top attack hosting countries given China's bad cyber security reputation. For example, Provos et al. [116] find that China hosts the largest number of web servers that deliver drive-by-downloads. However, given that China has the second largest number of computers in the world[3], the fact that China has the largest number of malicious web servers does not indicate that the computing environment in China is favorable to hosting attacks. Another reason why China has a bad cyber security reputation is the Chinese cyber espionage activities. More specifically, Chinese hackers, some of whom are affiliated with the Chinese government, are often suspected of launching sophisticated attacks on Western governments and industries to steal intellectual property. In this work, I examine the extent to which a country hosts cyber criminal computing infrastructure, and not to which a country hosts hackers.

Figure 3.5 presents a map visualization of the number of attacks launched per computer. From the figure, it can be seen that Eastern Europe tends to host the highest quantities of exploits, while Africa tends to host the smallest quantities of exploits. Other regions are somewhere in between Africa and Eastern Europe. When examining web attack and fake application hosting, the figures also show that Eastern European countries also host the highest quantities of these attacks. Other countries that host large quantities of web attacks and fake applications tend to be in Western Europe and North America. The figures also indicate that Africa hosts very small quantities of web attacks and fake applications. This is surprising given the scam reputation of African countries such as Nigeria and Ghana [17, 134]. The reason behind this surprising finding is that in this paper we study a different phenomenon than scams. Scams such as the 419 email scam involve people sending scam messages through email, social media or dating websites. In this paper we examine the likelihood that *computers* in a country are used to host and deliver web attacks, fake applications, and exploits. We do not examine the likelihood that *people* in a certain country send crafted scam messages.

3.7.2 Explanatory Analysis

Table 3.4 presents summary statistics of the variables to be used in the regression analysis of the number of attacks launched per computer and Table 3.5 is the correlation table between the ex-

[3] I estimate the total number of computers in different countries based on estimates of the number of computers by 100 people [145] and estimates of the population sizes [146].

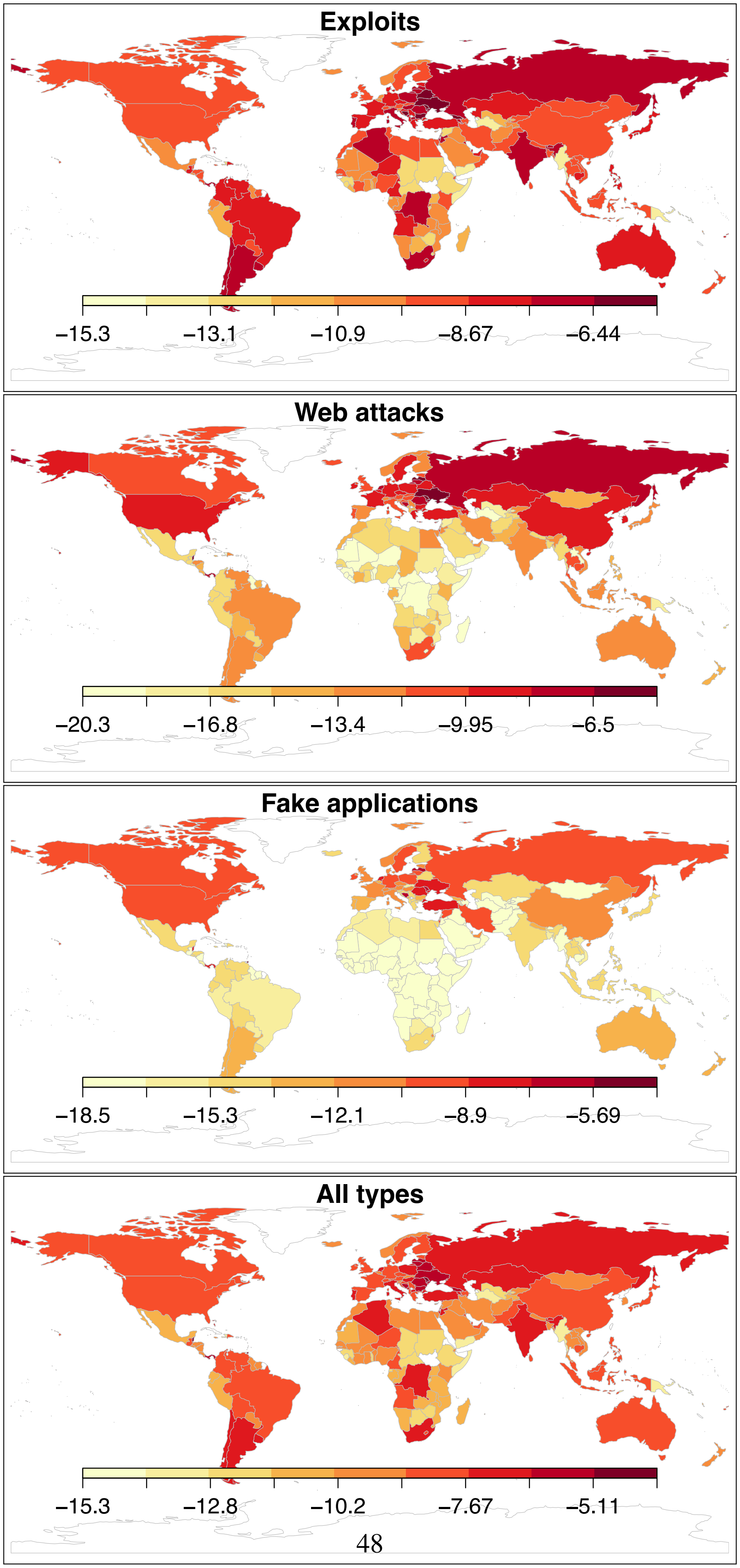


Figure 3.5: Attack hosting. Visualization of the number of attacks launched per computer (log scale).

planatory variables to be used in the regression analysis.

Table 3.10 presents the results of the regression analysis on the number of attacks launched per computer, my measure of attack hosting. The regression coefficients are standarized, allowing for a comparison of the effect of different factors. For example, when considering exposure to exploits, the fact that ICT's regression coefficient (1.2) is almost the double of the all attack types enc's regression coefficient (0.69) indicates that ICT has about double the effect of all attack types enc on exploit exposure.

Table 3.10 reveals that ICT and ICT x bribes are the factors with the strongest effect. The strong effect of ICT indicates that attackers prefer hosting attacks in countries with good computing infrastructure because these countries have fast computers with good Internet connection available at a cheap price. These cheap prices allow cyber criminals to maximize their profit margins. The strong effect of ICT x bribes indicates that a combination of widespread corruption and a good computing infrastructure is very favorable to hosting attack infrastructure[4]. Widespread corruption facilitates using computing infrastructure towards criminal ends. Even if such infrastructure is used to aggressively launch attacks, corrupt law officials and ISPs are likely to turn a blind eye on the situation. It is worth noting that, according to the table, corruption alone does not increase attack hosting. In the absence of a good computing infrastructure, the ability to conduct crime with impunity is not attractive to cyber criminals. As expected, attack exposure increases attack hosting. As computers encounter more attacks, these computers are more likely to become compromised and start launching attacks. Surprisingly, statistically significant regression coefficients on cyber security institutions and research are positive. This indicates that countries that have cyber security expertise are more likely to be used to host attacks. One possible explanation for this observation is that a certain number of hackers host their cyber attack infrastructure in their countries of residence because these hackers do not see the need to operate in foreign countries given that they can operate in their own countries where they have connections. Finally, international relations do not have a significant effect on attack hosting.

3.8 Attack Network

3.8.1 Descriptive Analysis

Tables 3.11 contain a regional aggregation of the cyber-attack networks. Table 3.11a reveals that exploits have a high tendency to propagate to geographically nearby countries. The table also reveals that E. Eur launches a large number of exploits on all regions, and that E. Eur and W. Eur.+ encounter a large number of exploits from all regions. From Tables 3.11b and 3.11c, it can be seen that E. Eur and W. Eur.+ launch large quantities of web attacks and fake applications primarily on W. Eur.+, but also on other regions. There is also a small tendency for intra-region propagation of web attacks and fake applications. Finally, the regional network of all attack types given in Table 3.11d is very similar to the regional network of exploits because exploits are the majority of attacks in the data (viz. Figure 3.2).

I now examine visualizations of the strongest edges in the attack networks in Figure 3.6. Such visualizations are less complex than visualizations of the entire attack networks. In order to be able

[4] The coefficient on ICT x bribes is negative because the bribes index is such that countries with widespread corruption have a small bribes index

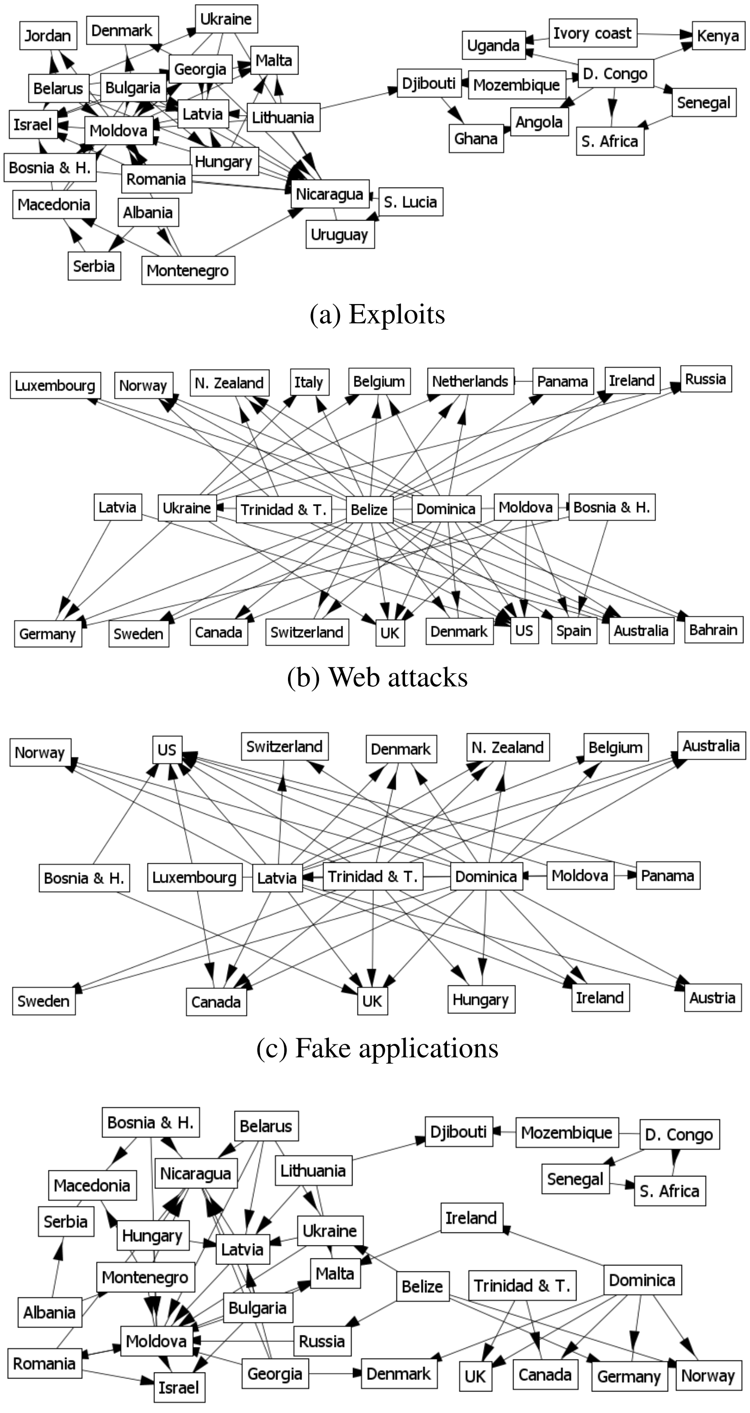
to meaningfully compare the 4 networks, we ensure that all visualized networks have density 0.003 where density is the ratio of edges present in the network to all possible edges in the network. In Figure 3.6a, one can see an Eastern European cluster, an African cluster and a small Latin American cluster confirming our previous observation that exploits tend to propagate to geographically nearby countries. Figures 3.6b and 3.6c show that the majority of edges are from Eastern European and Latin American countries to rich countries primarily in Western Europe. Finally, when taking into account all attack types in Figure 3.6d, we see patterns from the 3 previous networks.

3.8.2 Explanatory Analysis

Table 3.12 presents network level measures of the networks used in the explanatory analysis of the cyber attack networks. I only use attributes of victim countries and attacker countries that were significant in Sections 3.6.2 and 3.7.2 respectively[5]. A component is a maximal set of nodes that are connected. In the table, I 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 of the number of existing edges to the number of all possible edges in as explained previously. The clustering coefficient measures the extent to which a node's neighbors are themselves neighbors. From Table 3.12, one can 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 3.2. The attribute networks ICT att, bribes att, institutions att, ICT vic, institutions vic have 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 3.13 contains the results of the MrQAP regression on the cyber attack networks. One can see that attributes of attackers and victims have a similar effect to the effect discussed in Sections 3.6.2 and 3.7.2. For example, resources have a positive impact on the number of attacks encountered and launched. Similarly, a combination of good computing resources and high levels of corruption increases the number of attacks launched. From the table, one can also see that attacks tend to propagate to geographically nearby regions, which is consistent with Tables 3.11. 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.

[5] I compute the attribute network for the attacker by repeating that attribute column. For example, assume I 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, I compute the attribute network for the victim by repeating the attribute row. The ICT victim network is then $\begin{bmatrix} 3 & 8 & 6 \\ 3 & 8 & 6 \\ 3 & 8 & 6 \end{bmatrix}$.



(a) Exploits

(b) Web attacks

(c) Fake applications

(d) All types

Figure 3.6: Visualization of the strongest edges in the cyber attack networks.

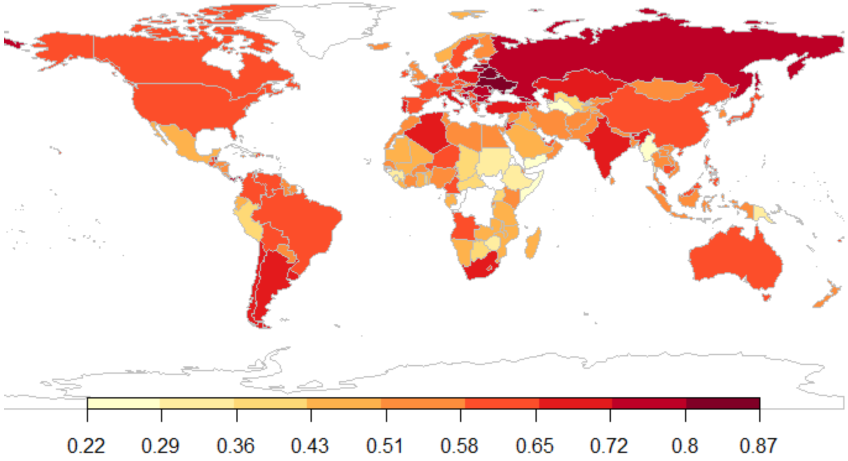


Figure 3.7: Prediction of attack hosting in different countries.

3.8.3 Predictive Analysis

In this section, I am interested in predicting countries' attack hosting in the near future based on the Friedkin model and the total attack network.

The Friedkin model is a recursive linear model that predicts change in attitudes of actors (such as people or countries) over time. The Friedkin equation model is $y^{(t)} = AWy^{(t-1)} + (I - A)y^{(1)}, t = 2, 3, ..,$ where $y^{(t)} = [y_i^{(t)}]$ is a $N * 1$ vector of actors' attitudes at time t and $A = diag(a_{11}, .., a_{ii}, .., a_{NN}), 0 \leq a_{ii} \leq 1$ is a $N * N$ diagonal matrix of actors' susceptibility to external influence. $W = [w_{ij}], (0 \leq w_{ij} \leq 1, \sum_{j=1}^{N} w_{ij} = 1)$ is a $N * N$ matrix of inter-actor influence. More specifically, w_{ij} is j's influence on i. The general formulation of the Friedkin model stipulates that $W = AC + I - A$ where $C = [c_{ij}], (c_{ii} = 0, 0 \leq c_{ij} \leq 1, \sum_{j=1}^{N} c_{ij} = 1)$ is a $N * N$ matrix of relative inter-actor influence and I is the identify matrix. Finally, $y^{(1)} = [y_i^{(1)}], 0 \leq y_i^{(1)} \leq 1$ is a $N * 1$ vector of the actors' initial attitudes.

I set $A = [a_{ii}]$ based on countries' GDP because countries' susceptibility to international influence is known to decrease with their GDP [56, 93]. More specifically, I set $a_{ii} = 0.4 + 0.6b_{ii}$ where $b_{ii} = 1 - log(GDP_i)/log(max(GDP))$. b_{ii} is such that $0 \leq b_{ii} \leq 1$ and b_{ii} is smaller for countries with large GDP. I set C as the all types attack network after dividing each row by the sum of that row. I then set W as $W = AC + I - A$. Finally, I set y^1 as the vector of the number of all attack types launched by computer in each country after dividing that vector by the maximum value of that vector.

I run the Friedkin model until equilibrium obtaining the scores in Figure 3.7. According to the Figure, Eastern Europe and Latin America will continue to be the main region hosting attacks. Other countries predicted to host considerate quantities of attacks are Luxembourg, Democratic Republic of Congo, Taiwan, Israel, and India.

3.9 Future Work

As future work, I intend to explore other measures such as the median of the number of attacks, and the percentage of computers that encounter or launch attacks in a country. I also intend to perform a longitudinal analysis. Another future work direction is to improve the regression analysis. For example, it would be interesting to collect and include data about other potential explanatory factors such as the categories of visited web sites and the number of applications installed. It would also be interesting to use techniques such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) to select regression models. Moreover, it would be interesting to perform lower granularity analyses such as studying attack exposure and hosting at the Internet Service Provide level. Finally, case studies about cyber security practices in countries that excessively host attacks in order to gain an in-depth understanding of problematic practices in these countries.

3.10 Conclusion

In this chapter, I empirically investigate how attack exposure and hosting vary across countries and statistically test the validity of expert opinions about factors behind such international variation. Such analysis allows policy actions aiming at reducing international attack encounters and hosting to rely on scientific empirical evidence instead of simply relying on expert opinions.

I use the Symantec IPS telemetry data collected from more than 10 million Symantec customer computers worldwide, and test for the relevance of factors such as computing and monetary resources, cyber security expertise, corruption, and international relations. I find that abundant computing and monetary resources are the main reason why some countries encounter more attacks. Criminals are interested in taking advantage of these resources in order to generate large monetary profits. This finding is an unfortunate side effect of the global nature of the Internet where criminals can target rich people worldwide, unconstrained by international boundaries and geographical distance. I recommend that developed countries make higher investments in cyber security defenses in order to increase the cost of successfully attacking them. As this cost increases, the profitability of attacking such countries will decrease and attackers' interest may shift to other places.

