Factors impacting botnet infections

By Matthijs Bijman, Max Pigmans, Daniël Vos, Rens Heddes

Introduction

The rise of the Internet of Things (IoT) has lead to widespread use of cheap, poorly secured devices. The volume of and lack of security present in these devices has made them a common vehicle for distributed denial of service (DDoS) attacks performed by botnets. In recent years, several high profile DDoS attacks originated from the Mirai IoT botnet. Since these botnets are mostly used for DDoS attacks or sending spam/phishing emails[1], the owners of these devices do not bear the damages and thus have low incentives to secure their devices[2]. Consequently, most mitigation strategies against botnet attacks are designed for the victims of such attacks[3][4].

In this report we select three actors related to this security issue, analyze how each actor is able to mitigate the risk, how they benefit from this mitigation, but also how others are affected. We then further analyse the performance of all ISPs in a country which is represented by the dataset. To do this, we identify various factors that may explain the variance in the dataset, collect data for these factors, and perform statistical analysis.

Chosen Actors

In this section, we will consider the security issue from the perspective of three related actors. For each of these actors, we identify one concrete countermeasure they could use to mitigate the issue. For each of these countermeasures we then discuss the distribution of costs and benefits among all actors. Based on these costs and benefits we can then analyze whether the actor is incentivized to take the countermeasure. Finally, we will reflect on the role of externalities around the will consider the following three actors:

1. **Internet Service Providers**. They are the problem owner, and the same factors that made them problem owner in our previous assignment, where we discussed how ISPs are in a good position to see if and how many new devices are being infected and they have the knowledge required to take action against growing botnets. These reasons make them well suited to analyse in this report.
2. **Device Owners** are at the core of the security issue, since their devices are the ones being infected causing negative effects for other parties. Therefore device owners have a large impact on the botnet security issue, and have a wide variety of options available to act on it**.**
3. **DDoS Victims** bear most of the cost of the security issue, it is interesting to look at the issue from their perspective and see what countermeasures are available for them, and how the decisions from other actors influence them.

Internet Service Providers

Mitigation

ISPs have several ways to mitigate the risk of botnets. Most of these options stem from the fact that ISPs are responsible for providing their clients with internet connectivity, and thus have access to the traffic transported from and to their clients. An ISP is thus able to mitigate the risk when botnet traffic flows through their network. This can happen either when an infected device is present in their network, or when outside devices are attacking a device in the network.

When an infected device is present in the network, the ISP can notify the device owner and block the malicious traffic. When a device is being attacked in the network the ISP can block the traffic. Another option for mitigation is updating the hardware provided to customers to become more resilient against infections, assuming that they provide such hardware and can update them.

Cost, Benefit, and Incentives

The cost for the ISP to implement a mitigation is relatively high while the benefits are limited. The primary loss associated with the Mirai botnet is unavailability due to a DDoS attack, however these attacks are not commonly targeted at ISPs themselves. Instead, the ISP only transports the data. This reduces the loss associated with having infected devices in their network to such a degree that ISPs do not have an inherent incentive to mitigate the risk.

Incentives to mitigate the risk may come from other places. The ISP may include DDoS protection in their product, thus creating a profit incentive to mitigate the risk. Alternatively, if ISPs become liable for DDoS attacks originating from their network, they will be incentivized to reduce the number of infected devices in their network and thus implement controls.

Externalities

Because ISPs have control over the traffic going from and to their network, they can have a strong impact on the risk. While the ISPs will not receive strong benefits from this, their choices can lead to strong externalities. If the service provider were to mitigate the risk of botnet infections by implementing one or multiple of the aforementioned controls, then it would have a positive effect for victims of DDoS attack carried out by the botnet since there will be fewer participating devices and thus a less powerful DDoS attack. Similarly, the victims inside their network will be largely protected from attacks. Attackers therefore receive negative externalities because their attacks become less effective. Device owners will receive slight positive externalities, since their devices are less likely to be infected or participate in attacks. These externalities are not significant because the owners do not experience strong negative consequences from being infected.

