



CPB Netherlands Bureau for Economic
Policy Analysis

CPB Discussion Paper | 306

What drives cybercrime?

Empirical evidence from DDoS attacks

Bastiaan Overvest
Bas Straathof

What drives cybercrime? Empirical evidence from DDoS attacks

Bastiaan Overvest Bas Straathof*

April 24, 2015

Abstract

The internet has become essential for advanced economies and the risk of disruption from cybercrime has increased accordingly. This paper focuses on a common type of cybercrime: Distributed Denial of Service (DDoS) attacks. We propose an economic model of DDoS attacks in which hackers choose the victim of an attack for economic or ideological goals, while the source country is chosen for its size, bandwidth and vulnerability. We use data on the frequency of attacks between the country of origin and the country of destination to estimate a "gravity" equation inspired on the international trade literature. Our results suggest that a ten percent increase in the number of internet users worldwide raises the number of attacks by about eight percent. Bandwidth in the country of origin and economic ties are also significantly related to attacks.

Keywords: Cybercrime, cybersecurity, DDoS attacks, poisson, gravity

JEL codes: L86, F14, F51

*Overvest: CPB Netherlands Bureau for Economic Policy Analysis and extramural fellow at Tilburg Law & Economics Center (TILEC), e-mail: b.overvest@cpb.nl. Straathof: CPB, e-mail: s.m.straathof@cpb.nl. We are grateful to Gaaitzen de Vries for helpful suggestions, to Bas ten Dam, Zhe Li and René Nieuwenhuizen for their excellent research assistance and to Michel van Eeten for a stimulating discussion on this topic. During the project we received valuable advice from many colleagues and from seminar participants at the University of Groningen, the ministry of Economic Affairs and the National Cyber Security Center. All errors and omissions are, naturally, made by evil cybermiscreants.

1 Introduction

The internet has become a critical part of the infrastructure in the past two decades – in particular for economically advanced countries. To illustrate this claim: the number of EU households with broadband access quadrupled between 2004 and 2013 from 15 to 78 percent and over 40 percent of all European citizens bank online.¹

As the dependence on the internet grows, cybersecurity² is becoming vital for the functioning of the economy. The importance of cybersecurity has not gone unnoticed to economists. Acemoglu, Malekian & Ozdaglar (2013), for instance, theoretically examine strategic investments in cybersecurity. And Athey & Stern (2013) empirically consider the international incidence of software piracy.

It is notoriously difficult to estimate the costs of cybercrime, but two recent studies suggest that the costs of cybercrime are substantial. Ponemon Institute (2013) estimates the cost of cybercrime at about \$11 million for the “average” US firm. In a report written for internet security firm McAfee, the Center for Strategic and International Studies (2014) reports an annual cost of \$400 billion for the global economy. Given the uncertainties that surround cybercrime and the associated output loss, both numbers should be seen as indicative.

This paper is the first empirical study on international DDoS attacks. We focus on the incidence of DDoS attacks between and within countries.³ Typically, DDoS attacks are conducted via a botnet, from which a hacker or a “bot herder” launches an attack against a target. By flooding a target website with parallel incoming requests, a DDoS attack tries to make the target unreachable for other, legitimate, users. Banks are often thought of as the main victims of DDoS attacks. The list of (publicly known) victims is actually quite diverse and includes non-governmental organizations, universities and newspapers.

There are several reasons why DDoS attacks are an interesting form of cybercrime. First, DDoS attacks occur frequently and the costs of an attack

¹Source: Eurostat. Codes TIN00089 (broadband access) and TIN00099 (online banking).

²Cybersecurity can be seen as the absence of disruption, failure or abuse of internet services. Cyber threats are acts (or crimes) that reduce the level of cybersecurity. Examples of cybercrime are intellectual property theft through computer hacks, identity theft, viruses, data hostage, and Distributed Denial of Service (DDoS) attacks.

³DDoS attacks are attempts to interrupt or suspend network devices or websites for other internet users. See Zuckerman, Roberts, McGrady, York & Palfrey (2010) for an extensive discussion of the history of DoS and DDoS attacks, various DDoS techniques and mitigation options.

due to lost business and IT-expenditure are substantial. Cyber security firm Kaspersky Lab (2015) reports that one third of firms in financial or public-facing online services have experienced a DDoS attack in the period between April 2013 and May 2014. For small and medium-sized firms the costs of an incident are \$52,000 on average. For larger enterprises the average costs per incident are \$444,000.

A second reason is that large DDoS attacks can have a disruptive impact on the receiving countries by distorting communication in important networks. Recent attacks on banks' websites illustrate this. Third, DDoS attacks use bandwidth and slow internet traffic even when attacks are not successful.⁴

We propose an economic model to explain the frequency of DDoS attacks between any pair of countries. In our model, hackers may deploy a botnet in one country to attack a website in another country. A hacker only executes the attack if the expected returns exceed the costs. The model's main prediction is that the frequency of attacks is proportional to the number of internet users in both countries.

The theoretical framework is used to specify an econometric model. The empirical results suggest that DDoS attacks can be largely explained by the number of internet users in the country of origin and destination, and by bandwidth in the country of origin. Trade is also significantly related to attacks, suggesting that attacks follow economic ties. The vulnerability of computers does not seem influential. Factors that matter for trade like GDP per capita and geographical distance do not appear to determine DDoS attacks.