Moreover, my analysis reveals that many Eastern European countries (such as Moldova, Bosnia & Herzegovina, Ukraine, Latvia and Romania) host disproportionate quantities of attacks. These countries are attractive for hosting these attacks because of a combination of reasonable computing resources and widespread corruption. Widespread corruption facilitates conducting cyber-criminal activities through the complicity of law officials and ISPs. In order to illustrate how corruption facilitates cyber crime, it is worth mentioning the Russian Business Network (RBN), a malicious ISP that hosted a major portion of cyber crime worldwide [14]. Part of the reason why the RBN was able to carry out cyber criminal operations of that magnitude was that the RBN's owner and operator was the nephew of a prominent Russian politician [120]. After the RBN was exposed and made headlines in late 2007, the RBN ceased its operations in St. Petersburg, but quickly resumed its operations in other places [137].

Implications of the finding that a combination of good computing infrastructure and widespread facilitates cyber crime in Eastern Europe is that traditional approaches to cyber security are likely to be ineffective in Eastern Europe. For example, designing more secure systems and providing cyber security training are likely to be of little help. These countries have the technical ability to reduce the likelihood that their systems are used to host attacks, but corrupt ISPs and law officials in these countries turn a blind eye to the problem. Similarly, the legislative route of increasing cyber crime penalties suggested in the United States [106] and the European Union [34] are also unlikely to be effective in Eastern Europe. Corruption greatly reduces the effectiveness of penalties because corruption reduces the likelihood that criminals are caught and prosecuted [9, 29, 80, 94]. Even worst, when corruption is widespread increasing penalities may result in higher crime rate [9, 29, 80, 94]. When fines are increased, corrupted law officials can obtain higher bribes from criminals and therefore have incentive to pamper crime [94]. Moreover, criminals have incentive to extend corruption rings which causes a decrease in the expected penalty of conducting crime and thus an increase in crime rate [80]. *When corruption is widespread, effectively reducing crime requires first identifying and cracking down on corrupt individuals and institutions that support crime* [9, 29, 80, 94]. For example, when the RBN was exposed in 2007, local authorities should have arrested the individuals that run these networks *and* law officials that permitted such activities. Local authorities in these countries may perceive little incentive to act because attacks they host affect users elsewhere. However, through the use of soft power [108], these countries should realize that they have incentive to collaborate because their practices are likely to have a deplorable effect on their local populations in the long term. Entire blocs of IP addresses from these countries may become blocked and honest users may become disconnected from parts of the Internet. This has happened to West African users because of the extensive scam activities originating from this part of the world [17]. While waiting for policy actions to take effect, I suggest that intrusion prevention systems should take into account the country where network traffic appears to be from when deciding whether the traffic is malicious. Network traffic from countries that host disproportionate quantities should be treated as more suspicious, but not necessarily always malicious.

I also find that many countries in Western Europe and North American host large quantities of attacks. Cyber criminals are interested in hosting attacks in these countries because, as long as such hosting remains undetected, these criminals can enjoy a high quality service at a cheap price, thus maximizing their profit margins. I suggest that honest hosting services that offer internationally competitive services and pricing should be extremely careful about the fact that their platforms could be used to host attacks.

Table 3.6: Attack exposure. Regression analysis. Regression coefficients are standardized.

	Exploits enc.	Web atks enc.	Fake apps enc.	All types enc.
Computing & Monetary resources				
Bandwidth	-0.074	-0.033	0.085	-0.060
	(0.083)	(0.064)	(0.070)	(0.077)
ICT	0.23^{**}	0.61^{***}	0.60^{***}	0.40^{***}
	(0.097)	(0.074)	(0.082)	(0.090)
Web browsing				
Web hits	-0.19^{**}	-0.10	-0.071	-0.19^{**}
	(0.091)	(0.070)	(0.077)	(0.085)
Web top	0.24^{***}	0.13^{*}	0.003	0.20^{**}
	(0.090)	(0.069)	(0.076)	(0.083)
Cyber security research and institutions				
Research	0.038	0.15^{**}	-0.074	0.036
	(0.096)	(0.074)	(0.081)	(0.089)
Institutions	0.36^{***}	0.13^{**}	-0.030	0.31^{***}
	(0.081)	(0.062)	(0.069)	(0.075)
International relations				
Alliance btw	0.024	0.15^{**}	0.15^{**}	0.038
	(0.082)	(0.063)	(0.069)	(0.076)
Hostility btw	-0.059	0.023	0.033	-0.064
	(0.079)	(0.061)	(0.067)	(0.073)
Extradition btw	-0.016	-0.073	0.097	-0.001
	(0.11)	(0.084)	(0.093)	(0.10)
F-Statistics testing coefficients (p-value)				
Resources	0.21	< 0.001	< 0.001	< 0.001
Web browsing	0.08	0.18	0.43	0.23
Research & institutions	< 0.001	0.004	0.56	< 0.001
International relations	0.60	0.17	0.007	0.16
N	184	184	184	184
R^2	0.25	0.56	0.47	0.36

Standard errors in parentheses
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 3.7: Attacks hosting. Top countries on the number of attacks launched per computer (log)

Country	**# Exploits**
Belarus	4.85
Moldova	2.88
Georgia	2.68
Bulgaria	2.52
Bosnia & H.	1.96
Ukraine	1.96
Latvia	1.56
Congo	1.49
Hungary	1.42
Romania	1.39

Country	**# Web attacks**
Belize	8.41
Dominica	3.76
Moldova	2.97
Ukraine	1.99
Latvia	1.57
Trinidad & T.	1.33
Lithuania	0.92
Bosnia & H.	0.75
Romania	0.74
Russia	0.52

Country	**# Fake apps**
Dominica	13.82
Trinidad & T.	5.27
Latvia	2.31
Bosnia & H.	1.01
Moldova	0.81
Luxembourg	0.76
Panama	0.67
Belize	0.55
Romania	0.50
Ukraine	0.35

Country	**# all types**
Dominica	21.66
Belize	9.08
Trinidad & T.	7.29
Moldova	6.75
Latvia	5.63
Belarus	4.91
Ukraine	4.06
Bosnia & H.	3.77
Georgia	2.72
Romania	2.67

Table 3.8: Attack hosting. Summary statistics of variables used in the regression.

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

All variables are of length 184

Table 3.9: Attack hosting. Correlation Table of explanatory variables used in the regression

	Bandwidth	ICT	Bribes	Research	Institutions	All types enc.	Alliance	Hostility
Bandwidth								
ICT	0.60***							
Bribes	0.57***	0.79***						
Research	0.09	0.23**	0.17*					
Institutions	0.34***	0.56***	0.39***	0.27***				
All types enc.	0.26***	0.51***	0.32***	0.16*	0.50***			
Alliance	0.10	0.21**	0.11	0.46***	0.20**	0.15*		
Hostility	0.03	0.02	-0.02	0.38***	0.07	-0.01	0.30***	
Extradition	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 3.10: Attack hosting. Regression analysis. Regression coefficients are standarized.

	Exploits lau.	Web atks lau.	Fake apps lau.	All types lau.
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				
BWxBribes	0.17	-0.57	-0.89	0.063
	(0.36)	(0.42)	(0.54)	(0.36)
ICTxBribes	-1.48***	-0.93*	-0.37	-1.35***
	(0.30)	(0.37)	(0.39)	(0.35)
Cyber security research & 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)
Attack exposure				
All attack types enc.	0.69***	0.28***	0.23***	0.61***
	(0.060)	(0.054)	(0.063)	(0.058)
International relations				
Alliance btw	-0.031	0.0054	0.042	-0.037*
	(0.016)	(0.023)	(0.032)	(0.015)
Hostility btw	-0.029	-0.014	-0.059	-0.045
	(0.035)	(0.074)	(0.035)	(0.034)
Extradition btw	0.0048	-0.046	0.0073	0.032
	(0.033)	(0.043)	(0.030)	(0.033)
F-Statistics testing coefficients (p-value)				
Resources	<.001	<.001	<.001	<.001
Resources x Corruption	<.001	.002	.034	<.001
Expertise	.151	<.001	.001	.612
International relations	.21	.107	.19	.069
N	184	184	184	184
R^2	0.65	0.57	0.49	0.61

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

Table 3.11: Regional aggregation of cyber attack networks. The edge weight from region A to region B represents the average number (log scale) of attacks from a computer in A on a WINE computer in B

(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

Color key

-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

Color key

-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

Color key

-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

Table 3.12: Cyber attack networks. Network Level Measures of Networks used in the MrQAP regression

Abbreviation	Network	Size	Components	Isolates	Density	Cluster. coef.
			(2+ nodes)	(1 node)		
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
Bribes att	Bribes attacker	184	1	0	1	1
Institutions att	Institutions attacker	184	1	0	0.36	0.77
ICT vic	ICT victim	184	1	0	1	1
Institutions vic	Institutions vic	184	1	0	0.36	0.77
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 3.13: MrQAP regression on cyber attack networks. Coefficients are standardized.

	Exploit	Web atks	Fake apps	All types
Attributes of attackers and victims				
ICT att	0.63***	0.57 ***	0.43*	0.05
Bribes att	0.13	0.18***	0.01	-0.10
ICT x Bribes att	-0.65***	-0.69***	-0.34	0.11
Institution att	0.21***	0.15***	0.17***	0.10
ICT vic	0.191***	0.19***	0.07*	0.03
Institution vic	0.23***	0.26***	0.05*	0.02
Interaction between attributes of attackers and victims				
ICT att x ICT vic	0.22***	0.18***	0.20***	0.29***
ICT att x Institution vic	0.05***	0.12***	0.16***	0.13***
Bribes att x ICT vic	-0.19***	-0.18***	-0.14***	-0.16***
Bribes att x Institution vic	-0.04	-0.10***	-0.11***	-0.07***
Institution att x ICT vic	-0.004	0.02	0.09***	0.05***
Institution att x Institution vic	-0.01	0.02***	0.05***	0.03***
Geographical proximity				
Regional	0.06***	0.07***	-0.01	-0.04***
International relations				
Hostility	0.03***	0.03***	0.03***	0.02
Extradition	0.04*	0.04*	0.11***	0.11***
Alliance	0.07***	0.06***	0.05***	0.05***
N	184	184	184	184
R^2	0.32	0.29	0.34	0.23

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

Chapter 4 Cyber Weapon Capabilities

Research questions: How to systematically identify countries that could pose a cyber weapon threat?

4.1 Introduction

Recent years have witnessed many state-sponsored cyber-attacks. For example, Stuxnet is a worm that the American and Israeli governments allegedly launched on the Iranian nuclear program in 2010 in order to halt that program [22, 122]. In 2007, Russia launched massive Denial of Service attacks that crippled websites of banks, the parliament, ministers, and news outlets [22] in Estonia. The Estonian attack followed the relocation of the "Bronze Soldier of Tallinn" statue, a symbol of the Soviet era. State-sponsored attacks are usually much more sophisticated than common malware [30, 84] and typically have political motivations instead of monetary motivations. Besides countries suspected of launching state-sponsored attacks, many countries have formally announced intent to develop cyber-warfare capabilities and included cyber warfare troops in their military institutions [87].

In a world where state-sponsored attacks are common and where many countries develop cyber warfare capabilities, it is important to assess countries' cyber warfare capabilities. Such assessment can help a country know where it stands compared to other countries. Such assessment can also help narrow down the list of actors capable of launching highly sophisticated attacks and thus help towards attributing future sophisticated attacks. Despite the importance of assessing countries' cyber warfare capabilities, such assessment is currently mainly done through case studies [16, 58, 63, 92]. Case studies provide a detailed picture of capabilities of one or a few countries. However, case studies require substantial country-specific expertise and effort. As a result, case studies typically focus on a small number of "obvious" countries and may thus miss less obvious countries that may develop capabilities "off-the-radar".

In this chapter, I develop a computational methodology [53] that assesses countries' cyber warfare capabilities. My methodology examines all countries in the world and can be used by non experts. I first identify factors that motivate countries to develop cyber warfare capabilities by statistically testing alternative hypotheses against historical data. Then, I incorporate the motivational factors identified by that hypothesis testing into the Friedkin socio-cultural model [55], an iterative model developed to capture people's change in opinion, but also used to predict many of the European Union decisions [46]. I populate the model using publicly available data about international relations and the list of countries that have included cyber warfare troops in their military. I run the model until equilibrium, obtaining an assessment of countries' motivations for these capabilities. I obtain an assessment of countries' cyber warfare latent abilities by examining countries' cyber security research, cyber security institutions, and general IT preparedness. I find that countries with the strongest cyber warfare capabilities are the United States, China, Israel, Russia, South Korea, Russia, Germany, France, the United Kingdom, and Australia. Other countries with strong capabilities are Canada, Italy, Denmark, Netherlands, Norway, Poland, and Japan.

It is worth noting that I am *not* interested in predicting which countries will engage in cyber war or how countries will use these capabilities. Countries may use such capabilities to cripple other countries' IT infrastructure, to access secret information about their opponents, or in a joint cyber physical attack. Countries will probably decide how to use such capabilities depending on their political goals and on who their political opponents are.

The remainder of this chapter is organized as follows. I provide background in Section 4.2 and test hypotheses about factors that motivate countries to develop cyber warfare capabilities in Section 4.3. I assess countries' motivations for cyber warfare capabilities in Section 4.4 and assess countries' latent abilities in Section 4.5. I present the results of the joint motivation and latent abilities assessment in Section 4.6. I discuss limitations and future work in Section 4.7 before concluding in Section 4.8.

4.2 Background

Cyber attacks vary greatly in sophistication. The simplest attacks require simply downloading attack tools from the Internet and using them. More sophisticated ones require building new malware, distributing malware, and managing infected machines. The most sophisticated attacks such as Stuxnet typically involve zero-day attacks and advanced knowledge of enemies' systems. In this work, I am interested in assessing countries capabilities to launch very sophisticated attacks.

4.2.1 Motivational Factors

The main factor discussed in the literature about why countries develop cyber warfare capabilities is the concern about the rise, sophistication, and impact of cyber attacks [32]. The literature also contains extensive discussion about the fact that cyber deterrence via retaliation is difficult [51, 89, 109] because of the attribution problem. Beyond these two discussions, I could not find in the literature a list of factors that motivate countries to develop cyber warfare capabilities. For this reason, I will present proliferation hypotheses from the nuclear proliferation theory and explain how those hypotheses may apply to cyber proliferation theory. I do *not* assume that these hypotheses apply to proliferation of cyber warfare capabilities. In Section 4.3, I will statistically test whether these hypotheses hold in the context of cyber warfare capabilities. I choose to draw hypotheses from the nuclear proliferation theory because nuclear proliferation theory is often used as a starting point to think about cyber warfare [33, 47, 89, 109, 113, 124]. For example, Nye [109] explains that despite major differences between cyber attacks and nuclear weapons, nuclear lessons can be used by governments and private actors to comprehend and address cyber space. Below, I give hypotheses about factors that motivate countries to develop cyber warfare capabilities. I will statistically test these hypotheses in Section 4.3.

In-kind deterrence. A country that has a nuclear enemy has a great incentive to acquire nuclear weapons in order to fill the security deficit [121, 155]. The hypothesis I derive in the context of cyber warfare capabilities is that a country that has cyber warfare capabilities may have an incentive to acquire cyber warfare capabilities in order to fill the security deficit.

Ally reassurance. A country is desicentivized from acquiring nuclear weapons if the country has a nuclear ally that promises retaliation in case the country is attacked [11, 44, 143]. A similar hypothesis in the context of cyber warfare capabilities is having an cyber weapon ally is a disincentive against developing cyber weapons.

Ally assistance. There are multiple instances where countries provided nuclear assistance to their allies so that these allies could develop their own nuclear weapons [79]. The hypothesis I derive in the context of cyber warfare capabilities is that having an ally with cyber warfare capabilities increases the likelihood that a country seeks cyber warfare capabilities.

Latent abilities. Countries that have nuclear latent ability such as nuclear engineers and Uranium deposits are more likely to develop nuclear weapons [69]. A similar hypothesis for cyber warfare is that a country that has cyber latent ability is more likely to develop cyber warfare capabilities.

High potential losses to cyber attacks. This hypothesis says that countries that have more to lose to cyber attacks may be more interested in developing cyber warfare capabilities. Countries may have more to lose to cyber attacks if these countries produce significant intellectual property or if these countries have substantial cyber critical infrastructure. This hypothesis is not derived from nuclear proliferation theory, but from the observation that countries often list concerns about intellectual property theft and worries about critical infrastructure when motivating the need to develop cyber warfare capabilities [32].

4.2.2 Latent Ability

Contrary to nuclear weapons, building cyber warfare capability does not require raw material such as Uranium deposits or major infrastructure such as nuclear centrifuges. Building cyber warfare capability mainly requires a team of highly skilled cyber security experts. A country can identify cyber security experts among cyber security researchers, cyber security practitioners or hackers. Such experts will probably receive additional training.

Cyber security researchers usually publish scientific papers in conferences and journals, unless these researchers work on classified projects in industry or within the government. Cyber security practitioners may, for example, work within cyber security organizations such as Computer Emergency Readiness Teams (CERTs) or work as security officers within large corporations. Unfortunately, hackers typically work in the underground world and their country of residence or origin is difficult to track in publicly-available data. It is worth noting, however, that hackers are more likely to be found in countries with high general computing expertise.

4.3 Hypothesis Testing about Motivational Factors

In this section, I test hypotheses presented in Section 4.2.1. I use survival analysis, an statistical analysis that handles over-time data. Survival analysis models the time duration until an event occurs. For example, survival analysis can be used to model the number of years until patients die after getting a certain disease. Survival analysis appropriately captures the case where observations are censored. The observation about a certain patient is censored if, for example, the experiment ends while the patient is still alive.

Survival analysis was initially developed to model how long individuals survive, but is often used in other contexts such as modeling the proliferation of weapons of mass destruction [64]. In this work, I use survival analysis to model the time until countries start developing cyber warfare capabilities. In other words, the dependent variable is the time until a country starts developing cyber warfare capabilities, and the independent variables are whether the country has an enemy that has such capability, whether the country has an ally that has such capability, the country's cyber capability, and the country's intellectual property.

In testing hypotheses about factors behind cyber weapon proliferation, I focus on the time period after year 2000. This is because international attitudes towards cyber warfare have dramatically evolved over time. In the 1990s, cyber warfare was a theoretical concept [125]. The situation changed in the 2000s, especially after state sponsored attacks such as the Estonian attack in 2007, the Georgia attack in 2008, and Stuxnet in 2010 [125]. Since, in this work, I am interested in identifying factors that *currently* motivate countries to motivate to seek cyber warfare capabilities, I will examine the time period starting from 2000 when testing hypotheses. In order to examine the robustness of my finding, I will also examine the results when using 2005 and 2007 as starting points.