Device Owners

Mitigation

Device owners can mitigate the issue by updating their Internet of Things (IoT) devices and make sure default authentication credentials from the manufacturer are changed to a more secure set of credentials. These controls can be extremely effective at preventing infections, because the methods used to infect devices are very basic, and mostly rely on default credentials.

Cost, Benefit, and Incentives

The cost for implementing these mitigation strategies is quite low for device owners, since it only requires a small amount of manual work. The benefits are also limited for the device owners, since the devices are primarily used for DDoS attacks which are of little influence to the owners. Furthermore, for these controls to have a noticeable effect for the DDoS victims, many device owners need to implement them. This is because a single IoT device is responsible for a very small part of a DDoS attack. To increase the effectiveness of such a control, legislation could be introduced that forces device owners to be more careful about their IoT security. This is however quite costly, especially since enforcement will be difficult. All in all, the owners of IoT devices have little incentive to mitigate the risks of the botnet security issues because their benefits are negligible.

Externalities

When device owners take actions to mitigate the risk, attackers will be negatively affected because the size of the botnet will shrink. This makes DDoS attacks performed by the botnet less potent.

As a result, the victims of the DDoS attacks executed by the botnet are positively impacted by having to protect against a smaller attack. This will only be a small effect if only a small part of device owners choose to do so.

DDoS Victims

Mitigation

The last actor, DDoS victims, can mitigate the issue by investing in a DDoS protection service, which monitors incoming traffic to filter out attacks and they could upgrade server hardware and software to work more efficiently under load. Such controls are never perfect at preventing DDoS attacks, but may reduce the impact enough to give a positive return.

Cost, Benefit, and Incentives

The costs of these solutions is generally quite high. This can be explained by the high losses that occur as a result of DDoS attacks. DDoS protection providers are able to ask high prices because as long as their solution is effective at reducing losses, the benefits outweigh the costs. These losses are generally a result of loss of revenue due to unavailable services, so the benefits strongly depend on the kind of service. Depending on the kind of service, the victim will be strongly incentivized to implement these controls, or may decide to accept the risk.

Externalities

Since the losses caused by DDoS attacks are largely paid for by the victims of attacks, the externalities of implementing contols are small. Implementing these mitigation solutions could however have a negative impact on other potential victims, because attackers might choose to target other potential victims with less protection against a DDoS attack. Similar to other actors, implementing these controls will have negative effects for the attacker, since their attacks become less effective.

Factors of the Security Metric

To compare countries' security performance regarding the Mirai botnet we looked at how many new IPs were infected in one week per country (the week from 2018-6-1 until 2018-6-7). This is then normalized by the number of internet users in the country. We chose to look at only one week because the effects of old infected devices having their IP rotated will be less severe. We combine all ISPs in a country because for this aggregation of actors we can a) analyse their performance from the dataset and b) find additional data that may explain the performance. For individual ISPs finding data explaining their performance is infeasible. We assume that ISPs in the same country will have similar policies since they are held to the same regulatory requirements.

We believe the following factors might influence the amount of new infections in a week per country:

- The country's overall cyber security performance, if a country performs worse overall it might also perform worse in having devices infected by Mirai.
- The country's ICT development level, as more ICT infrastructure might open up the possibility of having more devices infected.
- The country's Crime index, countries that have a higher crime rate might have more people trying to infect devices

In the rest of this section we collect data on the factors mentioned above and correlate these datasets with the Mirai infected IPs security metric to investigate whether these factors could be used to explain the security performance.

Normalizing infection rate

In order to analyze the correlations with the normalized infection rate, we first evaluate the infection rate statistic itself. On the left side of Figure 1 we show a histogram of the normalized infection rate data. We see that it is extremely bottom heavy, with some outliers. In order to improve the normality of the data we applied a logarithm to the dataset and obtained the data as shown in the right side of Figure 2. There is now a better spread between the countries, which will aid future analysis.