This study aims to contribute to a small literature on the economics of cybercrime. Closest to our work are Johnson, Laszka, Grossklags, Vasek & Moore (2014) and Vasek, Thornton & Moore (2014). Johnson et al. (2014) offer a game-theoretic perspective on DDoS attacks between Bitcoin mining pools. A Bitcoin mining pool is a collaboration between Bitcoin users to obtain freshly minted Bitcoins. They show that, to raise rivals' costs, a mining pool may launch a DDoS attack on a competing pool. Interestingly, they find that DDoS attacks are more likely to occur between larger mining pools than between smaller mining pools. Vasek et al. (2014) empirically study DDoS attacks on Bitcoin services. They collect data on 142 DDoS attacks on 40 Bitcoin services. Consistent with Johnson et al. (2014), they find that large mining pools are much more likely to be "DDoSed"

⁴In our dataset, we observe about 129 substantial attacks per day. See the DDoS monitoring website <http://map.ipviking.com> for a visualization of current attacks.

than small pools.

Another empirical study on cybersecurity is Van Eeten, Bauer, Asghari & Tabatabaie (2010), who analyze data on spam e-mail. Spam is not only a form of cybercrime itself, but it is also indicative of vulnerability, as spam is typically sent from infected machines. Van Eeten et al. (2010) find that a very large portion of spam (109 billion messages) originates from just 170 unique IP addresses. Moreover, they report that just 50 Internet Service Providers (ISPs) account for 50 percent of all infected machines worldwide.

More recently, Athey & Stern (2013) study the global incidence of illegal Windows software copies. Using cross-country regressions, they find that this case of software piracy decreases in the strength of legal institutions and increases in speed of broadband. They do not find a direct significant effect of GDP per capita.

Earlier papers on cybersecurity, such as Anderson & Moore (2006), often argue that cybersecurity investments are strategic complements. This means that if one internet user invests more, he creates positive externalities for other internet users and indirectly lowers the incentives for others to invest as well. An analogue example is the decision of an airline not to screen luggage transferred from airlines carriers with a strict security policy. Because agents can free ride on the efforts of others, the equilibrium level of security may be too low.

In a recent contribution, Acemoglu et al. (2013) reach a more nuanced conclusion. They study how the decision of interconnected agents to invest in cybersecurity depends on the structure of the network and the type of the cyber threat (a random or a strategic attack). They conclude that security investments may be strategic substitutes, instead of complements, and agents may invest too much compared to the social optimum.

In the next two sections, we present our economic model and explain our empirical strategy. Section 4 discusses the data. The empirical assessment of DDoS attacks can be found in Section 5. Section 6 summarizes the main conclusions.

2 A model of DDoS attacks

This section develops a simple tractable model of DDoS attacks. The focus here is to create a testable framework, and not an comprehensive economic theory. Our basic model can, however, easily be generalized.

There are $h \geq 1$ hackers in the world. Each hacker knows exactly one

computer software vulnerability, also known as a “zero day exploit”. A zero day exploit is a software flaw for which no patch is available (for example because the software developer is not aware of it), such that users have had no opportunity (“zero days”) to respond to the flaw. The zero day exploit of hacker l can infect a fraction c_l of all PCs worldwide. In any country $i = 1, 2, \dots, m$, the number of PCs is N_i and hacker l forms a botnet of $c_l N_i$ PCs. The number of botnets per country equals the number of hackers h .

A botnet with a fast internet connection is a more effective tool to launch an attack than a botnet with low bandwidth. The effective size of a botnet therefore depends on the number of infected PCs and the average internet connection of the infected PCs. If the average bandwidth in country i is denoted by w_i , the effective size of a botnet can be written as $w_i c_l N_i$.

For simplicity, we assume that hackers do not mutually combine botnets and do not attack a country from more than one country.⁵ Hacker l may use the botnet in country i to launch DDoS attacks on any country j . We suppose that botnets do not wear out: each botnet can be used repeatedly for new attacks, with no impact on the strength of the botnet. This allows hackers to subsequently attack websites in all countries.

When hacker l targets a website in country j , the expected utility of the hacker is

$$E[u_{ijl}] = Pr(\text{success}) N_j^\beta G_{ij}^\gamma - e_{il}. \quad (1)$$

If the DDoS attack successfully takes the target website offline, which occurs with a certain probability, the hacker receives a benefit N_j^β where N_j is the number of PCs in country j and β is a scalar. The idea behind this specification is that the hacker’s utility increases in the number of visitors of the target website. We assume that this number is proportional to the number of PCs. The variable G_{ij} indicates country-pair specific factors that may affect the hacker’s payoff. The hacker incurs a fixed cost of effort e_{il} when he attacks. This effort differs across hackers and source countries. We assume that e_{il} is drawn IID from a distribution with cumulative distribution function $F(e)$. Hackers observe the realization of e before they initiate the DDoS attack.

The probability of success is proportional to the effective size of the

⁵This is largely consistent with the empirical evidence. Of all 15,829 DDoS attacks in our raw dataset for which the source country or countries is known, only 824 (or 5 percent) originate from more than one country.

botnet:

$$Pr(success) = \frac{w_i c_l N_i^\alpha}{S_j}, \quad (2)$$

where α is a scalar and S_j is a measure of the relative level of cybersecurity in country j . The larger S_j , the more likely it is that the DDoS attack fails. We assume that S_j is sufficiently large such that the probability of success is strictly below one.