I present the data I use to test hypotheses in Section 4.3.1 and the results of my hypothesis testing in Section 4.3.2.

4.3.1 Data

Proliferation Timeline. Table 4.1 contains the year when countries started developing cyber warfare capabilities. I assume that a country starts developing cyber warfare capabilities when the country includes cyber warfare troops in its military or announces plans and budget to build cyber warfare capabilities. In this work, I mainly focus on the ability to develop and defend against malicious software. I do not consider the ability to launch information warfare i.e. propaganda as cyber warfare capability. Similarly, I do not consider the ability to develop or crack cryptographic algorithms as cyber warfare capability. For this reason, I do not consider that France and the United Kingdom worked on developing cyber warfare capabilities when they were working on cryptographic algorithms during World War II. Finally, it is worth noting that the timeline in Table 4.1 may contain errors because that timeline is only collected from publicly-available sources.

International Alliances. I use the alliance data from the Correlates of War project [36, 57, 131, 132]. The alliance data include defense pact treaties, entente agreements, non-aggression treaties, and neutrality pacts. Defense pact treaties commit countries to intervene military in case of an attack on any treaty member. Entente agreements pledge consultation in case of a crisis. In a non-aggression and neutrality treaties, countries commit to remaining neutral in case a treaty cosignatory is attacked. For each treaty, the data contain the lists of countries that signed the treaty and the years when the treaty was in force.

International Hostilities. I use the list of inter-state military disputes from the International Crisis Behavior (ICB) project [25]. For each dispute, the data contain the list of countries involved and the years when the dispute was active.

Latent Abilities. I use the number of Internet users per 100 people [67] as a measure of countries' cyber latent abilities. Such measure has been collected since early days of the Internet and is thus useful when testing hypotheses against historical data.

Table 4.1: Cyber Warfare Proliferation Timeline

Country	**Year**
United States	1990s [74]
Russia	1990s [13]
Israel	1990s [119]
China	1990s [13, 63]
N. Korea	1990s [13]
India	1990s [13, 87]
Iran	1990s [13]
Greece	1990s [32]
S. Korea	2001 [13]
Germany	2006 [88]
Malaysia	2007 [87]
Argentina	2008 [112]
Estonia	2008 [22, 32, 87, 119]
Poland	2008 [87]
Australia	2009 [22, 87, 119]
Denmark	2009 [87]
Austria	2009 [87]
Colombia	2009 [87]
France	2009 [22, 32, 87]
Burma	2009 [87]
United Kingdom	2009 [87]
Canada	2010 [22, 119]
Brazil	2010 [22]
Georgia	2010 [87]
Italy	2010 [119]
Norway	2010 [87]
Albania	2010 [87]
Turkey	2010 [32, 87]
Switzerland	2010 [87]
Netherlands	2011 [87]
Belarus	2011 [87]
Finland	2011 [32, 87]
Kazakhstan	2011 [87]
Spain	2014 [32]
Croatia	2004 [32]
Hungary	2014 [32]
Slovakia	2014 [32]

Table 4.2: Results of hypothesis testing about factors that motivate countries to develop cyber warfare capabilities

Independent variable	**2000 start**	**2005 start**	**2007 start**
Enemy developing cyber warfare capability	0.70	0.28	0.32
Ally developing cyber warfare capability	1.28**	1.24**	1.37**
# Internet users per 100 people	$3.94\ 10^{-2}$ ***	$3.74\ 10^{-2}$ ***	$3.74\ 10^{-2}$ ***
# Patents	$-6.37\ 10^{-7}$	$-1.02\ 10^{-6}$	$-2.14\ 10^{-6}$
# observations	2451	1535	1170
R^2	0.020	0.028	0.204

High potential losses to cyber attacks. As a measure of countries' intellectual property, I use the number of patents filled by country from the World Intellectual Property Organization [157]. As proxy for the extent to which countries have cyber critical infrastructure, I use the number of Internet users per 100 people [67]. It is worth noting that the number of Internet users per 100 people is also the measure I use for countries' latent abilities. Therefore, in my hypothesis testing, I will be unable to distinguish between whether countries develop cyber warfare capabilities because they have latent abilities or because they are interested in protecting their critical infrastructure.

4.3.2 Results

Table 4.2 shows the results of my hypothesis testing for starting points 2000, 2005, and 2007. From the table, one can see that the results are very similar for different starting points which indicates that the findings are robust. The table indicates that a country is more likely to develop cyber warfare capabilities if the country has an ally that is developing such capabilities. This is indicative of a collaborative effect among countries. For example, members of the NATO collaborate about cyber warfare through the NATO Cooperative Cyber Defense Center of Excellence in Estonia. The table also indicates that the number of Internet users per 100 people also has a significant effect. This may reflect the fact that countries are more likely to develop cyber weapon capabilities if it is easier for them to develop such capabilities or the fact that countries are more likely to develop cyber weapon capabilities if they worry about attacks on their critical infrastructure.

According to Table 4.2, having an enemy that develops cyber warfare capabilities does not have a significant effect on the likelihood that a country develops such capabilities. A possible explanation for this finding is that deterrence of cyber attacks is difficult because of the attribution problem. Another possible explanation is that fear of state-sponsored attacks is not the main reason why countries develop cyber warfare capabilities. Fear of non-state actors such as hackers and terrorists also motivates the development of such capabilities.

Finally, the number of patents does not have a significant effect. One possible explanation is that intellectual property as measured by the number of patents is not the resource that needs protection. Countries are also interested in protecting their government websites and their state secrets.

4.4 Motivation Assessment

4.4.1 Friedkin Model

The Friedkin model [55] is a linear recursive model developed to capture people's change in attitudes over time and adapted to predict many of the European Union's voting decisions [45, 46].

Table 4.3: Variables in the Friedkin model equation

Variable	Dimension	Interpretation
$y^{(t)}$	$N * 1$	actors' attitudes at time t
A	$N * N$	actors' susceptibility to external influence
W	$N * N$	inter-actor influence
I	$N * N$	identity matrix

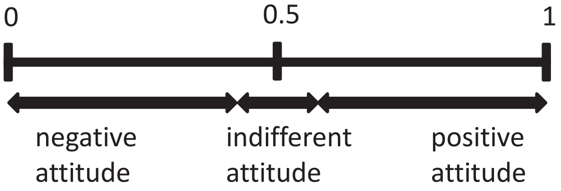


Figure 4.1: Interpretation of attitude values.

The Friedkin model stipulates that a person's opinion at time t is the result of the person's initial opinion and external influence on that person. The model also takes into account that some people are more susceptible to external influence than others. More formally, Equation 4.1 is the model equation for a group of N actors at time $t = 2, 3, ..$

$$y^{(t)} = AWy^{(t-1)} + (I - A)y^{(1)} \tag{4.1}$$

where $y^{(t)} = [y_i^{(t)}]$ is a $N{*}1$ vector of actors' attitudes at time t. $A = diag(a_{11}, .., a_{ii}, .., a_{NN}), 0 \leq a_{ii} \leq 1$ is a $N * N$ diagonal matrix of actors' susceptibilities to external influence where larger values of a_{ii} indicate larger susceptibilities to external influence. $W = [w_{ij}], (0 \leq w_{ij} \leq 1, \sum_{j=1}^{N} w_{ij} = 1)$ is a $N * N$ matrix of inter-actor influence where w_{ij} is the extent to which j influences i. W is computed as $W = AC + I - A$ where $C = [c_{ij}], (c_{ii} = 0, 0 \leq c_{ij} \leq 1, \sum_{j=1}^{N} c_{ij} = 1)$ is a $N * N$ matrix of relative inter-actor influence. Finally, $y^{(1)} = [y_i^{(1)}], 0 \leq y_i^{(1)} \leq 1$ is a $N * 1$ vector of the actors' initial attitudes. The notation of the Friedkin model is summarized in Table 4.3.

In the Friedkin model, it is always the case that $0 \leq y_i^{(t)} \leq 1$ because of the constraints on A, W, and $y^{(1)}$. Figure 4.1 depicts how these attitude values should be interpreted. 0.5 represent an indifferent attitude, smaller values represent negative attitudes (the actor is against the idea), and larger values represent positive attitudes (the actor is for the idea).

It is worth noting that the Friedkin model is *not* a rational actor model. Rational actor models tend to assume that actors have full information and are not influenced by their positions in the social network. In the Friedkin model, actors do not have full information and are influenced by their positions in the network.

4.4.2 Friedkin Model Adaptation

I adapt the Friedkin model to assess countries' motivations for cyber warfare capabilities. More specifically, I capture the fact that a country is more likely to develop cyber warfare capability if the country has an ally that develops such capability, which is the outcome of the my hypothesis testing in Section 4.3.

I compute a country-to-country alliance network $C = [c_{ij}]$ where c_{ij} encodes whether i and j are part of an alliance. I use international alliance data presented in Section 4.3.1 and only take into

Table 4.4: Countries with the highest increase in motivation according to the Friedkin model

Country	**Initial motivation**	**Motivation at equilibrium**
Pakistan	0.50	0.57
Philippines	0.50	0.57
Japan	0.50	0.55
Iceland	0.50	0.54
Luxembourg	0.50	0.53
Portugal	0.50	0.53
Czech Republic	0.50	0.53
Australia	0.68	0.71
Belgium	0.55	0.53
Kyrgyzstan	0.50	0.52

account alliances that are in force as of 2010, the most recent year in the data. I then compute W as $W = AC + I - A$. I set $A = [a_{ii}]$ based on countries' GDP because countries with high GDP are usually less susceptible to international influence [56, 93]. More specifically, I set $a_{ii} = 0.3 + 0.6b_{ii}$ where $b_{ii} = 1 - log(GDP_i)/log(max(GDP))$. b_{ii} is such that $0 \leq b_{ii} \leq 1$ and b_{ii} is smaller for countries with large GDP. Multiplying b_{ii} by 0.6 and adding 0.3 ensures that $0.3 \leq a_{ii} \leq 0.9$, which ensures that all countries are susceptible to both external and internal influence.

I set y^1 based on how long ago countries started developing cyber weapon capabilities and based on how fast countries can build their capabilities. More specifically, for a country i that has incorporated cyber warfare troops in its military, I compute $age_i = 2015 - year_i$ where $year_i$ is the year reported in Table 4.1[1]. I then compute $y_i^1 = 0.3 + 0.6 * \frac{age_i * GDP_i}{\text{max}(age_i * GDP_i)}$. This formula ensures that countries that have incorporated cyber warfare troops in their military have initial motivations between 0.6 and 0.9. According to the formula, countries that have high initial motivation either started developing cyber warfare capabilities early or started relatively late but are able to catch up because of their strong resources. For countries that have not included cyber warfare troops in their military, I consider that $y_i^1 = 0.5$.

Finally, I run the Friedkin model until equilibrium and obtain countries' motivation scores as $y^{(t)}$ at equilibrium. The change in motivations ($y^{(eq)}$ - $y^{(1)}$) has mean -0.07 and standard deviation 0.11. Table 4.4 contains the countries that undergo the biggest increase in their motivations.

4.5 Latent Ability Assessment

In this section, I present my measures of countries' latent cyber capabilities and how I combine these measures into a single latent ability measure.

Research. I measure the strength of countries' cyber security research using the number of cyber security research papers published during the period 2001-2011. The number of research papers of countries is a standard measure of the countries' scientific productivity and expertise [105].

I collect from SCOPUS [123] all research papers that contain the keyword "security" in the title or abstract and that belong to the computer science or engineering areas. I obtain a total of 28,400 research papers. I count a paper as being by a certain country if at least one of the authors has an affiliation in that country. I compute the research indicator of a country i as the logarithm of the number of papers of that a country scaled by the logarithm of the number of papers of the

[1] If i started developing cyber warfare capabilities in the 1990s, I consider that $year_i = 1995$

country that has the largest number of papers. Because of the scaling, the research indicator is always between 0 and 1.

Institutions. The institutions measure captures whether countries have civilian cyber security institutions such as CERTs that the countries can draw expertise from in order to build cyber warfare capabilities. I collect the list of countries that have such institutions from multiple sources [28, 66, 86]. The institutions measure is a binary variable that captures whether a country has such institutions.

IT preparedness. I use the Information and Communication Technology (ICT) development index from the International Telecommunication Union (ITU) branch of the United Nations [67]. The ICT development is a combination of 11 indicators: fixed-telephone subscriptions per 100 inhabitants, mobile-cellular telephone subscriptions per 100 inhabitants, international Internet bandwidth (bits/s) per user, percentage of households with a computer, percentage of households with Internet access, percentage of individuals using the Internet, fixed (wired)-broadband subscriptions per 100 inhabitants, wireless broadband subscriptions per 100 inhabitants. I compute the IT preparedness indicator of a country as the ICT index of the country scaled by the maximum ICT index.

I compute the cyber latent ability score using Equation 4.2. I give the research score and the institutions score a higher weight because those are specific to cyber security, whereas the IT preparedness is not cyber security specific.

$$\text{Latent ability} = (2 * \text{research} + 2 * \text{institutions} + \text{IT preparedness})/5 \tag{4.2}$$

4.6 Assessment Results

I compute cyber warfare capability scores as the product of the motivation scores and the latent capability scores. Figure 4.2 shows a heatmap of such scores for all countries. According to the figure, countries with the strongest cyber warfare capabilities are the United States, China, Israel, South Korea, Germany, the United Kingdom, Australia, Russia, France, Canada, Italy, India, Netherlands, Norway, Denmark, Poland, Spain, Iran, and Japan.

4.7 Limitations and Future Work

The main limitation of this work is that I did not validate the accuracy of this methodology by comparing the methodology's assessment to expert opinions or historical data. As future work, I intend to perform such validation. It is worth noting, however, that I validated a very similar methodology in the context of biological weapons in Chapter 5. Assessing biological weapon capabilities is more challenging than assessing cyber warfare capabilities because biological weapons are prohibited by international treaties and thus the development of biological weapons is surrounded by extreme secrecy.

Moreover, this work exclusively uses publicly-available data which may contain errors and do not accurately capture underground or classified activities. For example, countries may have started developing cyber warfare capabilities earlier than years reported in Table 4.1. If a country started

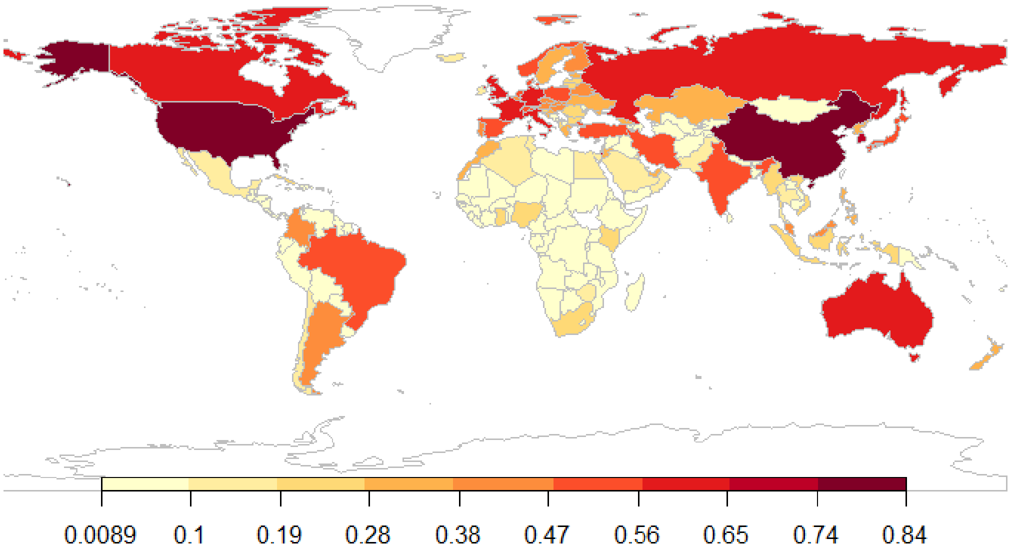


Figure 4.2: Cyber warfare capabilities scores

developing cyber warfare capabilities earlier than the dates reported in Table 4.1, then my assessment will probably under-estimate the country's cyber warfare capabilities because $y^{(1)}$ will be lower than what it should be. Similarly, this work may not accurately capture the availability of highly skilled hackers in a country. It is worth noting, however, that despite the limitations of publicly available data, such data are routinely used and valued by Intelligence agencies [27]. Another limitation is that I only investigate the effect of formal alliances on countries' motivations for cyber warfare capabilities, and do not consider the effect of informal alliances. This is consistent with prior work on proliferation of weapons of mass destruction [56, 64]. As future work, it would be interesting to investigate the effect of non-formal alliances as captured by trade relationships and arm sales. Moreover, this work does not distinguish between defensive and offensive cyber warfare capabilities. As future work, it would be interesting to examine countries' military doctrines to distinguish between those interested in developing defensive capabilities and those interested in developing offensive capabilities. Such approach will not, however, be perfect because some countries may not be upfront about their intentions to develop offensive capabilities. Finally, this work focuses on capabilities of governments, and does not assess the capabilities of non-state actors such as hacker and terrorist groups. Assessing capabilities of non-state actors will probably require a different methodology than the one presented in this chapter and is beyond the scope of this work. Similarly, this work does not capture the case where countries purchase ready-to-use malware or hire other entities to launch attacks on their behalf. This was, for example, the case when the Bahraini government allegedly hired Gamma, a German-British software firm, to spy on Bahraini activists [52]. It is worth noting, however, that a country that purchases a ready-to-use malware or hires another entity to launch attacks on its behalf does not acquire a cyber warfare capability in the same way a country that purchases conventional weapons or weapons of mass destruction acquires capabilities. A country that purchases conventional weapons or weapons of mass destruction can use such weapons repeatedly whereas a piece of malware looses its value after it is uncovered.

4.8 Conclusion

In this chapter, I develop a computational methodology to assess cyber weapon capabilities of all countries in the world. The methodology assess countries' motivations and latent abilities.

I start by testing alternative hypotheses about factors that motivate countries to develop cyber weapons against historical cyber weapon proliferation data. I find that a country is more likely to develop cyber weapons if the country has a military ally that develops such weapons. I also find that a country is more likely to develop cyber weapons if the country has high IT penetration.

In order to assess countries' motivations, I adapt the Friedkin model to capture the fact that countries that have cyber weapon allies are more likely to develop cyber weapons. I set the parameters of the adapted model using publicly available data about military alliances and countries that have included cyber warfare troops into their military. In order to assess countries' latent abilities, I examine countries' cyber security research, cyber security institutions, and general IT preparedness.

I find that countries to watch for about cyber warfare capabilities are the United States, China, Israel, South Korea, Germany, the United Kingdom, Australia, Russia, France, Canada, Italy, India, Netherlands, Norway, Denmark, Poland, Spain, Iran, and Japan. The United States, China, Israel, South Korea, Russia, and India started developing cyber warfare capabilities in the 1990s and have had enough time to build capabilities. The other countries started much later, but have strong economic power, strong cyber latent abilities, and military allies that help them quickly catch up.