We computed the normality of the transformed data using both the Shapiro–Wilk test and D'Agostino's K-squared test. We obtained from the first a p value of *0.025* and from the second a p value of *0.021*. This leads us to conclude that the transformed data is still not normally distributed. We will therefore use Spearman's rank correlation coefficient when determining correlations. Since we consider the transformed distribution more preferable than the original for analysis, we will continue to use it in the following sections.

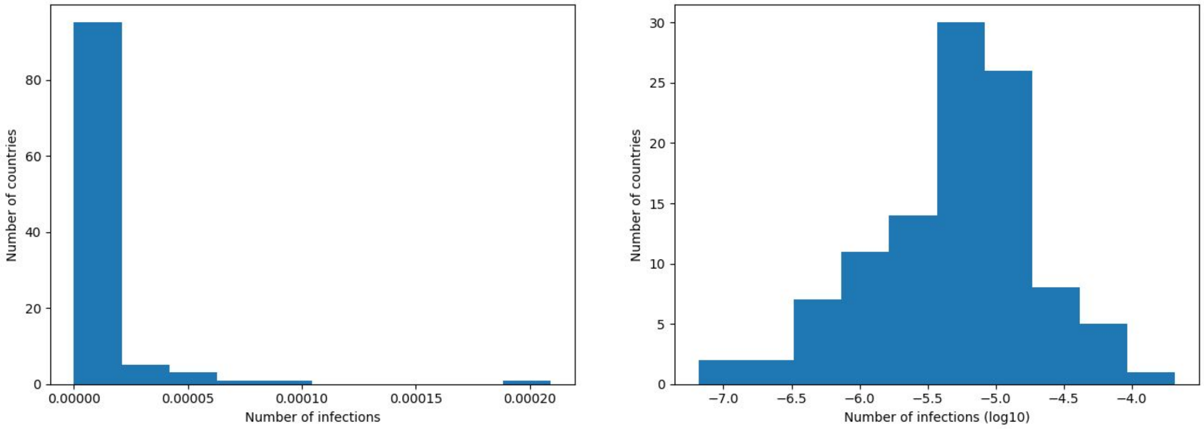

Figure 1: Distribution of the infection rate per country. On the left is the normalized but untransformed data, on the right a logarithm has been applied to the data.

Overall security performance

To measure a country's overall security performance, we make use of 2018's Global Cybersecurity Index (GCI) [5]**.** This index assigns a number between 0 and 1 to each country representing their cyber security performance. The GCI considers five distinct aspects: legal, technical, organizational, capacity building and cooperation. The legal pillar focuses on cybercrime legislation and cybersecurity regulations. The technical pillar looks at, among other things, the implementation of standards, technical measures in place and mechanisms to protect children online. The organizational pillar considers nationwide strategy and responsible central agency. The capacity pillar looks at the quantity and quality of training and research into future cybersecurity experts and measures. The cooperation pillar involves how much companies and agencies are cooperating both nationally and internationally. Combined, these metrics give a good comparison basis to determine how well countries are managing their cybersecurity.

When correlated with the Mirai security metric, Spearman's correlation gives a positive correlation of 0.194 with p=0.0274. The fact that GCI correlates positively with the number of infections is counter intuitive since before we expected that more security would lead to better protection against infections and therefore fewer infections. However, it could be that countries with a higher GCI also have better and more IT infrastructure which would be a valuable target to botnet operators. This could be because the GCI better captures a countries cybersecurity readiness against larger attacks focusing on corporations and the capability to respond to such large threats, rather than individual awareness of basic cybersecurity practices.