The utility of the hacker when he refrains from attacking country j from country i is zero. Then, hacker l launches an attack if the expected utility is greater than zero. The expected number (or frequency) of DDoS attacks from country i to country j can now be written as:

$$freq_{ij} = \sum_{l=1}^h Pr(E[u_{ijl}] \geq 0) = \sum_{l=1}^h F\left(\frac{w_i c_l N_i^\alpha N_j^\beta G_{ij}^\gamma}{S_j}\right). \quad (3)$$

Equation (3) relates the expected frequency of attacks to the number of PCs in both the country of origin i and the country of destination j . Based on this expression, we expect that DDoS attacks are more frequent between countries that both accommodate a large number of PCs. Additionally, we conjecture that the number of DDoS attacks increases in bandwidth bw_i and decreases in the level of cybersecurity S_j .

3 Estimation strategy

Our identification strategy is to estimate a structural equation based on (3). We make two additional assumptions to facilitate this. First, we impose symmetry on the infection rates: $c_l = c$ for all $l \in \{1, \dots, h\}$. Second, we let the effort levels be drawn from the uniform distribution, with $F(e) = e/b$, where b is an arbitrary but sufficiently large constant. These two assumptions allow us to rewrite equation (3) as

$$freq_{ij} = h \left(\frac{c w_i N_i^\alpha N_j^\beta G_{ij}^\gamma}{b S_j} \right). \quad (4)$$

The logarithm of this equation is linear in coefficients:

$$\ln(freq)_{ij} = \alpha_0 + \ln(w_i) + \alpha \ln(N_i) + \beta \ln(N_j) - \ln(S_j) + \gamma \ln(G_{ij}), \quad (5)$$

where $\alpha_0 \equiv \ln(hc/b)$ is a constant term.

The data set on DDoS attacks contains many zeros (see Section 4) which makes OLS less suitable for estimation. Taking logarithms leads to exclusion of all zero-valued observations and as it is likely that the incidence of zero-valued observations is non-random, OLS can yield biased results (Flowerdew & Aitkin, 1982; Silva & Tenreyro, 2006; Burger, van Oort & Linders, 2009).

We follow Silva & Tenreyro (2006) in using the Poisson pseudo-maximum likelihood (PPML) estimator instead of OLS. PPML rests on the assumption that the observed number of DDoS attacks between countries i and j has a Poisson distribution with a conditional mean θ_{ij} . The Poisson probability distribution is:

$$\Pr\{freq_{ij} = n\} = \frac{\exp(-\theta_{ij})\theta_{ij}^n}{n!}, \quad n = 0, 1, \dots \quad (6)$$

The conditional mean θ_{ij} is an exponential function of regression variables. In the empirical analysis, we work with specifications of the form:

$$\theta_{ij} = \exp(\alpha_0 + \beta' X_{ij}), \quad (7)$$

where α_0 is a constant, β is a vector of coefficients and X_{ij} is a vector of explanatory variables. The Poisson model of equations 6 and 7 is estimated with maximum likelihood.

As prior empirical research on DDoS attacks is scarce, there is little guidance on the factors that drive these attacks. For this reason we consider a variety of explanatory variables, subdivided over four groups: technological factors, geographical factors, economic factors and conflict factors.

Technological factors. DDoS attacks are undeniably technological phenomena. Our economic model suggests that the number of PCs connected to the internet in both countries matters for the size of the botnet and the attractiveness of a country as a DDoS victim. Because we could not obtain data on this particular variable, we use instead (the logarithm of) the number of internet users. Another relevant technological factor is bandwidth in the country of origin. Bandwidth may facilitate the creation of an effective botnet and could strengthen an attack. Finally, we consider the level of cybersecurity in both countries. A low level of cybersecurity may enable a hacker to form a large botnet and make websites more vulnerable to attacks.

Geographical factors. A classic stylized fact from international economics is

that geography is a key determinant of trade patterns between two countries.⁶ The distance between two countries, in particular, is an important factor that affects export levels. We aim to test whether distance also matters for DDoS attacks. Another geographic factor that we will include is the presence of a common border.

Economic factors. Hackers may have economic reasons for a DDoS attack, such as intellectual property theft or extortion. In addition, the sheer economic size of the victim may attract more attacks. To test this possibility, we include GDP per capita and the level of trade between both countries. We also use the “rule of law” index, which measures whether a country’s legal institutions are predictable, enforceable and, ultimately, benefit a market economy. One hypothesis is that extortion is more difficult in countries with a strong rule of law, which would lower the incidence of DDoS attacks.

Conflict factors. An often suggested motive for DDoS attacks are conflicts. This motive may for instance underlie the 2014 DDoS attacks on North Korea in the wake of the widely discussed infiltration of the network of Sony Pictures Entertainment. Zuckerman et al. (2010) reports cross-border DDoS attacks between China and Japan, Russia and Georgia, Russia and Estonia, China and US, Argentine and United Kingdom, Japan and South Korea, and Algeria and Egypt. Relatedly, several countries have established specialized cyber war units to strengthen the national digital defense and to engage cyber operations abroad.

We investigate the conflict hypothesis by including a measure of the level of military spending in both countries and dummies that indicate whether two countries were a single country or whether two countries have historical colonial ties. We also consider a dummy for countries with a shared language. A common language might facilitate a common understanding and thus prevent conflicts.