Chapter 5 Biological Weapon Capabilities

Research question: How to systematically identify countries that could pose a bioweapon threat?

5.1 Introduction

BW are weapons of mass destruction capable of causing major damage [76]. The development, production and stockpiling of these weapons is prohibited by the Bacteriological Weapons Convention (BWC) [103]. Yet, many countries are believed to have acquired BW despite their membership in the BWC.

Predicting countries that will seek BW gives the international community an opportunity to act early in order to prevent these countries from acquiring such weapons. Such prediction, however, is very challenging. Extreme secrecy surrounds BW related activities and political discussions. Moreover, BW technology is dual-use (i.e. has both civilian and military applications), enabling a country to pursue BW under the cover of civilian industries such as the pharmaceutical industry [153].

Despite the importance of BW proliferation, the problem has attracted limited attention in the literature. We find many case studies [3, 10, 49, 159] that provide an in-depth and important examination of a few countries' BW programs. Unfortunately, case studies typically focus on past or existing programs, and are not predictive. Moreover, case studies are impractical to perform for all countries because they require substantial effort and country-specific expertise. Finally, we find a small body of literature [31, 64, 76, 95, 117, 160] that addresses political science aspects of BW, but that has a different focus than assessing or predicting countries' BW capabilities. For example, Horowitz and Narang [64] empirically test whether countries perceive nuclear, chemical and biological weapons as complement or supplement to each other.

In this chapter, I develop a computational methodology that predicts countries that will seek BW. My methodology examines all countries in the world, and not just a subset. Moreover, non-experts can use my methodology using data collected from the Internet. My methodology assesses countries' BW motivations and abilities, and predicts that a country will seek BW if the country has sufficiently high motivation and ability. In order to assess motivations, I adapt the Friedkin model [55] in order to capture expert opinions about why countries seek BW. In order to assess abilities, I use indicators that include conventional arms purchase from BW countries, national material power and dual-use biological trade.

I validate our methodology by examining my methodology's ability to predict countries that started a BW program later than 1975, the year the Biological Weapon Convention (BWC) was ratified. I use my methodology to make predictions in five year increments starting from 1974 and compare these predictions to the ground truth BW proliferation timeline. I find that my methodology could have successfully predicted that Iran, Syria, Taiwan, Vietnam, Libya and Algeria would start BW programs. My methodology would also have predicted that India, Israel and Cuba would start BW programs. India, Israel and Cuba are suspected in some sources, but whether these countries have had BW programs is unclear. Finally, the only BW proliferator my methodology would have missed is Rhodesia/Zimbabwe.

The remainder of this paper is organized as follows. I provide background in Section 5.2 and an overview of my methodology in Section 5.3. I present my motivation assessment in Section 5.4 and my ability assessment in Section 5.5, and explain how I combine the two assessments in Section 5.6. In Section 5.7, I validate my methodology and make predictions about future BW proliferators. I discuss limitations in Section 5.8 before concluding in Section 5.9.

5.2 Background

I present factors that motivate countries to seek BW in Section 5.2.1 and then explain what it takes to acquire BW in Section 5.2.2. Finally, I cover the Friedkin model in Section 5.2.3.

5.2.1 Motivational Factors

Countries are motivated to develop BW mainly due to in-kind deterrence [76, 149] and deterrence of nuclear weapon use [149]. Having a BW enemy may cause a country to seek BW in order to fill the perceived security imbalance [76, 149]. Similarly, having a nuclear enemy may cause a country to seek BW in order to improve its deterrence posture [149]. This is particularly true when the country lacks the financial and technical infrastructure necessary to build nuclear weapons. It is worth noting that countries that seek BW to deter nuclear enemies are aware of the fact that BW are not perfect substitutes for nuclear weapons [10, 64] and may continue pursuing nuclear weapons even after acquiring BW [64].

It may be unclear at first that BW can play a deterrent role despite BW programs being secret. However, Tucker [149], the founding director of the Chemical and Biological Weapons program at the James Martin Center for Nonproliferation Studies of the Monetary Institute of International Studies, explains that BW can play such a role because countries may hint about their weapons without formally admitting to having them. For example, Israel is able to use nuclear weapons as a deterrent without formally acknowledging having such weapons.

The primary disincentive against seeking BW is the absence of a perceived security deficit [64, 76, 149]. A country that has nuclear reassurance may not see a need for BW. Nuclear reassurance results from the country's own nuclear weapons or from nuclear weapons of an ally that promises retaliation in case the country is attacked. Similarly, strong conventional weapons of the country or its allies may provide reassurance and reduce the need for BW. Other disincentives include the risk of provoking countermeasures, uncertain BW military utility, security problems associated with a BW capability, availability of BW defenses and moral constraints [149].

Pariah or dissatisfied countries are more likely to succumb to factors that motivate them to acquire BW [76]. This is because these countries are more likely to ignore international norms against BW and the risk of provoking counter-measures [76].

5.2.2 Acquiring BW

A country interested in BW may develop these weapons on its own, or may seek foreign assistance.

A report by the Office of Technology Assessment (OTA) [153] presents the steps a country is likely to take in order to develop BW on its own. The country would establish BW facilities and perform BW research before proceeding to pilot production of small quantities of BW agents. The country would then assess the military potential of the agent such as its stability and infectivity. The country would also need to develop and test delivery equipment. Subsequently, the country would produce the agent at a large scale, stabilize it and load it into delivery equipment such as munitions. Finally, the country would stockpile filled or unfilled delivery equipment, and train troops about the use of these weapons. The initial pilot production of BW is relatively easy for a country with modest pharmaceutical industry, but further steps are more complex.

Alternatively, the country may obtain foreign assistance that consists of ready-to-use BW supplies, or help in building its own program. Help in building a BW program can be material, technical or scientific. BW assistance is illegal under the Biological Weapons Convention (BWC), but has allegedly occurred historically. For example, the Soviet Union has allegedly supplied tricothecene mycotoxin to Vietnam for military purposes according to former U.S. Secretary of State Alexander Haig [107]. Similarly, China is suspected of having provided BW assistance [42].

5.2.3 Friedkin Model

The Friedkin model [55] is a simple recursive model that was developed to capture how people change their attitudes towards a certain topic over time. The model was also successfully adapted to predict countries' voting decisions within the European Union [45, 46].

The Friedkin model captures the fact that a person forms their attitudes based on their interactions with others and their own intrinsic beliefs. The model also captures the fact that people differ in their susceptibility to external influence. More specifically, the model stipulates that a person's attitude at time t is a weighted sum of external influence at time $t-1$ and intrinsic beliefs at time 1, where the weight can differ across people. Formally, Equation 5.1 encodes the model for a group of N actors at time $t = 2, 3, ..$

$$y^{(t)} = AWy^{(t-1)} + (I - A)y^{(1)} \tag{5.1}$$

Table 5.1 summarizes the notation used within the Friedkin model. $y^{(t)} = [y_i^{(t)}]$ is a $N * 1$ vector of actors' attitudes at time t and $A = diag(a_{11}, .., a_{ii}, .., a_{NN}), 0 \leq a_{ii} \leq 1$ is a $N * N$ diagonal matrix of actors' susceptibility to external influence. $a_{ii} = 1$ indicates that i's attitude exclusively depends on external influence, $a_{ii} = 0$ indicates that i's attitude exclusively depends on intrinsic beliefs, whereas intermediate values indicate that i's attitude depends on both external influence and intrinsic beliefs. $W = [w_{ij}], (0 \leq w_{ij} \leq 1, \sum_{j=1}^{N} w_{ij} = 1)$ is a $N * N$ matrix of inter-actor influence. More specifically, w_{ij} is j's influence on i. The general formulation of the Friedkin model stipulates that $W = AC + I - A$ where $C = [c_{ij}], (c_{ii} = 0, 0 \leq c_{ij} \leq 1, \sum_{j=1}^{N} c_{ij} = 1)$ is a $N * N$ matrix of relative inter-actor influence. Finally, $y^{(1)} = [y_i^{(1)}], 0 \leq y_i^{(1)} \leq 1$ is a $N * 1$ vector of the actors' initial attitudes.

In the Friedkin model, we always have $0 \leq y_i^{(t)} \leq 1$ because of the constraints on A, W and $y^{(1)}$. Attitude values are interpreted according to Figure 5.1 where values around 0.5 represent an indifferent attitude, larger values represent a positive attitude and smaller values represent a negative attitude.

The Friedkin model equation contains three main parts. $Wy^{(t-1)}$ represents actors' extrinsic attitudes resulting from external influence, $y^{(1)}$ encodes actors' intrinsic beliefs and A encodes actors' susceptibility to external influence. The constraint $0 \leq w_{ij} \leq 1$ implies that j's influence

Table 5.1: Variables in the Friedkin model equation

Variable	Dimension	Interpretation
$y^{(t)}$	$N * 1$	actors' attitudes at time t
A	$N * N$	actors' susceptibility to external influence
W	$N * N$	inter-actor influence
I	$N * N$	identity matrix

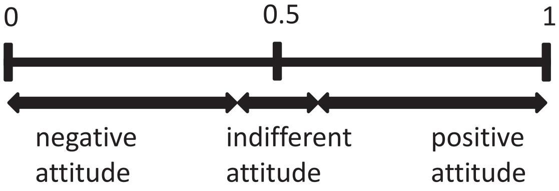


Figure 5.1: Interpretation of attitude values.

on i's attitude is in the same direction as j's attitude. For example, if j has a positive attitude, j will influence i to also have a positive attitude. This constraint reflects the dynamics of friendly human relationships, but unfortunately does not capture the dynamics of BW proliferation. In the context of BW proliferation, both hostile and friendly international relationships play a role into shaping a country's attitude towards BW. In Section 5.4.1, I derive a new extrinsic motivation term that I use to replace $Wy^{(t-1)}$ in Equation 5.1.

5.3 Methodology Overview

The goal of my methodology is to predict countries that will seek BW. More specifically, when used at year yr, my methodology predicts the set of countries that will start BW programs later than yr.

As depicted in Figure 5.2, my methodology consists of a motivation assessment component and an ability assessment component. I start by the Friedkin model considering the attitude of interest to be motivation for BW and the N actors to be all countries in the world. I modify the Friedkin model equation (viz. Equation 5.1) in order to incorporate factors that motivate countries to seek BW discussed in Section 5.2.1. I set the parameters of our adapted model equation based on data on international hostilities and alliances, countries' diplomatic isolation and BW suspected countries. I run the adapted model equation until equilibrium obtaining countries' BW motivation scores.

I identify and use 3 indicators of BW ability. The *arms* indicator captures conventional weapons purchase from BW countries, the *power* indicator captures national material power and the *trade* indicator captures dual-use biological trade. I combine these 3 indicators into a single ability score.

I predict that a country will start a BW program if the country has sufficiently high motivation and ability i.e. the motivation and ability scores are higher than some thresholds.

I note that my methodology predicts whether a country will start a BW program, but does *not* predict 1) the year the country will start such program, 2) whether a country will stop an existing BW program, and 3) whether the country will acquire BW or will simply keep pursuing them. I leave developing a methodology that predicts the above 3 points for future work.

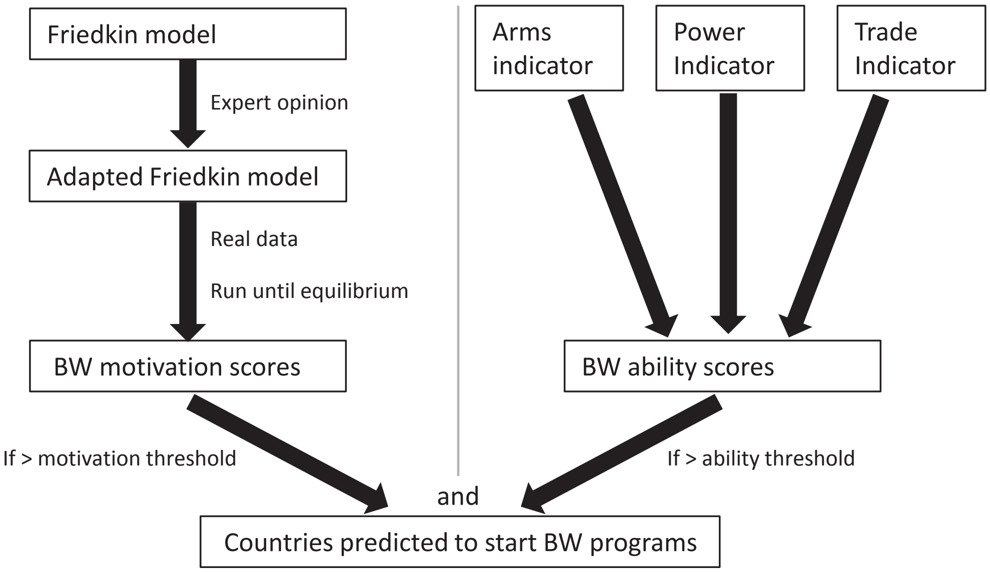


Figure 5.2: Methodology overview

5.4 BW Motivation Assessment Model

I explain how I adapt the Friedkin model equation to capture factors that motivate countries to seek BW in Section 5.4.1 and set the parameters of the adapted model equation in Section 5.4.2.

5.4.1 Model Equation Adaptation

In this work, I consider the N actors to be all countries in the world and the attitude of interest to be motivation for BW. I derive a new extrinsic motivation term to replace $Wy^{(t)}$ in Equation 5.1. $Wy^{(t)}$ is unable to capture factors that motivate countries to seek BW as explained in Section 5.2.3. The new extrinsic motivation term will capture two incentives: in-kind deterrence and deterrence of nuclear weapon use, and one disincentive: nuclear reassurance.

In order to simplify the discussion, I first consider a single country that has a single enemy and a single ally that promises retaliation in case the country is attacked. I derive an expression for the country's extrinsic motivation as a function of: 1) whether the enemy has BW, 2) whether the enemy has nuclear weapons and 3) whether the country has nuclear reassurance. Table 5.2 presents a qualitative description of the country's extrinsic motivation as a function of these parameters. The extrinsic motivation is high in scenarios 1, 2 and 3 where the enemy has BW and/or nuclear weapons, and the country has no nuclear reassurance. The country perceives a great security deficit and has incentive to acquire BW in order to fill that deficit. The extrinsic motivation is moderate in scenario 4 where both the country and the enemy have access to nuclear weapons, but the enemy has BW in addition. The country may be interested in BW despite having nuclear reassurance in order to be able to respond to BW in-kind [149]. Historically, some nuclear powers such as the United States did not perceive the need to maintain BW programs [76], whereas others such as Russia perceived such a need [64]. The extrinsic motivation is indifferent in scenarios 5, 6 and 7. In scenario 5, the enemy has no BW and no nuclear weapons, resulting in the absence of the in-kind deterrence and the nuclear weapon use deterrence incentives. In scenario 6, the country has access

Table 5.2: Qualitative description of a country's extrinsic BW motivation depending on the country's international security environment

Scenario	BW enemy	Nuclear enemy	Nuclear reassurance	Extrinsic BW motivation
1	yes	yes	no	very high
2	yes	no	no	very high
3	no	yes	no	high
4	yes	yes	yes	moderate
5	no	no	no	indifferent
6	yes	no	yes	indifferent
7	no	yes	yes	indifferent
8	no	no	yes	low

Table 5.3: Variables used in the derivation of the new extrinsic motivation term

Variable	Dimension	Interpretation
$m^{(t)}$	$1 * 1$	country's extrinsic motivation for BW at t
$b^{(t-1)}$	$1 * 1$	whether the country's enemy has BW at $t-1$
k	$1 * 1$	whether the country's enemy has nuclear weapons
r	$1 * 1$	whether the country has nuclear reassurance
$M^{(t)}$	$N * 1$	countries' extrinsic motivation for BW at t
$B^{(t-1)}$	$N * 1$	whether the countries' enemies have BW at $t-1$
K	$N * 1$	whether the countries' enemies have nuclear weapons
R	$N * 1$	whether the countries have nuclear reassurance

to weapons that are stronger than the enemy's weapons. In scenario 7, there is a security balance between the country and the enemy. Finally, the extrinsic motivation is low in scenario 8 as the country is in a much stronger security posture than the enemy.

I now derive the extrinsic motivation term for a single country. Let $m^{(t)}$ be the country's extrinsic motivations at time t, $b^{(t-1)}$ whether the enemy has BW at time $t-1$, k whether the enemy has nuclear weapons and r whether the country has nuclear reassurance. I assume k and r to be constant during the time period for which I make predictions because developing nuclear weapons takes much longer than developing BW [153]. $m^{(t)}$, $b^{(t-1)}$, k and r have the scaling depicted in Figure 5.1. Table 5.3 summarizes my notation.

Table 5.4 presents a quantitative encoding of Table 5.2. I find an expression for $m^{(t)}$ that satisfies the constraints in Table 5.4, obtaining the extrinsic motivation term in Equation 5.2.

Table 5.4: Quantitative description of a country's extrinsic motivation for BW depending on the country's international security environment

$b^{(t-1)}$	k	r	$m^{(t)}$
1	1	0.5	0.9
1	0.5	0.5	0.9
0.5	1	0.5	0.85
1	1	1	0.7
1	0.5	1	0.5
0.5	0.5	0.5	0.5
0.5	1	1	0.5
0.5	0.5	1	0.3

$$m^{(t)} = \big(2.6 - 2.8k - 2.2r + 2.8k.r\big)b^{(t-1)} - 1.1 + 2.4k + r - 2.6k.r \quad (5.2)$$

I now consider the case of N countries. Let $M^{(t)}$ be a $N * 1$ vector that represents countries' extrinsic motivation for BW at time t, $B^{(t-1)}$ a $N * 1$ vector that represents whether countries' enemies have BW at time $t-1$, K a $N * 1$ vector that represents whether countries' enemies have nuclear weapons and R a $N * 1$ vector that represents whether countries have nuclear reassurance. The notation is summarized in Table 5.3. The extrinsic motivation term in Equation 5.2 can be written for all countries as Equation 5.3 where $\circ$ denotes point by point multiplication of vectors.

$$M^{(t)} = \big(2.6 - 2.8K - 2.2R + 2.8K \circ R\big)B^{(t-1)} - 1.1 + 2.4K + R - 2.6K \circ R \quad (5.3)$$

$B^{(t-1)}$ can be written as $B^{(t-1)} = Wy^{(t-1)}$ where W captures international hostilities and $y^{(t-1)}$ captures whether countries have BW at time $t-1$. The extrinsic motivation term in Equation 5.3 can thus be written as Equation 5.4.

$$M^{(t)} = \big(2.6 - 2.8K - 2.2R + 2.8K \circ R\big)Wy^{(t)} - 1.1 + 2.4K + R - 2.6K \circ R \quad (5.4)$$

Finally, I substitute $Wy^{(t-1)}$ in Equation 5.1 by the new extrinsic motivation term obtaining the adapted Friedkin equation model in Equation 5.5.

$$y^{(t)} = A\Big[\big(2.6{-}2.8K{-}2.2R{+}2.8K{\circ}R\big)Wy^{(t-1)}{-}1.1{+}2.4K{+}R{-}2.6K{\circ}R\Big]{+}(I{-}A)y^{(1)} \quad (5.5)$$

5.4.2 Populating the Adapted Model

In this section, I explain how I set the parameters of Equation 5.5 when using the model at year yr.