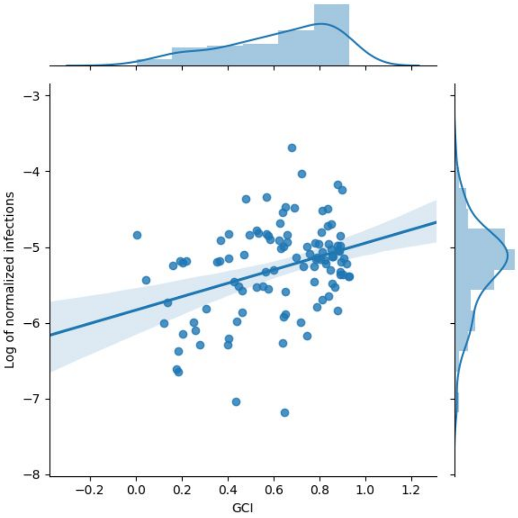

Figure 2: Correlation between Global Cybersecurity Index and amount of normalized newly infected devices per country. The blue line represents a linear regression estimate through the data points and the blue area around it a 95% confidence interval created using bootstrap samples.

Overall ICT development

To measure overall ICT development of each country we use the ICT Development Index (IDI) [6]. The IDI focuses on three key categories of metrics to compute their index. These are the ICT readiness, ICT use and ICT skills. ICT skills is weighed lower than the other two categories for the overall index. ICT access describes how well connected the population is to the internet. ICT use describes how active the population of a country is online. ICT skills focuses on how many years of schooling are standard in the country. Since we correlate with the normalized number of infected devices, any correlation that we find will be more so caused by the other factors in IDI, such as the number of devices, bandwidth and ICT skills.

When correlated with the Mirai security metric, Spearman's correlation gives a positive correlation of 0.472 with p=1.6*10^-8. The fact that IDI correlates positively with the number of infections is not surprising, since we predicted that a higher ICT development index means more internet devices that have the potential to be infected and there is no consideration for security in this index.

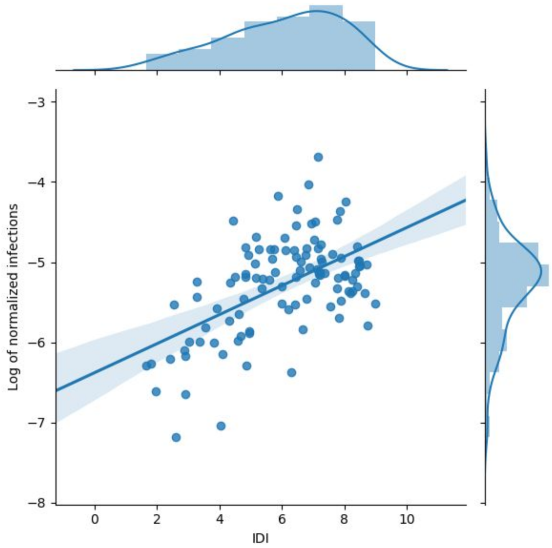

Figure 3: Correlation between ICT Development Index and amount of normalized newly infected devices per country. The blue line represents a linear regression estimate through the data points and the blue area around it a 95% confidence interval created using bootstrap samples.

Crime index rate

To measure overall non-cyber crime rate of each country we use the Crime Index Rate (Crime) [7]. This data is based on a survey in which people around the globe participated, factors like level of crime in that country and how worried participants were that they will be robbed. Other factors include drug problems in that country and the level of bribery, among others[8].
The expectation is that a higher crime index will correlate positively with the number of infected devices in a country, since we expect countries with a high crime index to first prioritize protection against non-cyber crime.

When correlated with the Mirai security metric, Spearman's correlation gives a negative correlation of -0.288 with p=0.0027. This is in contrast to our earlier expectations, and could possibly be attributed to the fact that countries with a high crime index probably have a lower ICT development score.

In order to investigate this possible correlation we plot the IDI score against the Crime Index Rate, see *figure 5*. Using Spearman's correlation this shows a very strong negative correlation of -0.618 with p=1.6*10^{-12}. This supports the idea that countries with a higher crime rate have a general low ICT development, and therefore explain why these countries receive less infections.