4 Data

We obtained detailed information on DDoS attacks from Digital Attack Map, which is a collaboration between Google and Arbor Networks.⁷ The data includes information on the number, duration, size, country of origin

⁶The seminal paper in this “gravity equation” field is Tinbergen (1962).

⁷See www.digitalattackmap.com for more information on this project.

and country of destination of DDoS attacks between April 2013 and August 2014. The raw dataset consists of 55,458 attacks. We can not discern between successful and unsuccessful attacks.

The stated source of the DDoS attack may or may not be the location of the actual attacker. The designer of the DDoS attack may reside elsewhere and use a foreign botnet to launch the attack. In a relatively small number of cases, 824, Digital Attack Map lists two or three countries simultaneously as the source. Just 86 attacks are targeted against more than one country. To maintain a bilateral square dataset, we interpreted each attack from countries *A* and *B* to country *C* as two distinct attacks, from *A* to *C* and from *B* to *C*. This results in a slightly larger set of 56,409 attacks. For our analysis, we need data on both the source country and the destination country. This requirement is met for a subset of 14,900 DDoS attacks.

Table 1: Top tens

| Source | Destination | Country pairs |
|-------------------|----------------|-------------------------|
| US (6256) | US (4065) | US-US (2039) |
| China (2851) | China (3050) | US-Poland (1689) |
| Netherlands (834) | Poland (2114)) | China-China (1183) |
| Germany (593) | Peru (1068) | China-US (959) |
| Korea (362) | France (571) | US-China (699) |
| France (350) | Brazil (476) | US-Peru (407) |
| UK (328) | UK (458) | Netherlands-China (366) |
| Brazil (286) | Russia (361) | US-Russia (305) |
| Peru (240) | Malaysia (232) | Peru-Peru (222) |
| Thailand (229) | Thailand (188) | US-Brazil (180) |

Source: Digital Attack Map and own calculations.

Note: This table shows the top ten countries and country pairs in terms of the highest number of DDoS attacks, during April 25 2013 and August 5 2014.

Table 1 lists the ten countries from which most DDoS attacks are launched, the ten countries with the highest number of incoming attacks and the ten most frequently observed country pairs. The United States and China head the list of source countries, which seems plausible as both countries have worldwide most internet users. However, the presence of the Netherlands and Peru in the top 10 of source countries is puzzling because these countries are relatively small, in terms of the number of internet users and absolute level of GDP. Overall, the list of destination countries seems more surprising than the list of source countries, with

Poland, Peru and Malaysia as unexpected entries. The third column reveals that many DDoS attacks occur within the same country. Of all 14,900 documented attacks, 4,282 take place within one country. Nine of ten bilateral attacks involve China or the United States.

The lists in table 1 can be seen as an empirical puzzle. Why are these countries involved in DDoS attacks? Why are China and US so dominant? Why does the list of destination countries appear more diverse? The remainder of the paper can be seen as an econometric approach to solving this puzzle.

Starting with a list of 186 countries, we calculated the number of DDoS attacks between any pair of countries i and j . This yields a list of 34,596 ($=186*186$) directed country pairs. We observe at least one DDoS attack between 883 country pairs and for 37 countries at least one “within” country attack. Thus, for a large number of countries we observe zeros.⁸

During the sample period Digital Attack Map has collected data on attacks from over 270 ISPs. Because it is likely that some countries are not represented by an ISP, we could erroneously interpret zero DDoS attacks between two countries as the actual absence of attacks, whereas DDoS attacks simply may not have been registered. To correct for this, we exclude all countries from our dataset for which we never observe a DDoS attack. This leads to a drop of 8,798 country pairs.

We supplement the DDoS data with information from the World Bank on the number of internet users, bandwidth, the rule of law, military expenditures as a fraction of GDP, and GDP per capita. To estimate the level of cybersecurity, we use data from Microsoft (2014). Microsoft collects and publishes data on the CCM (or *computers cleaned per mille*). This denotes the number of cleaned computers for every 1,000 executions of Microsoft’s malicious software removal tool which checks for many common types of malware and viruses. The CCM is therefore likely to give a fairly accurate estimate of the security level of tested PCs.

We obtained information on trade in 2012 from the United Nation’s Comtrade database. To calculate the level of trade, we added per country the import and export level.

Finally, we added data on bilateral factors, such as distance, common border or common language, from the CEPPI gravity database. “ComBor” indicates whether two countries share a common border. “Language” is

⁸This phenomenon of many zeros is not unique to DDoS data. In international economics, for instance, it is well-known that the majority of countries do not trade with each other. Helpman, Melitz & Rubinstein (2008), for instance, report that between 1970 and 1997, around 50 percent of the countries in their data did not export to each other.

a dummy that indicates whether two countries have the same ethnological language, “SMCTRY” indicates whether two countries were a single country, “ComCol” is 1 for countries with a common colonizer after 1945 and “Colony” is a dummy for countries with a colonial link. See Mayer & Zignago (2011) for a discussion of this database.

Table 2 summarizes our variables at the country level and, if applicable, at the country pair level.