$A = diag\big(a_{11}, ..., a_{ii}, ..., a_{NN}\big)$ I set A based on the diplomatic isolation index, the main indicator for countries' pariah status [60, 61, 81]. The diplomatic isolation index is computed as the ratio of the number of adjacent countries and major powers with whom a country does not have diplomatic exchanges to the total number of adjacent countries and major powers [69]. I consider two countries to be adjacent if they share a direct land or river border, or are within 400 miles of each other by body of water [37, 135]. I use the list of major powers in the Correlates of War state membership list data [39]. Finally, I use the diplomatic exchange data set [7], and consider that two countries have diplomatic exchanges if they exchange ambassadors or ministers.

I compute the average diplomatic index of each country during the 5 year period $[yr-4, yr]$. When I am able to compute the index for only a single year during that period, I simply use the index for that year. This approach is more robust to fluctuations in the index and to missing data. I obtain i's susceptibility to external influence as $a_{ii} = 0.1 + 0.7 *$ average diplomatic index. By definition, the average diplomatic index is within the range [0,1]. Multiplying that index by 0.7 and adding 0.1 narrows that range allowing all countries to be influenced by both external influence and intrinsic beliefs.

$K = [k_i]$ I identify countries' enemies based on the international hostility data from the International Crisis Behavior Project (ICB) [25] and identify whether these enemies have nuclear weapons using the nuclear proliferation timeline from Jo and Gartzke [69].

The ICB data cover military and non-military hostilities. I keep hostilities of type threat to existence, threat of grave damage, threat to influence, territorial threat, political threat and limited military damage, and disregard purely economic conflicts[1]. I consider two countries to be enemies if I find a hostility between the two countries anytime during $[yr - 14, yr]$. I set $k_i = 1$ if i has an enemy that possesses nuclear weapons anytime during $[yr - 4, yr]$ and set $k_i = 0.5$ otherwise. As a result, k_i satisfies the scaling in Figure 5.1.

$R = [r_i]$ I identify countries' allies using the alliance data from the Correlates of War project [36, 57, 131, 132] and whether these allies have nuclear weapons using the nuclear timeline from Jo and Gartzke [69].

From the alliance data, I exclusively keep alliances that belong to the defense pact category. Such alliances commit countries to intervene military on the side of any treaty partner that is attacked. I omit alliances that belong to the categories of neutrality, non-aggression treaty and entente agreement. Neutrality and non-aggression pacts specify that parties remain military neutral if any co-signatory is attacked, while ententes pledge consultation and/or cooperation in a crisis, including armed attack. I set $r_i = 1$ in case a country's ally possesses nuclear weapons anytime during the period $[yr - 4, yr]$ and set $r_i = 0.5$ otherwise.

$W = [w_{ij}]$ I first extract a hostility matrix $H = [h_{ij}]$ from the ICB hostility data [25] where $h_{ij} = 1$ indicates a hostility between i and j any time during $[yr - 14, yr]$ and $h_{ij} = 0$ indicates otherwise. I take into consideration hostilities of type threat to existence, threat of grave damage, threat to influence, territorial threat, political threat and limited military damage, and disregard purely economic threats. I divide each row of H by the sum of that row, obtaining a matrix C that satisfies the requirements $(c_{ii} = 0, 0 \leq c_{ij} \leq 1, \sum_{j=1}^{N} c_{ij} = 1)$ discussed in Section 5.2.3. Finally, I compute W using the formula $W = AC + I - A$ also given in Section 5.2.3.

$y^{(1)} = [y_i^{(1)}]$ I use the main BW proliferation timeline by Horowitz and Neil [64] after making a few changes to that timeline. First, I extend the timeline from 2000 to 2008 based on the same sources used by Horowitz and Neil. I stop at 2008 because a very small number of sources are newer than 2008. In addition, I remove Bulgaria, Cuba and Laos from the list of suspected countries. Horowitz and Narang suspect these 3 countries based only on a 1993 OTA report [152] [2], but the report in question does not suspect these countries. More specifically, the OTA report examines 6 sources and decides to suspect a country if $2/3$ of these 6 sources suspect that country. The OTA report finds that only a single source suspects Bulgaria, Cuba and Laos, and decides to exclude these 3 countries from the list of suspected countries[3]. The final main BW proliferation timeline is given in Table 5.5. I set $y_i^{(1)} = 1$ in case i has a BW program anytime during the period $[yr - 4, yr]$ and set $y_i^{(1)} = 0.5$ otherwise.

It is worth noting that Horowitz and Neil also include an alternative BW proliferation timeline[4]. Horowitz and Neil include such timeline because of the uncertainty surrounding countries' BW

[1] I make that selection using the "gravcr" variable.

[2] The sources Horowitz and Narang used to compile the timeline are given in Appendix Table 12 of their paper.

[3] The final list of countries suspected by the OTA report is given in Table 2-8 of the report.

[4] Horowitz and Neil provide the alternative timeline in Appendix Table 12 of their paper.

Table 5.5: “Ground-truth” BW proliferation timeline 1970-2008

Country	Main Timeline	Alternative Timeline
Algeria	1999-2008	1999-2008
China	1970-2008	-
Egypt	1970-2008	1970-2008
France	1970-1973	1970-1973
India	-	1970-2000
Iran	1981-2008	1981-2008
Iraq	1974-2003	1974-2003
Israel	-	1970-2008
Libya	1988-2003	1988-2003
N. Korea	1970-2008	1970-2008
Rhodesia/Zimbabwe	1975-1980	-
USSR/Russia	1970-2008	1970-2008
S. Africa	1970-1993	1970-1993
Syria	1990-2008	-
Taiwan	1975-1993	-
Vietnam	1988-1993	-

programs. I extend that timeline to 2008 and include Israel [35, 68, 85, 149, 152]. I will *not* use the alternative timeline to set $y^{(1)}$, but I will use it when validating the methodology in Section 5.7. The alternative timeline is also given in Table 5.5.

It it worth noting that the timeline by Horowitz and Narang distinguishes between pursuit and acquisition. In this work, however, I refrain from making that distinction because of the uncertainty about the progress of countries' BW programs. More specifically, a country can become aware that its enemy has a BW program, but it is more difficult for the country to know whether its enemy has already acquired BW or is still pursuing them.

5.5 Ability to Acquire BW

In this section, I present the indicators I suggest and use to assess countries' abilities to acquire BW. I put more emphasis on assessing countries' abilities to obtain foreign assistance than on assessing their abilities to develop BW on their own. Countries that started BW programs in recent decades are typically developing countries that may have difficulty building BW on their own and may prefer to seek foreign assistance. In my discussion, I assume that I am interested in assessing countries' abilities at year yr.

5.5.1 Conventional Arms Purchase from BW Countries

The first indicator I use is the value of conventional arms purchased from countries that have BW. A country that purchases large amounts of conventional weapons from a supplier that has BW may be able to obtain BW from that supplier. Conventional arms trade is often legal and public, contrary to BW trade.

I use the Stockholm International Peace Research Institute (SIPRI) arms trade database [136] which covers trade of aircrafts, air defense systems, anti-submarine warfare weapons, armored vehicles, artillery, military engines, missiles, radar and sonar systems, reconnaissance satellites and military ships. The SIPRI data is reported in constant 1990 USD. I compute the *arms* score as the value of conventional arms purchased from BW countries during the period $[yr - 4, yr]$. I identify BW countries using the main BW proliferation timeline in Table 5.5.

5.5.2 National Material Power

The second indicator is countries' national material power. This indicator is useful because powerful countries may encounter less difficulty developing or purchasing weapons. I measure national material power using the standard Composite Index of National Capability (CINC) score [38, 130, 133]. The CINC score is computed based on countries' total population, urban population, iron and steel production, energy consumption, military personnel, and military expenditure. I compute the *power* score as the average CINC score over the period $[yr - 4, yr]$.

5.5.3 Dual-Use Biological Trade

As third indicator, I use the trade of dual-use biological commodities i.e. commodities with both military and civilian biological applications. Examples of these commodities are sterilization equipment and delivery mechanisms. Dual-use biological trade is relevant because a country may acquire equipment to develop BW under the cover of civilian industries such as the pharmaceutical industry.

I collect the dual-use biological trade from the UN Comtrade database [150], a publicly available depository of international trade data. Countries inform the United Nations Statistics Division (UNSD) of their international trade at the end of each year and the UNSD makes the data available though the UN Comtrade database. I collect the trade data by specifying the codes of dual-use biological commodities [62], and adjust for inflation using the Producer Price Index (PPI) [110]. I compute the *trade* score as the value of dual-use biological trade during the period $[yr - 4, yr]$.

5.6 Combined Motivation and Ability Assessment

In this section, I explain how I use the motivation assessment model and the ability scores in order to predict future BW proliferators. More specifically, at a given year yr, the goal of the methodology is to predict countries that will start BW programs later than yr.

I set the parameters of the adapted model equation (viz. Equation 5.5) as explained in Section 5.4.2. I then run that iterative equation until equilibrium i.e. $y^{(t+1)} = y^{(t)}, t >= eq$. My assessment of countries' motivation scores is $y^{(eq)}$.

I combine the arms, power and trade scores into a single ability score using the formula in Equation 5.6.

$$\text{ability} = 8\log(\text{arms}) + 5\log(10^7\text{power}) + \log(\text{trade}) \tag{5.6}$$

I use the logarithm because the arms, power and trade scores differ by several orders of magnitude across countries. I multiply the power score by 10^7 because the minimum power score is 2.3 10^{-7} and I need to make sure I obtain values larger than 1 before applying the logarithm. I weigh the 3 indicators depending on their relevance to acquiring BW. I give the arms score the highest weight because a strong military relationship with a BW country can be the perfect channel for obtaining BW. I give the power score the second highest weight because that score captures countries' military power and investment. Finally, the trade score has the lowest weight because dual-use biological trade contains, by definition, considerable amount of civilian trade.

My methodology predicts that i will start a BW program if i has sufficiently high BW motivation i.e. $y_i^{(eq)} > 0.53$ and sufficiently high BW ability i.e. ability score larger than 132. I obtain similar results if I use any motivation threshold between 0.521 and 0.534, and any ability threshold between 131 and 133.

5.7 Validation and Prediction

5.7.1 Validation

In order to validate our methodology, I evaluate whether my methodology could have predicted historical BW proliferators. More specifically, I use my methodology to make predictions in 5 year increments starting at 1974. I make the first prediction at 1974 because the BWC was ratified in 1975, and make the last prediction as early as 1999 in order to have a ground-truth "future" to compare the methodology's prediction to.

Table 5.6 summarizes the accuracy measures I use to evaluate our methodology. I decide whether a country is a true positive (TP), a false positive (FP), a true negative (TN) or a false negative (FN) by comparing the methodology's prediction to the main timeline. For example, a country is a TP if the methodology predicts that the country will start a BW program after yr and the country starts such program in the main timeline.

Regular measures reflect the case where the alternative timeline agrees with the main timeline about how to label a country, and contentious measures reflect otherwise. For example, a regular TP will also be considered a TP if the methodology's prediction is compared to the alternative timeline instead of the main timeline. On the other hand, a contentious TP will be considered a FP if the methodology's prediction is compared to the alternative timeline. In other words, according to the alternative timeline, a regular TP starts a BW program after yr, but a contentious TP does not. Because of my definition of regular and contentious measures, all countries on which the two timelines disagree will necessarily be contentious. I note that from a policy standpoint, a contentious TP is preferable to a contentious FN. A contentious TP is a warning that causes the international community to closely monitor a country in order to reduce the uncertainty, whereas a contentious FN may cause the international community to miss a potential BW proliferator.

In the table, positive predictions (countries predicted to start BW programs) are compared against the entire subsequent timeline, whereas negative predictions are compared against the subsequent 5 years. In other words, the methodology is allowed (and encouraged) to provide a very early warning about a BW program, but is only penalized when missing a BW program at the prediction year immediately preceding the start of that program.

Finally, accuracy measures are only defined for countries that do not have BW programs as of yr according to the main timeline. For other countries, the motivation assessment model sees through the initial condition that these countries already have BW programs. Therefore, the methodology should not be rewarded for "predicting" that these countries will have BW programs.

Table 5.7 summarizes accuracy results of the methodology's predictions. From the table, it is possible to see that the methodology successfully predicts all countries that started BW programs after 1975 with the exception of Rhodesia/Zimbabwe. The methodology typically provides a very

Table 5.6: Accuracy measures for prediction made at year yr.

Measure	**Methodology: country will start BW program after** yr	**Main timeline: country starts BW program**	**Alternative timeline: country starts BW program**
Regular true positive	yes	yes after yr	yes after yr
Regular false positive	yes	never	never
Regular true negative	no	not during $[yr+1, yr+5]$	not during $[yr+1, yr+5]$
Regular false negative	no	yes during $[yr+1, yr+5]$	yes during $[yr+1, yr+5]$
Contentious true positive	yes	yes after yr	never
Contentious false positive	yes	never	yes anytime
Contentious true negative	no	not during $[yr+1, yr+5]$	yes during $[yr+1, yr+5]$
Contentious false Negative	no	yes during $[yr+1, yr+5]$	not during $[yr+1, yr+5]$

Table 5.7: Accuracy of the methodology's predictions.

	1974	**1979**	**1984**	**1989**	**1994**	**1999**
True positives	Iran Syria* Taiwan* Vietnam*	Algeria Iran Libya Syria* Vietnam*	Algeria Libya Syria* Vietnam*	Algeria Syria*	Algeria	
False positives	India* Israel*	India*	Cuba			
False negatives	Rhodesia*					
True negatives	132 countries	138 countries	141 countries	145 countries	171 countries	180 countries

Countries with a (*) are contentious cases, other countries are regular cases

early warning, predicting that countries will start BW programs many years before the start of these programs. Syria, Taiwan and Vietnam are contentious TP in the sense that these countries appear in the main timeline, but not in the alternative one. The methodology also predicts that India, Israel and Cuba will start BW programs. India and Israel are contentious FP because India and Israel do not appear in the main timeline, but appear in the alternative one. Cuba is a regular FP because neither the main nor the alternative timeline includes Cuba. It is worth noting that Cuba is sometimes suspected of having had a BW program [99], but is excluded from both timelines because the sources are not conclusive. Finally, TN countries at a given prediction year are the countries that have not started a BW program as of that year according to the main timeline, and that are neither TP nor FP nor TN. The number of TN increases over time because the total number of countries in the world increases over time.

5.7.2 Prediction

When running the methodology for years 2004 and 2008, the methodology predicts that no additional country will start a BW program.

5.8 Limitations

The methodology's accuracy when used on future data is not guaranteed to be the same as the accuracy reported in this paper. For example, factors that motivate countries to seek BW might change in the future. Disease outbreaks such as the recent Ebola outbreak may motivate countries to acquire BW even if these outbreaks are unrelated to BW activities. This is because these outbreaks revive the potential damage and casualties of BW. Alternatively, as newer technologies such as cyber weapons become more relevant, countries may become less interested in BW. Another limitation is that I set many numerical parameters relatively arbitrarily. As future work, I intend to perform a full sensitivity analysis that investigates how varying these parameters affects the results.

Finally, I focus on countries and overlook terrorist groups. Biological terrorism poses a real threat, but is beyond the scope of this work. Such terrorism needs to be studied using a methodology different than the one presented in this paper. Typically, factors that motivate terrorist groups are different from factors that motivate countries. Moreover, data about terrorist groups is more difficult to collect than data about countries. I refer to Koblentz [77] for discussion about biological terrorism, and to Tucker [148] and the START database [144] for open-access data sets about terrorist and criminal use of weapons of mass destruction.

5.9 Conclusion

In this chapter, I develop a novel computational methodology to predict countries that will seek BW. My methodology assesses countries' BW motivations and abilities, and predicts that a country will start a BW program if the country has sufficiently high motivation and ability. Assessing both motivations and abilities makes my methodology robust despite the error-prone nature of our data sources.

I develop a socio-cultural model to assess countries' motivations by capturing expert opinions about why countries seek BW into the Friedkin model. I set the parameters of my adapted model based on publicly-available data about international hostilities and alliances, countries' diplomatic isolation and the list of countries that have BW. Moreover, I suggest and use publicly-available indicators of countries' abilities to acquire BW. My indicators are conventional arms purchase from BW countries, national material power and dual-use biological trade.

I validate our methodology by assessing our methodology's ability to predict countries that started a BW program after 1975, the year the BWC was ratified. I find that my methodology could have successfully predicted BW programs of Iran, Syria, Taiwan, Vietnam and Algeria. The methodology would also have suspected India, Israel and Cuba. Some sources suspect these 3 countries, but it is uncertain whether these countries have had such programs. The only country the methodology would have missed is Rhodesia/Zimbabwe. It is worth noting that despite the excellent accuracy of my methodology, such methodology should not be the sole basis of sanctioning countries. Detailed investigation and on-site visits should be carried out before taking action against countries.

Chapter 6 Joint Cyber and Biological Weapon Capabilities

Research questions: How to systematically identify countries that could pose both a cyber weapon and a bioweapon threat?

6.1 Introduction

Both cyber and biological weapons are considered an asymmetric threat in the sense that a small actor can use these weapons to inflict unacceptable damage on a much stronger military opponent. This is because both weapons are relatively easy to acquire while being capable of causing major damage. A country that has both weapons may be tempted to use two concurrently in order to amplify the damage. For example, a country can use a cyber attack to access classified information about bioweapon preparedness of its opponent such as vaccine stocks and protective equipment. Using that information, the country can optimize its biological attack in order to incur maximal damage. Similarly, the country can use malware to disrupt hospitals' operations while concurrently launching a biological attack. In this context, it is important to identify countries that could acquire both cyber weapons and bioweapons.