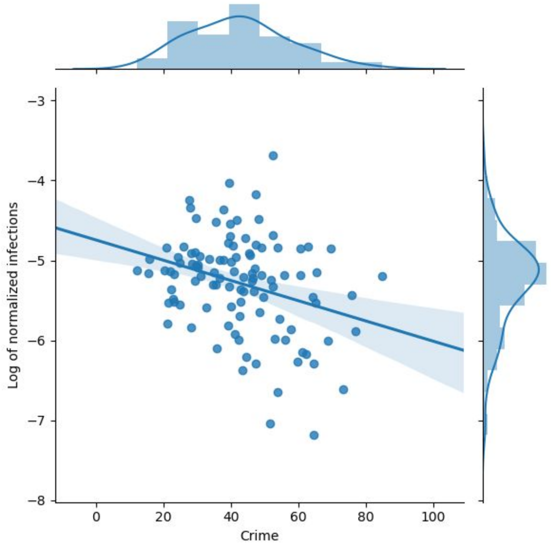

Figure 4: Correlation between Crime rate Index and amount of normalized newly infected devices per country. The blue line represents a linear regression estimate through the data points and the blue area around it a 95% confidence interval created using bootstrap samples.

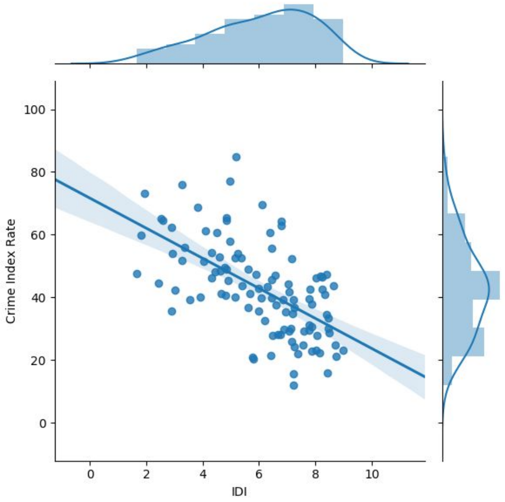

Figure 5: Correlation between Crime rate Index and ICT Development Index The blue line represents a linear regression estimate through the data points and the blue area around it a 95% confidence interval created using bootstrap samples.

Effect of Combined Factors on Security Metric

In order to see the combined effect of all variables we performed a multiple linear regression on the variables after standardization, of which the results can be seen in table 1 below. Given the

R^2 value of 0.422 we conclude that these metrics capture a decent portion of the variance, but there is still a lot that is not explained by these coefficients. We see that all terms involving the IDI are statistically significant with larger coefficients. This matches our expectations, as we have previously determined that IDI is the best correlating factor in our individual analysis. We do, however, see that both the IDI:GCI and IDI:Crime factors are statistically significant and so the addition of these factors into the model does improve our predictive capability compared to only IDI. Since we determined earlier that the Crime index rate is very heavily correlated with the IDI we perform a second analysis using only the variables of GCI and IDI as shown in table 2.

We see in table 2 that we achieve almost the same results. The same terms are significant, and we lose some predictive power as can be seen in the lower R^2 value of 0.345, which is lower but not by a large amount. We think that since the IDI:Crime combination was so significant, the removal of this from the model is not worth it in combination with the lower R^2 value.

	coef	standard error	t	P>|t|	[0,025	0,975]
Intercept	**0.2823**	**0.104**	**2.718**	**0.008**	**0.076**	**0.488**
IDI	**0.4561**	**0.113**	**4.037**	**0.000**	**0.232**	**0.680**
GCI	0.0660	0.115	0.574	0.567	-0.162	0.294
Crime	-0.0403	0.107	-0.376	0.708	-0.253	0.172
IDI:GCI	**-0.2610**	**0.104**	**-2.521**	**0.013**	**-0.466**	**-0.056**
GCI:Crime	-0.1856	0.117	-1.586	0.116	-0.418	0.047
IDI:Crime	**0.3910**	**0.112**	**3.505**	**0.001**	**0.170**	**0.612**
IDI:GCI:Crime	0.1188	0.087	1.373	0.173	-0.053	0.291