Table 2: Summary statistics

| | count | mean | sd | min | max |
|-----------------------------------|-------|--------|-------|------|--------|
| No. of DDoS attacks as source | 186 | 57.09 | 340.1 | 0 | 4217 |
| No. of DDoS attacks as victim | 186 | 57.09 | 264.6 | 0 | 2102 |
| Trade (*billion dollars) | 186 | 170.41 | 456.6 | 0.0 | 3502.4 |
| No. of internet users (million) | 186 | 10.86 | 40.1 | 0.0 | 463.3 |
| GDP per capita (*1,000 dollars) | 186 | 10.39 | 15.3 | 0.2 | 80.3 |
| Military exp. (%) of GDP | 143 | 1.96 | 1.6 | 0.0 | 11.5 |
| CCM | 125 | 9.41 | 7.2 | 0.0 | 54.6 |
| Bandwidth | 182 | 0.10 | 0.5 | 0.0 | 6.4 |
| Rule of law | 182 | -0.08 | 1.0 | -1.7 | 1.9 |
| No. of DDoS attacks per pair | 25684 | 0.41 | 13.9 | 0 | 1689 |
| Trade per pair (*billion dollars) | 23773 | 1.33 | 13.0 | 0 | 629.5 |
| Distance (*1,000 km) | 23773 | 7.83 | 4.4 | 0.1 | 19.8 |
| Common border | 23773 | 0.02 | 0.1 | 0 | 1 |
| Common ethnic language | 23773 | 0.13 | 0.3 | 0 | 1 |
| Same country | 23773 | 0.01 | 0.1 | 0 | 1 |
| Common colonizer | 23773 | 0.07 | 0.3 | 0 | 1 |
| Colonial link | 23773 | 0.02 | 0.1 | 0 | 1 |

Source: Digital Attack Map, CEPII, Microsoft and the World Bank.

Note: GDP per capita measured in constant US dollars. CCM: number of computers cleaned for every 1,000 executions of Microsoft’s malicious software removal tool. Bandwidth is displayed as megabits per second per internet user. Trade is the sum of exports and imports.

For “within” country observations, we applied a few transformations. We replaced the level of trade with twice the level of GDP; “ComBor” and “Language” are set at 1; “SMCTRY”, “ComCol” and “Colony” are 0, and “Distance” obtains the value 1.⁹.

⁹In one of the robustness checks in section 5.2, we restrict the regression to bilateral

In the regressions we use a logarithmic transformation of all dependent variables, except for the dummies. This allows us to interpret the coefficients as elasticities. Before we proceed to the results, it is useful to note that our empirical setup only allows us to observe the general patterns in the data, and does not measure the causal relationships between economic variables and the incidence of DDoS attacks.

5 Results

5.1 Baseline

We consecutively test the groups of regressor outlined above and retain those variables that turn out to be significant (a p-value below 5 percent) by estimating versions of (7) with PPML. Residuals can be correlated both by source country and by destination country. As we can not correct for both kinds of correlation simultaneously, we report two types of standard errors: one clustering residuals by source country and one clustering residuals by destination country. To constrain the impact of outliers, we winsorized all continuous variables for the 1 percent highest and lowest observations.¹⁰

attacks and find that the main results are unaffected.

¹⁰In section 5.2 the baseline results are reported for unwinsorized data. The signs and significance remains unaltered.

Table 3: Baseline results

| No. DDoS | Technology | Geography | Economy | Conflict | Final |
|-----------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Users(i) | 0.677*** (0.03) (0.05) | 0.689*** (0.05) (0.06) | 0.452*** (0.04) (0.06) | 0.409*** (0.04) (0.05) | 0.412*** (0.04) (0.05) |
| Users(j) | 0.680*** (0.05) (0.05) | 0.687*** (0.05) (0.06) | 0.453*** (0.06) (0.05) | 0.402*** (0.05) (0.06) | 0.408*** (0.05) (0.06) |
| Bandwidth(i) | 0.171*** (0.05) (0.02) | 0.180** (0.07) (0.02) | 0.054 (0.06) (0.02) | 0.115* (0.05) (0.02) | 0.118* (0.05) (0.02) |
| Bandwidth(j) | 0.148 (0.02) (0.08) | | | | |
| CCM(i) | -0.17 (0.09) (0.04) | | | | |
| CCM(j) | -0.24 (0.02) (0.13) | | | | |
| Distance | | -0.09* (0.03) (0.04) | -0.03 (0.02) (0.02) | | |
| ComBor | | 0.231 (0.18) (0.25) | | | |
| GDPcap(i) | | | 0.051 (0.09) (0.04) | | |
| GDPcap(j) | | | 0.129 (0.04) (0.13) | | |
| Trade | | | 0.249*** (0.03) (0.05) | 0.336*** (0.03) (0.05) | 0.333*** (0.03) (0.05) |
| Law(i) | | | 0.093 (0.11) (0.06) | | |
| Law(j) | | | 0.068 (0.06) (0.18) | | |
| Military(i) | | | | -0.02 (0.08) (0.06) | |
| Military(j) | | | | 0.007 (0.06) (0.15) | |
| Language | | | | -0.01 (0.08) (0.12) | |
| SMCTRY | | | | 0.560* (0.28) (0.23) | 0.408 (0.27) (0.24) |
| ComCol | | | | -1.50 (0.52) (0.81) | |
| Colony | | | | -0.21 (0.16) (0.13) | |
| N | 10366 | 10366 | 10366 | 10366 | 10366 |
| Pseudo R ² | 0.369 | 0.355 | 0.394 | 0.389 | 0.386 |

Note: This table presents the regression estimates based on the Poisson pseudo-maximum likelihood estimator. The standard errors are between round brackets and are clustered both at country i (left) and at country j (right). Stars indicate significance of the estimates based on the highest p-value. Legend: *: $p < .05$, **: $p < .01$ and ***: $p < .001$.