Prior work [3, 10, 16, 49, 58, 63, 92, 159] mainly examines countries' cyber weapon and bioweapon capabilities separately and mainly consists of case studies on one or a few countries. Unfortunately, such approach overlooks assessing the risk of joint cyber-biological capabilities. Moreover, such approach is not systematic in the sense that it does not cover all countries in the world. Other prior work discusses past cyber-physical attacks such as the joint cyber physical attack on Georgia [22] and the joint cyber physical attack on Syria [1]. In 2008, Russia allegedly launched a massive Denial of Service attacks on websites in Georgia concurrently with the physical attack on Georgia. In 2007, Israel launched an air strike on Syria while allegedly using a cyber attack to blind Syrian radars [1]. The cyber attack allowed Israeli jets to complete their mission without being detected or challenged. Besides discussion of joint cyber physical attacks that have happened in real life, the literature also contains discussion of how cyber and physical attacks could be combined. The National Research Council report [113] discusses possible scenarios of joint cyber physical attacks. For example, the report explains that a cyber attack on microelectronics that control the military munitions could degrade these munitions. Sharma [126] explains that physical force should be thought of as enhancer of cyber attacks, instead of the other way around. Whether physical force enhances cyber attacks, or cyber attacks enhance physical force is irrelevant to this thesis. McConnel [98] explains that physical attacks can be a vector to launch cyber attacks. Examples include a missile attack on satellites, or electromagnetic pulses to cripple power lines. Joint cyber biological attacks have never been reported in real life. However, these attacks are still important to consider proactively. I was unable to find discussion of such joint attacks in the literature.

In this chapter [53], I identify countries to watch for about joint cyber biological capabilities.

Table 6.1: Motivational factors

Biological	Cyber
Incentive: enemy has nuclear weapons or biological weapons Disincentive: country or ally has nuclear weapons Disincentive: risk of provoking counter-measures	Incentive: Ally develops cyber weapons

In order to identify these countries, I leverage the methodologies and results developed in Chapters 5 and 4. The remainder of this chapter is organized as follows. In Section 6.2, I compare my approaches to assessing countries' motivations for bioweapons and cyber weapons. In Section 6.3, I compare my approaches to assessing countries' bioweapon and cyber weapon latent abilities. Finally, In Section 6.4, I identify countries to watch for about joint cyber biological capabilities.

6.2 Motivation Assessment

In order to assess countries' motivations for biological weapons (cyber weapons), I adapt the Friedkin socio-cultural model to capture factors that motivate countries to develop these weapons. Table 6.1 presents these motivational factors for biological weapons and cyber weapons. For bioweapons, the main incentive is deterrence of a nuclear or a biological enemy. Bioweapons are sometimes called the "poor man's atomic bomb" in reference to the fact that poor countries that cannot afford nuclear weapons may be interested in bioweapons in order to improve their deterrence posture. The main disincentive against developing bioweapons is having nuclear weapons or having an ally that has such weapons and that promises retaliation in case the country is attacked. Such access to nuclear weapons deters biological attacks and reassures the country. Finally, the risk of provoking counter-measures is an important disincentive against developing bioweapons. Bioweapons are prohibited by international treaties and countries that develop them run a serious risk of provoking counter-measures such as economic sanctions.

The main incentive for developing cyber weapons is having an ally that develops such weapons. Such ally can assist the country in its efforts to develop indigenous capabilities. Capabilities of allies do not provide a disincentive against developing indigenous capabilities. An ally cannot promise retaliation against cyber attacks because of the attribution problem. Moreover, an ally can provide limited help defending against cyber weapons because this defense may require the ally to have full access to the country's sensitive systems. Countries may be uncomfortable giving such full access to their allies. Finally, a country that develops cyber warfare capabilities does not run a risk of provoking counter-measures because cyber warfare capabilities are not prohibited by international treaties.

6.3 Latent Abilities Assessment

Table 6.2 presents the indicators that I use to measure countries' biological weapon and cyber weapon capabilities. For biological weapons, I mainly focus on the ability of countries to purchase the capability or equipment necessary to build such capability. In recent decades, biological weapons have mainly been attractive to third-world countries with limited biological infrastructure and expertise. It is easier for those countries to seek external assistance than to develop biological

Table 6.2: Latent abilities indicators

Biological	Cyber
- Conventional arms purchase from bioweapon countries	- Cyber security research
- Dual-use biological trade	- Cyber security institutions
- National material power	- IT preparedness

weapons entirely by themselves. External assistance can consist of dual-use biological equipment or ready-to use weapons. I do not use a research indicator for bioweapon latent capabilities because the scientific knowledge required to build bioweapons is readily available. Moreover, bioweapon research is dual-use i.e. has both civilian and military applications. The most active countries in such research are countries with strong pharmaceutical industries that are interested in developing new medicines and vaccines.

For cyber weapons, I mainly focus on cyber security expertise gained through research and institutions. Cyber security is a relatively new area and developing cyber weapons requires cutting-edge expertise. Developing cyber weapons does not require raw material or expensive equipment. Finally, purchasing ready-to-use cyber warfare capability is difficult. A country can purchase sophisticated malware. However, this is different from acquiring a capability because malware is single-use in the sense that malware looses its value after it is discovered. If the country wants to "purchase" the ability to repeatedly create sophisticated attacks from other countries, the country needs to hire international cyber security experts to work for its military. This is risky because these international experts may not be loyal to the country. Finally, the country can seek international assistance training its own experts. Such situation is already captured by the motivation assessment part of the model.

6.4 Joint Cyber Biological Capabilities

The goal of this section is to identify countries to watch for about joint cyber and biological capabilities. Countries to watch for about cyber weapon capabilities are listed in Section 4.6. Countries with the highest capabilities are the United States, China, Israel, South Korea, Germany, the United Kingdom, Australia, Russia, France, Canada, Italy, India, Netherlands, Norway, Denmark, Poland, Spain, Iran, and Japan.

In order to identify countries to watch for about biological weapon capabilities, I use the methodology described in Section 5.6 to make predictions at 5 year increments starting from 1979 similarly to the technique described in Section 5.7. I consider that a country should be watched for about biological weapons if the methodology predicts that the country will develop biological weapons at any of the increments or if the country has biological weapons as of 1979 according to Table 5.5. I go back in time when making the predictions because countries may keep biological weapons developed many years ago even if, currently, the motivation and latent abilities are not very strong. Countries to watch for about biological weapons are China, Egypt, Iraq, North Korea, Zimbabwe, Russia, South Africa, Iran, Syria, Taiwan, Vietnam, Libya, India, Israel, and Cuba.

Thus, countries to watch for about joint cyber biological capabilities are China, Israel, India, Russia, and Iran.

6.5 Conclusion

Combined cyber biological attacks could be order of magnitudes more serious than cyber only attacks or biological only attacks. Many countries are to be watched for about joint cyber biological capabilities. These countries are China, Israel, India, Russia, and Iran.

Countries that invest in cyber weapons but not biological weapons tend to be developed countries whereas countries that invest in biological weapons but not cyber weapons tend to be developing countries that are isolated internationally. This is because developing cyber weapons requires cutting edge cyber security expertise and is not prohibited under international law. On the other hand, the expertise required to develop biological weapons is relatively readily available and developing biological weapons is prohibited under international law (making these weapons mainly appealing to countries that are willing to break such law).

Chapter 7 Conclusion

In this thesis, I take an empirical systematic approach towards assessing the global cyber and biological threat. The first part of the thesis focuses on cyber crime and aims at identifying factors that cause some countries' environments to be conducive to cyber crime. The second part of the thesis focuses on cyber and biological weapon capabilities and aims at identifying countries that develop such capabilities.

In the first part of the thesis, I empirically examine international variation in cyber crime hosting and exposure and test alternative hypotheses about factors behind such international variation. Countries that offer a favorable environment to cyber crime hosting pose an international threat because cyber crime hosted in these countries affects end-users around the world. Countries where computers are highly exposed to cyber crime also pose an international threat because computers in these countries are more likely to become infected and to serve as bots. I study malware exposure (trojans, viruses, and worms) using the Symantec AV telemetry data and network-based attack (exploits, web attacks, and fake applications) exposure and hosting using the Symantec IPS telemetry data. I find that malware is most prevalent in Sub-Saharan Africa because of high computer piracy rates in this region. End-users in this poor region mainly use pirated software and music purchased from street merchants. These pirated products have a high tendency to contain malware. On the other hand, I find that web attacks and fake applications are most prevalent in North America and Western Europe and that exploits are most prevalent in countries with emergent economies. My regression analysis reveals that cyber criminals target these countries in order to take advantage of the abundant economic and computing resources in these countries. Finally, I find that many Eastern European countries host disproportionate quantities of network-based attacks because of a combination of reasonable computing infrastructure and widespread corruption in these countries. Corrupt law officials and internet service providers tend to turn a blind eye on malicious hosting.

The findings of my analysis provide an opportunity to fight cyber crime using a novel approach that is complementary to the traditional approach of improving cyber defenses. My analysis indicates that addressing corruption in Eastern Europe or computer piracy in Sub-Saharan Africa could reduce the global cyber crime. As an analogy, the approach this thesis suggests is similar to fighting malaria by modifying and eliminating aquatic habitats where mosquitoes breed. Modifying and eliminating these aquatic habitats was critical in eradicating malaria from parts of Israel, the United States, and Italy [75]. In other words, making the environment less favorable to a threatening activity (mosquito breeding) helped eradicate the threat of malaria in these countries. Such approach is still commonly used today in countries affected by malaria [154].

Addressing corruption in Eastern Europe and computer piracy in Sub-Saharan Africa is challenging. I believe that this is best done using a soft power approach [108] that persuades and attracts parties into collaborating rather than a hard power approach that relies on sanctions or military force to force parties to collaborate. For example, Eastern European countries should understand that cracking downing on malicious hosting and corrupt individuals that permit such hosting is in their interest in the long term. If malicious activities in these countries continue and grow, entire blocs of IP addresses from these countries may become blocked and honest users and businesses may become virtually disconnected from parts of the Internet. This has happened in West Africa because of massive scam activities from that region [17]. Similarly, a win-win solution could be found if the software industry adjusts prices to countries' income level. The software industry could collect some income instead of loosing all potential income to piracy. At the same time, users in these countries can benefit from safe and reliable software. Price adjustment is already common for medicines and textbooks.

In the second part of the thesis, I develop two methodologies: one that identifies countries that could pose a biological weapon threat and one that identifies countries that could pose a cyber weapon threat. The two methodologies leverage the fact that the strength of countries' weapon capabilities depend on countries' motivations for these weapons and countries' latent abilities to acquire these weapons. In order to assess countries' motivations for biological (cyber) weapons, I adapt the Friedkin socio-cultural model to capture factors that motivate countries to acquire these weapons and set the parameters of that adapted model using publicly available data. The main factor that motivates countries to develop biological weapons is deterrence of nuclear weapon and biological weapon enemy. On the other hand, the main factor that motivates countries to develop cyber weapons is having an ally that develops these weapons. In order to assess countries' latent biological weapon abilities, I examine conventional arms purchase from biological weapon countries, dual-use biological trade, and countries' military power. In order to assess countries' latent cyber weapon capabilities, I examine countries' cyber security research, cyber security institutions, and general IT preparedness.

The methodologies I developed can be used as a prediction and a policy planning tool. I assess the prediction accuracy of my biological weapon assessment methodology by comparing the methodology's predictions against historical data. Surprisingly, the methodology has very high accuracy despite the high secrecy surrounding biological weapons and the fact that I only use publicly available data to set the methodology's parameters. The reason behind such surprising finding is that countries can conceal their weapon programs and purchases, but are unable to conceal their international relations. These international relations drive many of countries' decisions and are public information. Countries' hostilities affect the risk perceived by these countries and their motivations to pursue biological weapons. Similarly, countries' alliances provide assistance opportunities. Validating the prediction ability of my cyber weapon capabilities assessment methodology by comparing my assessment to historical data is the subject of future work.

I find that countries that could pose both a cyber and a biological weapon threat are China, India, Iran, Israel, and Russia. These countries are typically interested in developing all types of weapon capabilities, and not just cyber and biological weapons. Countries that could pose a cyber weapon threat but not a biological weapon threat are mainly developed countries such as the United States, South Korea, France, Germany, the United Kingdom, Australia, and Canada. On the other hand, countries that could pose a biological weapon threat but not a cyber weapon threat are primarily developing countries that are (or were at some point) isolated internationally such as Algeria and Syria. The difference in the profile of countries that invest in cyber weapons and countries that invest in biological weapons is mainly due to two reasons: 1) acquiring cyber weapons requires significant expertise whereas acquiring biological weapons does not and 2) developing cyber weapons is not prohibited by international law whereas acquiring biological weapons is prohibited by the Biological Weapon Convention (BWC). In the future, I expect a massive cyber weapon arms race. Cyber weapons are attractive because such weapons can achieve their goals remotely and without incurring casualties. Given the interest of powerful countries in these weapons, I do not expect these powerful countries to take action to prohibit the development of these weapons. Instead, I expect many countries to invest in cyber weapons to keep up with other countries and with non-state actors.

I believe that developing computational models that capture expert opinions is a paradigm that can benefit many problems in political science beyond biological weapon and cyber weapon proliferation. For example, it is possible to develop a computational to predict nuclear proliferation by capturing expert opinions about factors that motivate countries to develop nuclear weapons. Another example is to develop computational models to predict the occurrence of inter-state conflicts by capturing theories [23] about factors that cause such conflicts.

Cyber crime vs. cyber weapons. Cyber security research usually does not distinguish between cyber crime and cyber weapons. In this thesis, however, Chapters 2 and 3 on cyber crime focus on cyber attacks on end-users that can be detected by security products, while Chapter 4 focuses on cyber attack capabilities developed by governments. It is worth noting that such distinction is not perfect. For example, many cyber crime attacks are undetected by commercial products. Another example is that governments may also launch cyber attacks that are considered crime, and not an act of war. Having said that, the distinction I make in this thesis helps differentiate between attacks that have different perpetrators and sophistication levels *in general*.

I find that while some country-level social and technical factors have an important impact on both cyber crime and cyber weapons, other factors are relevant for one type of attacks, but not the other. Countries' information technology sophistication plays an important for both cyber crime and cyber weapons. Countries with higher information technology sophistication are more likely to encounter and to host network-based cyber crime attacks. Such countries are also more likely to develop cyber warfare capabilities. On the other hand, countries leverage their cyber security research and institutions when building cyber weapon capabilities. Unfortunately, I did not find evidence that cyber security and institutions reduce cyber crime exposure or hosting. Finally, international relations have an important effect on the likelihood that countries develop cyber weapons, but no significant effect on cyber crime exposure and hosting. More specifically, having a military ally with cyber weapon capabilities increases the likelihood that a country develops cyber weapon capabilities. On the other hand, military alliances, military hostilities, and extradition treaties have no significant effect on countries' cyber crime exposure and hosting.

Russia Russia is an active ground for both cyber crime and cyber weapons. In terms of cyber crime, Russia provides a favorable environment to cyber crime hosting. For example, the Russian Business Network (RBN) was a major malicious Internet Service Provider that hosted up disproportionate quantities of cyber crime (up to 60% of worldwide cyber crime according to some estimates). The Russian government also has cyber warfare capabilities and has allegedly launched massive Denial of Service attacks on Estonia and Georgia.

In Russia, cyber crime and cyber weapons are probably not disconnected. The government likely relies on the extremely talented local hackers when interested in developing or launching cyber weapons. In return, these hackers likely get a free pass to carry on their criminal activities as long as these hackers do not attack local targets.

Appendix A Attack Catalogs

In this appendix, I explain how I build and populate the IPS and AV catalogs that I use to complement the Symantec IPS and AV telemetry data respectively. I build and populate these catalogs by processing the Symantec IPS and AV threat description corpora [141, 142]. The corpora consist of semi-structured descriptions provided by Symantec, a major security vendor providing end-host security products. My catalogs consist of structured descriptions that are easy to analyze using automated techniques. More specifically, I build the AV catalog based on the anti virus (AV) corpus [141] and the Intrusion Prevention System (IPS) catalog based on the IPS corpus [142]. The AV corpus contains descriptions of threats detected by Symantec's commercial AV and the IPS corpus contains descriptions of attacks detected by Symantec's commercial IPS. The AV and IPS are two end-host systems that often run side-by-side, but do not interact. The AV examines the end-host's files, whereas the IPS examines the end-host's network activity. The AV corpus contains more than 12,400 threat descriptions and the IPS corpus contains more than 2,700 attack descriptions.

The AV catalog contains, for a given threat, the threat name, threat family name, type, discovery date, signature release date and severity measures. Threat severity measures [138] are distribution, damage, threat containment and removal levels. The values of these measures belong to three categories: high, medium and low. The IPS contains, for a given attack, the attack name, attack family name and type. The AV catalog contains more attributes than the IPS catalog because the AV corpus contains more detailed information than the IPS corpus.

Symantec's threat corpora contain rich information, but extracting information from these corpora to populate our catalogs is challenging. Symantec provides these corpora in order to help threat victims remove these threats from their machines. Symantec does not provide these corpora in order to help researchers perform a systematic analysis. Threat descriptions are created by different people and contain large amounts of unstructured text. While I used a semi-automated process to extract some attributes from these descriptions, I had to rely on a semi-manual process to extract other attributes. I also often contacted Symantec to inquire about their internal conventions, and to validate our approaches.

I describe the AV catalog in Section A.1 and the IPS catalog in Section A.2.

A.1 AV Catalog

I build and populate the AV catalog based on information available in the Symantec AV online corpus [141]. In order to extract information from the corpus, I mainly look for relevant keywords in the threat descriptions. Table A.1 presents two examples of AV catalog entries.

A.1.1 Catalog Attributes

The catalog contains the following attributes:

Threat name. The unique name Symantec gives to the threat.

Table A.1: Examples of AV catalog entries

Entry	Field	Value
1	Threat name	W32.Aimdes.A@mm
	Specific/Generic	Specific
	Threat family name	Aimdes
	Type	Worm
	Discovered	February 11, 2005
	Initial rapid release	February 11, 2005
	Initial daily certified version	February 11, 2005
	Discovery year	2005
	Distribution	High
	Damage	Medium
	Threat containment	Easy
	Removal	Moderate
2	Threat name	Backdoor.Trojan
	Specific/Generic	Generic
	Threat family name	not available
	Type	backdoor, trojan
	Discovered	February 11, 1999
	Initial rapid release	February 11, 1999
	Initial daily certified version	February 11, 1999
	Discovery year	1999
	Distribution	Low
	Damage	Medium
	Threat containment	Easy
	Removal	Easy

Specific/Generic. This variable indicates whether this is a specific or a generic threat. A specific threat is a particular threat variant, whereas a generic threat may correspond to multiple threat variants that either belong to the same family or that have a common characteristic such as the packer software. The first entry in Table A.1 corresponds to a specific threat, whereas the second entry corresponds to a generic threat.

Threat family name. A generalized name that I derive from the threat name. I am always able to extract a threat family name for specific threats. On the other hand, I am mostly unable to associate a threat family name with generic threats.

Type. The type such as worm and virus. Some threats have more than one type. The main types are trojan, worm, virus, macro, adware/spyware and fake application.

Discovered. The date Symantec finds out about the threat.

Initial Rapid Release. The date the rapid release virus definition is released. Rapid release signatures are subject to basic testing before their release. These signatures defend against newly emerging threats, but are more susceptible to false positives [139].