Table 1: Summary of multiple linear regression on our metrics. Statistically significant factors have been marked in bold. $R^2 = 0.422$

	coef	standard error	t	P>|t|	[0,025	0,975]
Intercept	0.1450	0.098	1.476	0.143	-0.050	0.340
IDI	**0.5397**	**0.104**	**5.208**	**0.000**	**0.334**	**0.745**
GCI	-0.0618	0.107	-0.577	0.565	-0.274	0.151
IDI:GCI	**-0.2289**	**0.090**	**-2.553**	**0.012**	**-0.407**	**-0.051**

Table 2: Summary of multiple linear regression on our statistically significant terms. Statistically significant factors have been marked in bold. $R^2 = 0.345$

Conclusion

In this report we discussed several actors involved with our security issue: ISPs, device owners, and victims. For each actor, we discussed the costs and benefits of possible countermeasures, and the externalities caused by them. We then performed a more in-depth analysis on the correlation between various datasets (Global Cybersecurity index, Crime Rate Index, and ICT Development Index) against the logarithm of normalized number of infections in each country. We conclude that there is a positive correlation between the number of infections per user, and each of the datasets we analysed.

Our analysis shows that the main factor influencing the normalized number of infected devices by Mirai within a country is the number of devices per person within that country. This is reflected in the correlation between ICT Development Index and the number of normalized infections. In an ideal situation we would have normalized the number of infected devices within a country over the total number of infectable devices in that country. However, there is no dataset available which describes the amount of infectable devices. In order to account for the size of a country we therefore normalized using the number of internet users within a country. This, however, does not take into account the number of devices per internet user. This could therefore explain the significant correlation between number of infected devices and the ICT Development Index, since the ICT Development Index does likely include the number of devices as one of its factors.

The GCI, surprisingly, negatively correlated with the number of infections. We think that this is largely due to the GCI focussing more on the cybersecurity capabilities of the country overall and of the corporations within, rather than that of individuals who might own infectable IoT devices, and additionally the fact that more developed countries are generally having a better GCI score.

The overall crime rate does have a correlation with the rate of infected devices, but we believe that this is largely because the crime rate is heavily correlated with the ICT Development Index. Since the combination of IDI and Crime rate does improve the prediction of a multiple regression model, it does appear that crime rate has some predictive capability that is not solely based on its correlation with IDI.

Overall, we find the best results using all three of our metrics to explain the variance in our dataset, obtaining an R^2 of 0.422 with each variable having a significant contribution in some combination. However, we note that this still leaves a large part of the variance in the security metric unexplained.

References

[1] Yury Namestnikov, “The economics of Botnets”.

[2] James A. Jerkins, “Motivating a Market or Regulatory Solution to IoT Insecurity with the Mirai Botnet Code”.

[3] S. T. Zargar, J. Joshi and D. Tipper, "A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks," in *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2046-2069, Fourth Quarter 2013.

[4] M. Özçelik, N. Chalabianloo and G. Gür, "Software-Defined Edge Defense Against IoT-Based DDoS," *2017 IEEE International Conference on Computer and Information Technology (CIT)*, Helsinki, 2017, pp. 308-313.

[5] International Telecommunication Union. Global Cybersecurity Index 2018/2019. Retrieved from https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx Annex B

[6] International Telecommunication Union. ICT Development Index 2017. Retrieved from https://www.itu.int/net4/ITU-D/idi/2017/index.html

[7] Crime index Rate, Numbeo. Retrieved from https://www.numbeo.com/crime/rankings_current.jsp

[8] About Crime Indices, Numbeo. https://www.numbeo.com/crime/indices_explained.jsp