Table 3 presents the baseline results. As noted above, the estimated coefficients of the non-dummy regressors are elasticities. The first model features the technological variables. The number of internet users in both countries is highly significant, as well as the level of bandwidth in the source country. Consistent with our economic model, both variables relate positively with the observed frequency of DDoS attacks. The coefficient for CCM, an indicator for cybersecurity, is insignificant.

In the second model, we retain the significant variables from the first specification and add two geographical factors: distance and common border. Only distance turns out to be significant. The coefficients for the number of internet users and the level of bandwidth in the source country remains significant, though.

In model "Economy" we include (the logarithm of) GDP per capita, (the logarithm of) the Rule of Law and (the logarithm of) the level of trade. We do not see an effect of GDP per capita or the Rule of Law indicator on DDoS patterns. Perhaps the incidence of botnets and the attractiveness of victims depends more on technological factors than on country specific economic development. The level of trade, however, has a significant and positive sign. This suggests that economic ties, as indicated by trade relations, are a determinant of DDoS patterns. Bandwidth is no longer significant in this specification. We nevertheless include bandwidth in the next model, as we think that the insignificant result may be driven by multicollinearity of bandwidth and GDP per capita.

As a test of the cyberwar hypothesis, we introduce six indicators for "conflict" in the fourth specification. The only significant variable is the same country dummy "SMCTRY".

Finally, in the fifth model, we re-estimate the model with all significant variables from the previous specification. Overall, the findings are consistent with the economic model in section 2.¹¹ The number of internet users and bandwidth in the source country all are positively related to the frequency of DDoS attacks.

For both the source and the destination country, the elasticity of the number of attacks with respect to the number of internet users is 0.4. This implies that a ten percent increase in the number of internet users in both the source and destination country would lead to an increase of eight percent in the total number of attacks between these countries. The relation between the number of attacks and source country bandwidth is inelastic and the same holds for trade.

¹¹As a robustness exercise (not shown in this paper), we tested the four factors in a reverse order and obtained that the same variables remain significant.

5.2 Robustness

We consider whether the results for the baseline model are sensitive to changes in specifications, estimation techniques and data samples.

Table 4: Alternative cybersecurity indicators

| No. DDoS | I | II | III | IV |
|-----------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Users(i) | 0.427*** (0.04) (0.04) | 0.430*** (0.04) (0.04) | 0.211** (0.07) (0.04) | 0.388*** (0.04) (0.05) |
| Users(j) | 0.394*** (0.04) (0.05) | 0.383*** (0.05) (0.06) | 0.371*** (0.05) (0.06) | 0.398*** (0.04) (0.05) |
| Bandwidth(i) | 0.129** (0.05) (0.02) | 0.131* (0.05) (0.02) | | 0.121** (0.04) (0.02) |
| Trade | 0.358*** (0.03) (0.04) | 0.355*** (0.03) (0.04) | 0.355*** (0.03) (0.06) | 0.355*** (0.03) (0.05) |
| Port23(i) | | 0.018 (0.05) (0.03) | | |
| Port23(j) | | -0.02 (0.03) (0.13) | | |
| No. hub members | | | 0.166* (0.07) (0.03) | |
| Throughput(i) | | | | |
| ≤ 1000 Gb | | | | 0.369 (0.14) (0.37) |
| 1000 – 2000 Gb | | | | 0.478 (0.12) (0.38) |
| ≥ 2000 Gb | | | | 0.153 (0.11) (0.37) |
| N | 18805 | 16907 | 5913 | 18805 |
| Pseudo R ² | 0.451 | 0.436 | 0.412 | 0.452 |

Note: Estimates based on the Poisson pseudo-maximum likelihood estimator. The standard errors are between round brackets and are clustered both at country i (left) and at country j (right). Stars indicate significance of the estimates based on the highest p-value. Legend: *: $p < .05$, **: $p < .01$ and ***: $p < .001$.

Alternative specifications. First, we consider alternative technological variables. In the baseline set-up, we measured the level of cybersecurity with the CCM indicator. In addition to the CCM, we developed a measure of

cybersecurity ourselves. We obtained data on the status of ports of about 1.3 billion IPv4 addresses from the Internet Census (2012).¹² We focus on the status of port 23, because computer security experts generally recommend to close port 23 for remote login attempts. An open status of port 23 may therefore indicate a weak security policy. We were able to calculate for all countries in our dataset the fraction of addresses with an open port 23.

The results are in table 4. The first column re-estimates the baseline model with the number of internet users, bandwidth and trade as the regressors and reports similar estimates as in table 3. The number of observations is larger than in table 3, because here we do not maintain a constant sample. In the second column, we add the logarithm of the fraction of addresses with an open port 23 in both the source country and the destination country. Just as the CCM, this indicator is not significant.

A key variable in our economic model is bandwidth in the source country. The importance of bandwidth should not depend on how exactly internet speed is measured. An alternative indicator is the size of internet exchange points, which provide infrastructure services to ISPs to lower costs and enhance bandwidth. One of the largest internet exchange point, AMS-IX, is located in the Netherlands and serves over 679 clients and boasts a peak data rate of 3200 Gbits per second. According to some observers, the excellent internet infrastructure in the Netherlands explains why so many DDoS attacks originate from that country.