Initial Daily Certified Version. The date the threat definition is included in the daily release. Threat signatures undergo thorough testing before being included in the daily release [139].

Discovery year. The year the threat is discovered. In most cases, the discovery year is simply the year in the discovered attribute. However, when the discovered attribute is missing, I use the year in the initial rapid release or the initial daily certified version attributes when available.

Table A.2: AV. Guidelines on threat distribution and damage measures [138]

	Distribution
High	Some worms, network aware executables, uncontainable threats (due to high virus complexity or low AV ability to combat)
Medium	Most viruses, some worms
Low	Most trojans
	Damage
High	File destruction/modification, very high server traffic, large scale non repairable damage, large security breaches, destructive triggers
Medium	Non critical settings altered, buggy routines, easily repairable damage, non destructive triggers
Low	No intentionally destructive behavior

Distribution level. A measure of the aggressiveness of the threat propagation mechanism [138]. There are three distribution levels: high, medium and low according to guidelines in Table A.2. It is worth noting that the distribution level is not always a perfect indicator of the number of computers infected by the threat. This is due to multiple reasons such as the fact that the threat may target an old software vulnerability that is currently patched in most systems.

Damage level. A measure of the damage that an infection is capable of causing [138]. There are also three damage levels: high, medium and low as described in Table A.2.

Threat containment. A measure of the difficulty to contain the threat.

Removal. A measure of the difficulty to remove the threat from a machine.

Threat containment. This measure takes 3 possible values: easy, moderate and difficult.

Removal. This measure also takes 3 possible values: easy, moderate and difficult. Easy removal may only require running a full system scan and deleting detected malicious files. Difficult removal may require starting the machine in trouble shooting mode and following detailed instructions.

The AV corpus also contains the number of infections, number of sites and wild level, but I choose not to include these severity measures in the catalog because drawing conclusions based on these measures is difficult. These measures vary over time, and are reported at different times and different life stages for different threats.

A.1.2 Attribute Extraction Methodology

I automatically extract the values of some attributes because these values immediately follow a fixed keyword in threat descriptions. These attributes are the threat name, type, initial rapid release, initial daily certified version, discovered, distribution level, damage level, threat containment and removal. Later, I remove inconsistencies from these values. More specifically, I merge the types “trojan horse” and “trojan” into “trojan”. I also merge the types “adware”, “spyware” and “trackware” into “adware/spyware”.

For threats that do not have the threat type listed as an attribute in the threat description, I leverage Symantec’s virus naming conventions [140] in order to identify the type. Leveraging naming conventions allows me to determine the type of about 5% of the threats. I determine that the type is "worm" if the name contains "worm" or ends with "@m" or "@mm", that the type is "adware/spyware" if the name contains "adware", "spyware" or "infostealer", and that the type is "trojan" if the name contains "trojan" or "backdoor". I decide to assign type "trojan" when the threat name contains "backdoor" because in more than 98% of the cases when the threat name contains "backdoor" and the threat type is given in the corpus, the threat type is "trojan". I do not use the other naming conventions because these conventions do not match the usage in the AV corpus, probably because the convention description is imprecise. For example, one convention says that a "W32" prefix indicates that the threat is a virus that can infect Windows 32 platforms. However, worms are the majority of threats that have a "W32" prefix and that have a type listed in the corpus. The "W32" prefix probably indicates the platform targeted by the threat, but not necessarily the threat type.

The specific/generic attribute is not explicitly given in the corpus, but can be inferred from the threat name or description. I consider a threat to be generic when the threat name ends with "!gen" or "Family" [140], or if the description contains one of the keywords "generic signature", "detection name", "detection technology", "cloud-based detection", "heuristic", "without traditional signatures", and "new malware threats". I determine the keywords to search for by reading a large number of threat descriptions, and manually extracting the relevant keywords.

I extract the threat family name from the threat name by taking advantage of the threat name structure. According to Symantec's virus naming conventions [140], a threat variant name consists of a prefix that designates the platform targeted by the threat or the threat type, the family name and a suffix that differentiates among variants of the family name. I automatically search for and remove the prefixes and suffixes listed in the naming convention. I manually review and correct the resulting family names as some threat variant names do not strictly follow the naming convention.

A.2 IPS Catalog

I build and populate the IPS catalog based on descriptions in the IPS corpus [142]. The IPS corpus contains descriptions of more than 2,700 attacks. The majority of attacks in the IPS corpus are attempts to exploit software vulnerabilities. However, in the IPS corpus, there are also threats such as worms and trojans. Some of these threats may also appear in the AV corpus.

Descriptions in the IPS corpus are less rich than descriptions in the AV corpus, and therefore the IPS catalog contains fewer attributes than the AV catalog. Table A.3 presents an example of an IPS catalog entry.

A.2.1 Catalog Attributes

The catalog contains the following attributes:

Attack name. The unique name given by Symantec to the attack. The name may contain the CVE code of the corresponding exploit [43] as in the first entry in Table A.3, or the name of the corresponding threat as in the second entry in the table.

Family name. This is a generalization of the attack name. I obtain the family name by removing the exploit variant identifier as in the first entry in Table A.3, or by extracting the threat family

Table A.3: Examples of IPS catalog entries

Entry	Field	Value
1	Attack name	Attack: Apache Struts CVE-2013-2251 Code Execution 2
	Family name	Attack: Apache Struts CVE-2013-2251 Code Execution
	Type	Exploit
	Severity	High
2	Attack name	BugBear B Worm FileShare Propagation
	Family name	BugBear
	Type	Worm
	Severity	High

Table A.4: IPS. Guidelines on attack severity levels

Level	Interpretation
High	- Widespread worms or viruses - Or arbitrary code execution as superuser - Or high-impact Denial of Service - Or some backdoors
Medium	- Arbitrary code execution not as superuser - Or write access to important or arbitrary data - Or medium-impact Denial of Service - Or some backdoors - Or most invasive scanning tools
Low	- Reconnaissance tools - Or policy violation such as P2P networks and instant messenger - Or troubleshooting signatures - Or authorized activity

name as in the second entry in the table.

Type. The attack type. Some attacks have more than one type. The main types are web attack, exploit, worm, adware/spyware, trojan, backdoor and OS exploit.

Severity. The severity level of the attack is assigned based on a combination of the attack prevalence among users, and the potential malicious impact of the attack. The Symantec guidelines for assigning the severity level, used for informational purposes, are given in Table A.4.

A.2.2 Attribute Extraction Methodology

Attack name. I extract this name automatically as it is the title of the online attack description.

Family name. I extract this name manually from the attack name or description. Although IPS attack names follow some patterns, these names lack a precise structure. It is worth noting that I am unable to associate an accurate family name with exploits, OS exploits and web attacks. For these attacks, the family name is equal to the signature name after removing the variant identifier as illustrated in the first entry in Table A.3.

Type. I extract the type based on keywords that appear in the attack name or description. More specifically, we assign type “worm”, “virus”, “rootkit”, “trojan” or “backdoor” when these exact keywords appear in the attack name or description. I assign the type “exploit” when the attack name starts with “Attack:” [1], and type “OS exploit” when the attack name starts with “OS Attack:”. We assign type “adware/spyware” when the attack name or description contains “adware” or “spyware”. We assign type “web attack” when the attack name starts with “Web Attack:” or contains “MS IE” or “MSIE” (for Microsoft Internet Explorer). I also assign type “Web Attack” when the attack name contains “HTTP” and is not already assigned another type such as worms or adware. Some worm or adware attack names contain “HTTP” because machines infected with these threats communicate over HTTP.

Severity. I extract this attribute automatically as it always follows the keyword “severity” in the descriptions.

[1] I choose to assign type “exploit” instead of assigning type “attack” because assigning type “attack” would be confusing given that all entries in the IPS corpus are considered attacks.

Bibliography

[1] Sally Adee. The hunt for the kill switch. *IEEE Spectrum*, May 2008. 6.1

[2] Akamai. Akamai's state of the internet report, Q1 2014. 2.2, 3.1, 3.2

[3] Ken Alibek and Stephen Handelman. *Biohazard: The chilling true story of the largest covert biological weapons in the world - Told from the inside by the man who ran it*. Dell publishing, 1999. 5.1, 6.1

[4] Ross Anderson, Chris Barton, Rainer Bhme, Richard Clayton, Michel J. G. van Eeten, Michael levi, Tyler Moore, and Stefan Savage. Measuring the Cost of Cybercrime. In *Workshop on the Economics of Information Security*, pages 1–31, Berlin, Germany, June 2012. 1

[5] Kallol Bagchi, Peeter Kirs, and Robert Cerveny. Global software piracy: can economic factors alone explain the trend? *Communications of the ACM*, 49(6):70–76, June 2006. ISSN 00010782. doi: 10.1145/1132469.1132470. URL `http://portal.acm.org/citation.cfm?doid=1132469.1132470`. 2.7

[6] Michael Bailey, Jon Oberheide, Jon Anderen, Z. Morley Mao, Farnam Jahanian, and Jose Nezario. Automated classification and analysis of internet malware. In *International Symposium on Research in Attacks, Instrusions and Defenses (RAID)*, September 2007. 2.5, 3.5

[7] Reat Bayer. Diplomatic exchange data set, v2006.1. http://correlatesofwar.org, 2006. 5.4.2

[8] Ulrich Bayer, Paolo Milani Comparetti, Clemens Hlauschek, Christopher Kruegel, and Engin Kirda. Scalable, behavior-based malware clustering. In *Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, February 2009. 2.5, 3.5

[9] Gary S. Becker and George J. Stigler. Law enforcement, malfeasance, and compensation of enforcers. *The Journal of Legal Studies*, 3(1): 1–18, January 1974. 3.3.3, 3.10

[10] Joseph Bermudez Jr. The democratic People's Republic of Korea and unconventional weapons. In Peter Lavoy, Scott Sagan, and James Wirtz, editors, *Planning the Unthinkable. How New Powers Will Use Nuclear, Biological, and Chemical Weapons*. Cornell University Press, Ithaca and London, 2000. 1, 5.1, 5.2.1, 6.1

[11] Richard K. Betts. Paranoids, pygmies, pariahs and nonproliferation revisited. In *The proliferation Puzzle: Why nuclear weapons spread (and what results)*. F. Cass, London, England; Portland, OR, 1993. 4.2.1

[12] Leyla Bilge and Tudor 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. 2.5, 3.5

[13] Charles G. Billo and Welton Chang. Cyber warfare. an analysis of the means and motivations of selected nation states. Technical report, Institute for Security Technology Studies at Darmouth College, 2004. 4.1

[14] David Bizeul. Russian business network study. `http://www.bizeul.org/files/RBN_study.pdf`, 2007. 3.2, 3.10

[15] BIZTECH AFRICA. Nigeria at the mercy of software pirates. `http://www.biztechafrica.com/article/nigeria-mercy-software-pirates` 2012. 2.5, 2.7

[16] Krekel Bryan. Capability of the people's republic of china to conduct cyber warfare and computer network exploitation. Technical report, Northrop Grumman Corporation, October 2009. 4.1, 6.1

[17] Jenna Burrell. *Invisible users: youth in the Internet cafes of urban Ghana*. Acting with technology. MIT Press, Cambridge, Mass, 2012. 3.7.1, 3.10, 7

[18] Business Software Alliance. 2010 piracy study. Technical report, May 2011. 2.4.2, 2.7

[19] Juan Caballero, Chris Grier, Christian Kreibich, and Vern Paxson. Measuring pay-per-install: The commoditization of malware distribution. In *The 20th USENIX Security Symposium*, San Francisco, CA, August 2011. 2.1, 2.2, 2.3, 2.6, 2.9, 3.1, 3.3.1, 3.3.2

[20] Davide Canali, Leyla Bilge, and Davide Balzarotti. On the effectiveness of risk prediction based on users browsing behavior. In *ACM symposium on Information, computer and communications security (ASIA CCS)*, pages 171–182. ACM Press, 2014. 2.1, 2.2, 2.3, 2.4.2, 3.1, 3.2, 3.3.2, 3.4.3

[21] Julio Canto, Marc Dacier, Engin Kirda, and Corrado 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. 2.5, 3.5

[22] Jeffrey Carr. *Inside Cyber Warfare. Mapping the Cyber Underworld.* O'reilley, 2nd edition edition, 2012. 1, 4.1, 4.1, 6.1

[23] Greg Cashman. *What causes war?: an introduction to theories of international conflict.* Rowman & Littlefield, Lanham, Maryland, second edition edition, 2014. ISBN 9780742566507 9780742566514. 7

[24] Center for International Development and Conflict Management. International crisis behavior project. `http://www.cidcm.umd.edu/icb/`. Last accessed: December 2011. 2.4.2, 3.4.3

[25] Center for International Development and Conflict Management. International crisis behavior project. http://www.cidcm.umd.edu/icb/, 2010. Last accessed: December 2011. 4.3.1, 5.4.2, 5.4.2

[26] Central Intelligence Agency. The World Factbook. `https://www.cia.gov/library/publications/the-world-factbook/`. Last accessed: March 2012. 2.4.2, 2.9, 3.4.3

[27] Central Intelligence Agency. INTelligence: Open source intelligence. `https://www.cia.gov/news-information/featured-story-archive/20`

2010. 4.7

[28] CERT. National computer security incident response teams. `http://www.cert.org/csirts/national/contact.html`, 2014. Last accessed: January 2014. 2.4.2, 3.4.3, 4.5

[29] Juin-jen Chang, Ching-chong Lai, and C.C. Yang. Casual police corruption and the economics of crime: Further results. *International Review of Law and Economics*, 20(1):35–51, March 2000. 3.3.3, 3.10

[30] T M Chen and S Abu-Nimeh. Lessons from Stuxnet. *Computer*, 44(4):91–93, April 2011. 4.1

[31] Marie Isabelle Chevrier. Impediment to proliferation? Analyzing the Biological Weapons Convention. *Comtemporary Security Policy*, 16: 77–102, 1995. 5.1

[32] Carmen-Cristina Cirlig. Cyber defense in the EU. Preparing for cyber warfare? Technical report, European Parliamentary Research Service, October 2014. URL `http://www.europarl.europa.eu/EPRS/EPRS-Briefing-542143-Cyber-defence-in-the-EU-FINAL.pdf`. 4.2.1, 4.2.1, 4.1

[33] Richard A. Clarke and Robert Knake. *Cyber War: The Next Threat to National Security and What to Do About It*. Harper Collins, 2010. 4.2.1

[34] cnet. EU increases penalties for cybercriminals and hackers. `http://www.cnet.com/news/eu-increases-penalties-for-cybercriminals-a` 2013. 3.10

[35] Avner Cohen. Israel and chemical/biological weapons: History, deterrence, and arms control. *The Nonproliferation Review*, 8(3):27–53, September 2001. 5.4.2

[36] Correlates of War. Alliances v4.1. http://www.correlatesofwar.org/. Last accessed: December 2011. 4.3.1, 5.4.2

[37] Correlates of War. Direct contiguity data, 1816-2006. version 3.1. http://correlatesofwar.org, 2006. 5.4.2

[38] Correlates of War. National material capabilities (v4.0). http://www.correlatesofwar.org/, 2010. 5.5.2

[39] Correlates of War. State system membership list, v2011. http://correlatesofwar.org, 2011. 5.4.2

[40] Correlates of War Project. Alliances v3.03. `http://www.correlatesofwar.org/`. Last accessed: December 2011. 2.4.2, 3.4.3

[41] Marco Cova, Corrado Leita, Olivier Thonnard, Angelos D. Keromytis, and Marc Dacier. An analysis of rogue av compaigns. In *International Symposium on Research in Attacks, Instrusions and Defenses (RAID)*, pages 442–463, Ottawa, Ontario, Canada, September 2010. 3.2, 3.3.3

[42] Eric Croddy. China's role in the chemical and biological disarmament regimes. *The Nonproliferation Review*, 9(1):16–47, March 2002. 5.2.2

[43] CVE-Common Vulnerabilities and Exposures. CVE-Common Vulnerabilities and Exposures. `http://cve.mitre.org/`. Last accessed: October 2012. A.2.1

[44] Zachary S Davis. The realist nuclear regime. In *The proliferation Puzzle: Why nuclear weapons spread (and what results)*. F. Cass, London, England; Portland, OR, 1993. 4.2.1

[45] Bruce Bueno de Mesquita. Decision-making models, rigor and new puzzles. *European Union Politics*, 5:125–138, 2004. 4.4.1, 5.2.3

[46] Bruce Bueno de Mesquita and Frans Stockman. *European Community decision-making: models, applications and comparisons*. University Press, New Haven, Yale, 1994. 4.1, 4.4.1, 5.2.3

[47] Dorothy Denning. Reflections on cyberweapons controls. *Computer Security Journal*, XVI(4):43–53, 2000. 4.2.1

[48] Department of Peace and Conflict Research. Uppsala University. Ucdp dyadic dataset. `http://www.pcr.uu.se/research/ucdp/datasets/ucdp_dya` Last accessed: December 2011. 2.4.2, 3.4.3

[49] Zuhair Diab. Syria's chemical and biological weapons: Assessing capabilities and motivations. *The Nonproliferation review*, 5:104–111, 1997. 5.1, 6.1

[50] Tudor Dumitras and Darren Shou. Toward a standard benchmark for computer security research. The worldwide intelligence network environment (wine). In *Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, Salzburg, Austria, April 2011. 1, 2.1, 2.4.1, 3.1

[51] David Elliott. Deterring strategic cyberattack. *IEEE Security and Privacy*, 5(9):36–40, October 2011. 4.2.1

[52] European Center for Constitutional and Human Rights. UK rebukes german-british software company Gamma. `http://www.ecchr.de/surveillance-technology/articles/human-rights-organisations-file-oecd-complaints-against` 2015. 4.7

[53] William Frankenstein, Ghita Mezzour, Kathleen M. Carley, and L. Richard Carley. Remote assessment of countries nuclear, biological, and cyber capabilities: joint motivation and latent capability approach. *Social Network Analysis and Mining*, 5(1), December 2015. 4.1, 6.1

[54] Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35, March 1977. ISSN 00380431. 2.4.2

[55] Noah Friedkin and Eugene Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15:193–205, 1990. 1, 4.1, 4.4.1, 5.1, 5.2.3

[56] Erik Gartzke. The Capitalist Peace. *American Journal of Political Science*, 51(1):166–191, January 2007. 3.8.3, 4.4.2, 4.7

[57] Douglas M. Gibler. *International military alliances, 1648-2008*. Correlates of war series. CQ Press, Washington, D.C, 2009. ISBN 9781568028248. 4.3.1, 5.4.2

[58] Keir Giles. Information troops - a russian cyber command? In *3rd International Conference on Cyber Conflict*, Tallinn, Estonia, June 2011. 4.1, 6.1