We test the importance of internet exchange points in two ways. First, in the third column of table 4, we consider the logarithm of the number of members of internet exchange points in the source country. This variable has a similar effect as bandwidth. Second, we consider the size of the internet exchange point, as measured by the peak data rates (or throughput). We considered three classes: exchange points with a throughput up to 1,000 Gbit, between 1,000 and 2,000 Gbit and over 2,000 Gbit per second. We include the dummies for the three classes in the baseline model, in the final column, and find that throughput has no extra explanatory power over bandwidth.

Alternative estimation techniques. The baseline model has been estimated with PPML. A closely related technique is zero-inflated Poisson regression. This technique is well suited for count data with a large number of zeros. In particular, if the underlying true data generating process of the

¹²The Internet Census was conducted by an anonymous hacker, who developed a botnet of 420,000 devices to scan the ports of all worldwide active IPv4 addresses.

zeros differs from the remaining observations, zero-inflated Poisson is appropriate because it explicitly models two latent groups within the population: observations with zero counts and observations having a non-zero probability of having a positive count. See e.g. Cameron & Trivedi (2010) for more details.

We re-estimated the baseline model with zero-inflated Poisson (ZIP). From column ZIP in table 5 we can see that the number of internet users and the level of trade remain significantly related to DDoS attacks. Bandwidth in the source country is no longer significantly associated with the number of DDoS attacks. The second part of column ZIP presents the results from a logit regression that explains the probability of no attacks between countries i and j . The negative coefficients indicate that internet users, bandwidth and trade relate positively with the probability of an attack.

The fact that bandwidth is significant here, sheds more light on the role of bandwidth in the emergence of DDoS attacks. Apparently, more bandwidth makes an attack more likely, but is not related with the number of attacks. This suggests that a sufficient amount of bandwidth is a necessary condition for a “bot herder” to form a botnet.

The Poisson model assumes equidispersion, i.e. that the conditional variance is equal to the conditional mean. The negative binomial regression model is a modification of the Poisson model and allows for overdispersion, i.e. that the conditional variance is higher than the conditional mean. To consider whether this matters, we also ran the baseline model with the negative binomial regression model (NBREG).

The estimates from the negative binomial regression are very similar to the baseline Poisson estimates, as can be seen in table 5. The assumption of equidispersion therefore does not seem critical for the results.

The positive coefficient for trade may be a spurious relation if we failed to include the “true” drivers of DDoS attacks. To test for the possibility of omitted variable bias, we estimate a baseline Poisson model with fixed effects for country i and country j . This specification allows us to control for any country specific variation. The inclusion of country fixed effects implies that other country specific factors drop out and we can only retain pairwise variables.

Column FE in table 5 for the results of the fixed effects estimator. If omitted variable bias were present, we would expect a noticeable change in the estimated coefficient for the level of trade. The coefficient on trade is significant, but one third smaller than the estimate from our baseline

Table 5: Different estimation techniques

| No. DDoS | Poisson | ZIP | NBREG | FE |
|-----------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Users(i) | 0.407*** (0.04) (0.05) | 0.174*** (0.03) (0.03) | 0.474*** (0.05) (0.05) | |
| Users(j) | 0.403*** (0.05) (0.06) | 0.131*** (0.03) (0.03) | 0.478*** (0.05) (0.07) | |
| Bandwidth(i) | 0.116* (0.05) (0.02) | 0.035 (0.02) (0.01) | 0.143** (0.06) (0.03) | |
| Trade | 0.337*** (0.03) (0.05) | 0.063* (0.02) (0.03) | 0.310*** (0.04) (0.05) | 0.223*** (0.03) (0.03) |
| <i>inflation</i> | | | | |
| Users(i) | | -0.44*** (0.07) (0.06) | | |
| Users(j) | | -0.52*** (0.04) (0.11) | | |
| Bandwidth(i) | | -0.14* (0.06) (0.03) | | |
| Trade | | -0.33*** (0.05) (0.05) | | |
| N | 10366 | 10366 | 10366 | 10366 |
| Pseudo R ² | 0.385 | 0.362 | 0.330 | 0.545 |

Note: The dependent variable is the number of DDoS attacks between country i and country j . For Poisson, ZIP and NBREG: the standard errors are clustered at country i (left) and at country j (right). For FE: the standard errors are robust and between round brackets. Legend: *: $p < .05$, **: $p < .01$ and ***: $p < .001$.

results. This suggests that omitted variable bias might be responsible for overestimation of the relation between DDoS attacks and trade relations, but that it is unlikely that this relation is spurious.

Alternative samples. In the empirical analysis so far, we included “within” country attacks. And, as we observed earlier the “within” country attacks constitute a sizeable portion of all attacks in our data. As it is possible that “within” country attacks are somehow different from bilateral attacks, it would be interesting to check whether the estimates differ if we exclude “within” country attacks.

The estimates for exclusively bilateral attacks can be found in the sec-

ond column in table 6. Compared to the estimates from the baseline sample, the estimates for the number of internet users and bandwidth is slightly higher, and the estimate for trade is slightly lower. Overall, we see little differences with the baseline model.

The dataset contains information on 186 countries and the differences between countries with respect to the use of ICT are large. In particular, it is possible that the incentives and scope for hackers to launch DDoS attacks are different between developing and developed countries. To test whether a more homogeneous sample leads to different results, we estimated the baseline model for developed countries.¹³ The results for the sample of developed countries are presented in the third column of table 6. The estimates do not differ substantially from those of the baseline.

The final column in table 6 re-estimates the baseline model with un-winsorized data. This has no impact on the signs and coefficients of the estimates, but does inflate the coefficients of the technological variables.

Table 6: Sample robustness checks

| No. DDoS | Baseline | Bilateral | Developed | No winsor. |
|-----------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Users(i) | 0.427*** (0.04) (0.04) | 0.470*** (0.04) (0.05) | 0.428*** (0.04) (0.04) | 1.075*** (0.26) (0.24) |
| Users(j) | 0.394*** (0.04) (0.05) | 0.436*** (0.05) (0.06) | 0.381*** (0.05) (0.05) | 0.672*** (0.07) (0.09) |
| Bandwidth(i) | 0.129* (0.05) (0.02) | 0.132* (0.05) (0.02) | 0.121* (0.05) (0.02) | 0.244* (0.10) (0.09) |
| Trade | 0.358*** (0.03) (0.04) | 0.324*** (0.03) (0.05) | 0.341*** (0.03) (0.05) | 0.342*** (0.07) (0.08) |
| N | 18805 | 18692 | 14492 | 18805 |
| Pseudo R ² | 0.451 | 0.449 | 0.423 | 0.730 |

Note: Estimates based on the Poisson pseudo-maximum likelihood estimator. The standard errors are between round brackets and are clustered both at country i (left) and at country j (right). Stars indicate significance of the estimates based on the highest p-value. Legend: *: $p < .05$, **: $p < .01$ and ***: $p < .001$.

¹³Our classification of developed countries is based on the definition of the World Bank.

6 Concluding remarks

Distributed Denial of Service (DDoS) attacks are a frequently occurring type of cybercrime, with potentially large costs to the real economy. We propose a simple model of the size and direction of DDoS attacks. The main predictions of the model are that effective botnets are located in countries with many internet users and high internet speeds, and that the most attractive targets of DDoS attacks are countries with many internet users.

We use a theoretical framework to derive a structural equation that resembles the "gravity equations" common in the literature on international trade. The empirical results are consistent with the predictions of the model. The number of internet users is strongly related to the number of international DDoS attacks: our results suggest that a ten percent increase in the number of internet users worldwide would raise the total number of DDoS attacks by eight percent. Bandwidth in the country of origin is also significantly related to attacks, but quantitatively not very important. The vulnerability of computers does not seem influential.

Trade relations are significantly related to attacks, while other economic factors including GDP per capita do not appear to play a role. The geographical distance between countries is not relevant, while historical ties between countries are significantly related to the number of attacks.

This paper is one of the first to explore possible determinants of cybercrime at an aggregate level. We hope that by uncovering some general patterns in the data, our research may contribute to the growing and exciting field of cybersecurity economics.

References

- Acemoglu, D., Malekian, A., & Ozdaglar, A. (2013). Network security and contagion. *NBER Working Paper, 19174*, 1 – 69.
- Anderson, R. & Moore, T. (2006). The economics of information security. *Science*, 314, 610 – 613.
- Athey, S. & Stern, S. (2013). The nature and incidence of software piracy: Evidence from windows. *NBER Working Paper, 19755*, 1 – 50.
- Burger, M., van Oort, F., & Linders, G.-J. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4, 167–190.

- Cameron, A. & Trivedi, P. (2010). *Microeconometrics Using Stata*. Stata Press.
- Center for Strategic and International Studies (2014). Net losses: Estimating the global cost of cybercrime. McAfee Report on the Global Cost of Cybercrime.
- Flowerdew, R. & Aitkin, M. (1982). A method of fitting the gravity model based on the poisson distribution. *Journal of Regional Science*, 22, 191–202.
- Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics*, 123, 441–487.
- Internet Census (2012). Port scanning /0 using insecure embedded devices. Available at internetcensus2012.bitbucket.org/paper.html.
- Johnson, B., Laszka, A., Grossklags, J., Vasek, M., & Moore, T. (2014). Game-theoretic analysis of DDoS attacks against bitcoin mining pools. mimeo.
- Kaspersky Lab (2015). Global it security risks survey 2014: distributed denial of service (ddos) attacks. Technical report, Kaspersky.
- Mayer, T. & Zignago, S. (2011). Notes on CEPII's distances measures: The geodist database. *MPRA Paper*, 36347, 1 – 19.
- Microsoft (2014). *Microsoft Security Intelligence Report*. Microsoft Corporation.
- Ponemon Institute (2013). Cost of cyber crime study: United states. Report sponsored by HP Enterprise Security.
- Silva, J. S. & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88, 641 –658.
- Tinbergen, J. (1962). *The World Economy. Suggestions for an International Economic Policy*. Twentieth Century Fund.
- van Eeten, M., Bauer, J. M., Asghari, H., & Tabatabaie, S. (2010). The role of internet service providers in botnet mitigation: An empirical analysis based on spam data. OECD STI Working Paper no. 5.
- Vasek, M., Thornton, M., & Moore, T. (2014). Empirical analysis of denial-of-service attacks in the bitcoin ecosystem. mimeo.

Zuckerman, E., Roberts, H., McGrady, R., York, J., & Palfrey, J. (2010). Distributed denial of service attacks against independent media and human rights sites. Report of the *Berkman Center for Internet & Society at Harvard University*.



Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (070) 3383 380

April 2015 | ISBN 978-90-5833-688-0