[59] Chris Grier, Lucas Ballard, Juan Caballero, Neha Chachra, Christian J. Dietrich, Kirill Levchenko, Panayiotis Mavrommatis, Damon McCoy, Antonio Nappa, Andreas Pitsillidis, Niels Provos, M. Zubair Rafique, Moheeb Abu Rajab, Christian Rossow, Kurt Thomas, Vern Paxson, Stefan Savage, and Geoffrey 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. 3.3.1, 3.3.2

[60] Robert Harkavy. The pariah state syndrome. *Orbis*, 21(3):623–650, 1977. 5.4.2

[61] Robert Harkavy. Pariah states and nuclear proliferation. *International Organization*, 35(1):135–163, 1981. 5.4.2

[62] Harmonized System Committee. Classification of the biological dual-use items of the convention on the prohibition of the development, production and stockpiling of bacteriological (biological) and toxin weapons and their destruction. Technical Report NC1264E1a, World Customs Organization, Brussels, Belgium, 2008. 5.5.3

[63] Steven A. Hilderth. Cyberwarfare. Technical report, CRS Report for Congress, June 2001. 4.1, 4.1, 6.1

[64] Michael C. Horowitz and Neil Narang. Poor man's atomic bomb? Exploring the relationship between weapons of mass destruction. *Journal of Conflict Resolution (Forthcoming)*, 2014. 1, 4.3, 4.7, 5.1, 5.2.1, 5.4.1, 5.4.2

[65] Xin Hu, Tzicker Chiueh, and Kang G. Shin. Large-scale malware indexing using function-call graphs. In *Computer and Communication Security Conference (CCS)*, Chicago, IL, November 2009. 2.5, 3.5

[66] International Cyber Center. George Mason University. Certicc home. `http://internationalcybercenter.org/certicc`, 2014. Last accessed: January 2014. 2.4.2, 3.4.3, 4.5

[67] International Telecommunication Union. Measuring the information society. `http://www.itu.int/en/ITU-D/Statistics/Documents/publicati` 2012. 2.4.2, 3.4.3, 4.3.1, 4.3.1, 4.5

[68] James Martin Center for Non Proliferation Studies. Chemical and biological weapons: Possession and programs past and present. http://cns.miis.edu/cbw/possess.html, 2008. 5.4.2

[69] D.-J. Jo and E. Gartzke. Determinants of nuclear weapons proliferation. *Journal of Conflict Resolution*, 51(1):167–194, February 2007. 4.2.1, 5.4.2, 5.4.2, 5.4.2

[70] Wolfgang John and Sven Tafvelin. Analysis of internet backbone traffic and header anomalies observed. In *ACM Internet Measurement Conference (IMC)*, San Diego, CA, August 2007. 3.4.2

[71] Benjamin Johnson, John Chuang, Jens Grossklags, and Nicolas 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. 3.2

[72] Andrew J. Kalafut, Craig A. Shue, and Minaxi Gupta. Malicious hubs: Detecting abnormally malicious autonomous systems. In *Proceedings of the 29th Conference on Information Communications (INFOCOM)*, San Diego, CA, March 2010. 3.2

[73] Markus Kammerstetter, Christian Platzer, and Gilbert Wondracek. Vanity, cracks and malware: insights into the anti-copy protection ecosystem. In *Computer and Communication Security Conference (CCS)*, 2012. 2.3

[74] Richard J. Kilroy. The U.S. Military Response to Cyber Warfare. In Lech Janczewski and Andrew M. Colarik, editors, *Cyber warfare and cyber terrorism*. Information Science Reference, Hershey, 2008. ISBN 1591409918. 4.1

[75] U. Kitron and A. Spielman. Suppression of Transmission of Malaria Through Source Reduction: Antianopheline Measures Applied in Israel, the United States, and Italy. *Clinical Infectious Diseases*, 11(3):391–406, May 1989. 7

[76] Gregory Koblentz. Pathogens as weapons. The international security implications of biological warfare. *International security*, 28, April 2003-2004. 1, 5.1, 5.2.1, 5.4.1

[77] Gregory Koblentz. Predicting peril or the peril of prediction? assessing the risk of CBRN terrorism. *Terrorism and Political Violence*, 23: 501–520, 2011. 5.8

[78] David Krackhardt. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10(5):359–381, December 1988. 3.3.5

[79] Matthew Kroenig. *Exporting the bomb technology transfer and the spread of nuclear weapons*. Cornell University Press, Ithaca, 2010. 4.2.1

[80] Maurice Kugler, Thierry Verdier, and Yves Zenou. Organized crime, corruption and punishment. *Journal of Public Economics*, 89(9-10): 1639–1663, September 2005. 3.3.3, 3.10

[81] Anthony Lake. Confronting backlash states. *Foreign Affairs*, 73(2):45–55, 1994. 5.4.2

[82] Fanny Lalonde Levesque, Jude Nsiempba, Jos M. Fernandez, Sonia Chiasson, and Anil Somayaji. A clinical study of risk factors related to malware infections. In *ACM SIGSAC conference on Computer and communications security (CCS)*, pages 97–108, Berlin, Germany, November 2013. ACM Press. 2.1, 2.2, 2.3, 3.1, 3.2, 3.3.2

[83] Carl Landwehr. Cybersecurity: From engineering to science. *The Next Wave. The National Security Agency's review of Emerging Technologies*, 19(2), 2012. 3.1

[84] Ralph Langner. Stuxnet: Dissecting a Cyberwarfare Weapon. *IEEE Security & Privacy Magazine*, 9(3):49–51, May 2011. 4.1

[85] Milton Leitenberg. Biological weapons in the twentieth century: A review and analysis. *Critical Reviews in Microbiology*, 27(4):267–320, 2001. 5.4.2

[86] James Andrew Lewis and Katrina Timlin. Cybersecurity and cyberwarfare. Preliminary assessment of national doctrine and organization. Technical report, Center for Strategic and International Studies, 2011. 2.4.2, 3.4.3, 4.5

[87] James Andrew Lewis and Katrina Timlin. Cybersecurity and cyberwarfare. {P}reliminary assessment of national doctrine and organization. Technical report, Center for Strategic and International Studies, 2011. 1, 4.1, 4.1

[88] John Leyden. Germany reveals secret techie soldier unit, new cyberweapons. http://www.theregister.co.uk/2012/06/08/germany_cyber_offe
2012. 4.1

[89] Martin Libicki. Cyberdeterrence and cyberwar. Technical report, Rand, 2009. 4.2.1

[90] Stuart Madnick, Xitong Li, and Nazli 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. 3.3.3

[91] Gregor Maier, Anja Feldmann, Vern Paxson, Robin Sommer, and Matthias Vallentin. An assessment of overt malicious activity manifest in residential networks. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, volume 6739, pages 144–163. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. 2.1, 2.2, 3.1, 3.2

[92] Mandiant. APT1: Exposing one of china's cyber espionage units. Technical report, February 2013. 4.1, 6.1

[93] Z. Maoz. Structural Equivalence and International Conflict: A Social Networks Analysis. *Journal of Conflict Resolution*, 50(5):664–689, October 2006. 3.8.3, 4.4.2

[94] Sugata Marjit and Heling Shi. On controlling crime with corrupt officials. *Journal of Economic Behavior & Organization*, 34(1):163–172, January 1998. 3.3.3, 3.10

[95] Susan Martin. The role of biological weapons in international politics: The real military revolution. *Journal of strategic studies*, 25:63–98, 2002. 5.1

[96] Maxmind. Geolite free downloadable databases. Geolite country. http://dev.maxmind.com/geoip/legacy/geolite/, November 2012. 2.4.1, 3.4.2

[97] McAfree. Mcafee labs threats report. http://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q1-2014.pdf, June 2014. 2.1, 2.2, 3.1, 3.2

[98] Mike McConnel. America's cyber future. security and prosperity in the information age. Technical report, 2011. 6.1

[99] Harvey J. McGeorge. Chemical addiction. *Defense and Foreign Affairs*, April 1989. 5.7.1

[100] Ghita Mezzour, Kathleen M. Carley, and L. Richard Carley. An empirical study of global malware encounters. In *Symposium and Bootcamp on the Science of Security*, Urbana, IL, April 2015. 2.1

[101] Microsoft. Windows 8.1. http://www.microsoftstore.com/store/msusa/en_US/pdp/Windows-8.1/productID.288401200,2015.htt
2.9

[102] Micrsoft. Micrsoft security intelligence report. Worldwide threat assessment, July-December 2013. 2.2, 3.1, 3.2

[103] PD Millet. The biological and toxin weapons convention. *Revue Scientifique et Technique - Office International Des Epizooties*, 25:35–52, 2006. 5.1

[104] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. Inside the slammer worm. *IEEE Security and Privacy*, 4(1):33–39, July 2003. 2.5, 3.5

[105] F. Narin, D. Olivastro, and K. A. Stevens. Bibliometrics/Theory, Practice and Problems. *Evaluation Review*, 18(1):65–76, February 1994. 3.4.3, 4.5

[106] New York Times. Obama calls for new law to bolster cybersecurity. http://www.nytimes.com/2015/01/14/us/obama-to-announce-new-cybe
2015. 2.3, 3.10

[107] Nuclear Threat Initiative. Vietnam. http://www.nti.org/country-profiles/vietnam/, 2014. 5.2.2

[108] Joseph S. Nye. *Soft power: the means to success in world politics*. Public Affairs, New York, 1st ed edition, 2004. 3.10, 7

[109] Joseph S. Nye. Nuclear Lessons for Cyber Security? *Strategic Studies Quarterly*, 5(4):18–38, 2011. 4.2.1

[110] Bureau of Labor Statistics. Producer price indexes. http://www.bls.gov/ppi/. 5.5.3

[111] Kaan Onarlioglu, Yutku Ozan Yilmaz, Engin Kirda, and Davide Balzarotti. Insights into user behavior in dealing with internet attacks. In *Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, February 2012. 2.1, 2.2, 2.3, 3.1, 3.2, 3.3.2

[112] Javier Ulises Ortiz. Argentina: The challenge of information operation. 2008. 4.1

[113] William O. Owens, Kenneth W. Dam, and Herbert S. Lin. Technology, policy, law, and ethics regarding u.s. acquisition and use of cyberattack capabilities. Technical report, National Research Council (NRC), 2009. 4.2.1, 6.1

[114] Evangelos E. Papalexakis, Tudor Dumitras, Duen Horng (Polo) Chau, B. Aditya Prakash, and Christos Faloutsos. Spatio-temporal mining of software adoption &amp; penetration. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 878–885. ACM Press, 2013. ISBN 9781450322409. 2.4.1

[115] J. R. Platt. Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, 146(3642):347–353, October 1964. 3.1

[116] Niels Provos, Panayiotis Mavrommatis, Moheeb Abu Rajab, and Fabian Monrose. All your iframes point to us. In *17th Usenix Security Symposium*, San Jose, CA, July 2008. 1, 2.3, 3.2, 3.3.1, 3.3.3, 3.5, 3.7.1

[117] George Quester. Chemical and biological warfare. *American Political Science Review*, 68:1285–1291, 1974. 5.1

[118] Konrad Rieck, Thorsten Holz, Carsten Willems, Patrick Düssel, and Pavel 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. 2.5, 3.5

[119] Marco Roscini and Leverhulme Trust. *Cyber operations and the use of force in international law*. Oxford University Press, Oxford, 2014. ISBN 9780199655014. 4.1

[120] Paul Rosenzweig. *Cyber warfare: how conflicts in cyberspace are challenging America and changing the world*. The changing face of war. Praeger, Santa Barbara, Calif, 2013. ISBN 9780313398957. 3.10

[121] Scott Douglas Sagan. *The spread of nuclear weapons: an enduring debate*. W.W. Norton & Co, New York, 3rd ed edition, 2013. ISBN 9780393920109. 4.2.1

[122] David E. Sanger. Obama order sped up wave of cyberattacks against Iran. http://www.nytimes.com/2012/06/01/world/middleeast/obama-ord June 2012. 3.3.2, 4.1

[123] SCOPUS. www.scopus.com. Last accessed: October 2012. 2.4.2, 3.4.3, 4.5

[124] Scott J. Shackelford. From nuclear war to net war: Anologizing cyber attacks in international law. *Berkeley Journal of International Law*, 27 (1):191–250, 2009. 4.2.1

[125] Paulo Shakarian. *Introduction to cyber-warfare: a multidisciplinary approach*. Syngress, Amsterdam ; Boston, 2013. ISBN 9780124078147. 4.3

[126] Amit Sharma. Cyber wars: A paradigm shift from means to ends. *Strategic Analysis*, 34(1), 2010. 6.1

[127] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)*, page 373, Atlanta, GA, April 2010. ACM Press. 2.1, 2.2, 3.1, 3.2

[128] Seung Kyoon Shin, Ram D. Gopal, G. Lawrence Sanders, and Andrew B. Whinston. Global software piracy revisited. *Communications of the ACM*, 47(1):103–107, January 2004. 2.7, 2.9

[129] Adam Shostack. The evolution of information security. *The Next Wave. The National Security Agency's review of Emerging Technologies*, 19(2), 2012. 3.1

[130] J. David Singer. Reconstructing the correlates of war dataset on material capabilities of states, 18161985. *International Interactions*, 14(2): 115–132, May 1988. 5.5.2

[131] J. David Singer and Melvin Small. Formal alliances, 1815-1939. *Journal of Peace Research*, 3(1):1–31, 1966. 4.3.1, 5.4.2

[132] J. David Singer and Melvin Small. Formal alliances, 1816-1965: An extension of the basic data. *Journal of Peace Research*, 6(3):257–282, 1969. 4.3.1, 5.4.2

[133] J. David Singer, Stuart Bremer, and John Stuckey. Capability distribution, uncertainty, and major power war, 1820-1965. In *Peace, War, and Numbers*, pages 19–48. SAGE Publications, Beverly Hills, 1972. 5.5.2

[134] Frank Stajano and Paul Wilson. Understanding scam victims: seven principles for systems security. *Communications of the ACM*, 54(3):70, March 2011. ISSN 00010782. doi: 10.1145/1897852.1897872. 3.7.1

[135] Douglas M. Stinnett, Jaroslav Tir, Philip Sschafer adn Paul F. Diehl, and Charles Gochman. The correlates of war project direct contiguity data, version 3.0. *Conflict Management and Peace Science*, 19(2):59–67, 2002. 5.4.2

[136] Stockholm International Peace Research Institute. SIPRI arms transfers database. http://www.sipri.org/databases/armstransfers, 2013. 5.5.1

[137] Brett Stone-Gross, Christopher Kruegel, Kevin Almeroth, Kevin Almeroth, Andreas Moser, and Engin Kirda. FIRE: Finding rogue networks. In *Annual Computer Security Applications Conference (ACSAC)*, Honolulu, HI, December 2009. 3.2, 3.3.3, 3.10

[138] Symantec. Threat severity assessment. http://www.symantec.com/security_response/severityassessment.jsp, . Last accessed: October 2012. (document), A, A.2, A.1.1, A.1.1

[139] Symantec. Types of virus definitions available for download. http://www.symantec.com/popup.jsp?popupid=sr_help_popup, . Last accessed: October 2012. A.1.1, A.1.1

[140] Symantec. Symantec naming conventions. http://www.symantec.com/security_response/virusnaming.jsp, 2013. Last accessed: June 2013. A.1.2

[141] symantecAV. Symantec threat explorer. http://www.symantec.com/security_response/landing/azlisting.jsp. Last accessed: October 2012. 2.4.1, A, A.1

[142] symantecIPS. Symantec attack signatures. http://www.symantec.com/security_response/attacksignatures/. Last accessed: October 2012. 3.4.1, A, A.2

[143] Bradley A. Thayer. The Causes of Nuclear Proliferation and the Utility of the Nuclear Non-Proliferation Regime. *Security Studies*, 4(3): 463–519, March 1995. 4.2.1

[144] The National Consortium for the Study of Terrorism and Responses to Terrorism. Global terrorism database. http://www.start.umd.edu/gtd/, 2014. Last accessed: May 2014. 5.8

[145] The World Bank. The little data book on information and communication technology. http://data.worldbank.org/products/data-books/little 2011. 3.4.2, 3

[146] The World Bank. World development indicators (wdi) 2012. http://data.worldbank.org/data-catalog/world-development-indicators/ April 2012. 2.4.2, 3.4.2, 3.4.3, 3

[147] Olivier Thonnard, Leyla Bilge, Gavin O'Gorman, Seán Kiernan, and Martin 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. 2.5, 2.7, 3.5, 3.6.2

[148] Jonathan Tucker. Historical trends related to bioterrorism: An empirical analysis. *Emerging and Infectious Diseases*, 5:498–504, 1999. 5.8

[149] Jonathan Tucker. Motivations for and against proliferation: The case of the middle east. In Raymond Zilinkas, editor, *Biological Warfare:*

Modern Offense and Defense. Lynne Rienner Publishers, 2000. 1, 5.2.1, 5.4.1, 5.4.2

[150] United Nations Commodity Trade Statistics. UN comtrade. http://comtrade.un.org/db/, 2011. Last accessed: August 2011. 5.5.3

[151] 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`. 2.4.2, 3.4.3

[152] U.S. Congress, Office of Technology Assessment (OTA). Proliferation of weapons of mass destruction: Assessing the risks. Technical Report OTA-ISC-559, U.S. Government Printing Office, Washington, DC, August 1993. 5.4.2

[153] U.S. Congress, Office of Technology Assessment (OTA). Technologies underlying weapons of mass destruction. Technical Report OTA-BP-ISC-115, U.S. Government Printing Office, Washington, DC, December 1993. 5.1, 5.2.2, 5.4.1

[154] Jurg Utzinger, Yesim Tozan, and Burton H. Singer. Efficacy and cost-effectiveness of environmental management for malaria control. *Tropical Medicine and International Health*, 6(9):677–687, September 2001. 7

[155] Kenneth N Waltz. *Theory of international politics*. Waveland Press, Long Grove, Ill., 2010. ISBN 1577666704 9781577666707. 4.2.1

[156] World Economic Forum. The global competitiveness report. `http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2012-13` 2012-2013. 3.4.3

[157] World Intellectual Property Organization. World intellectual property indicators. `http://www.wipo.int/ipstats/en/wipi/`, 2015. 4.3.1

[158] Ting-Fang Yen, Victor Heorhiadi, Alina Oprea, Michael K. Reiter, and Ari Juels. An epidemiological study of malware encounters in a large enterprise. In *ACM SIGSAC conference on Computer and communications security (CCS)*, pages 1117–1130. ACM Press, 2014. 2.1, 2.2, 2.3, 3.1, 3.2, 3.3.2

[159] Raymond Zilinskas. Iraq's biological weapons. The past as future? *The Journal of American Medical Association (JAMA)*, 278:418–424, 1997. 5.1, 6.1

[160] Raymond Zilinskas. *Biological Warfare: Modern Offense and Defense*. Lynne Rienner, 1999. 5.1