Assignment Block 2

SpamHaus

DRAFT

Group 8
Jorrit van den Spek, Hugo Bijmans, Eveline Pothoven, Ben Hup, Lisette Altena

WM0824 Economics of Security

Table of Contents

1. Security Issues

Spam, it is a major source of frustration for internet users. More than halve of the e-mail traffic consists of spam (Figure 1). It causes high costs for companies. Employees who could otherwise spend their time on more productive work, waste their time. Network errors, which are among others caused by employees responding to spam mail, largely impacts the productivity. Furthermore spam uses storage and bandwidth, which could otherwise be used for useful purposes, spam filters might unfairly block emails, because of which emails are delayed. (http://www.windowsecurity.com/whitepapers/anti_spam/Impact-Reducing-SPAM-Part1.html)

Besides the fact that spam impacts businesses by wasting a lot of time, it affects the environment as well. Spam expert Richi Jennings calculated together with climate change consultant ICF International the environmental impact of spam. According to the study, the energy consumed in transmitting and deleting spam last year (62 trillion) is similar to the amount of electricity used in 2.4 million U.S. homes. (https://resources2.secureforms.mcafee.com/LP=2968).

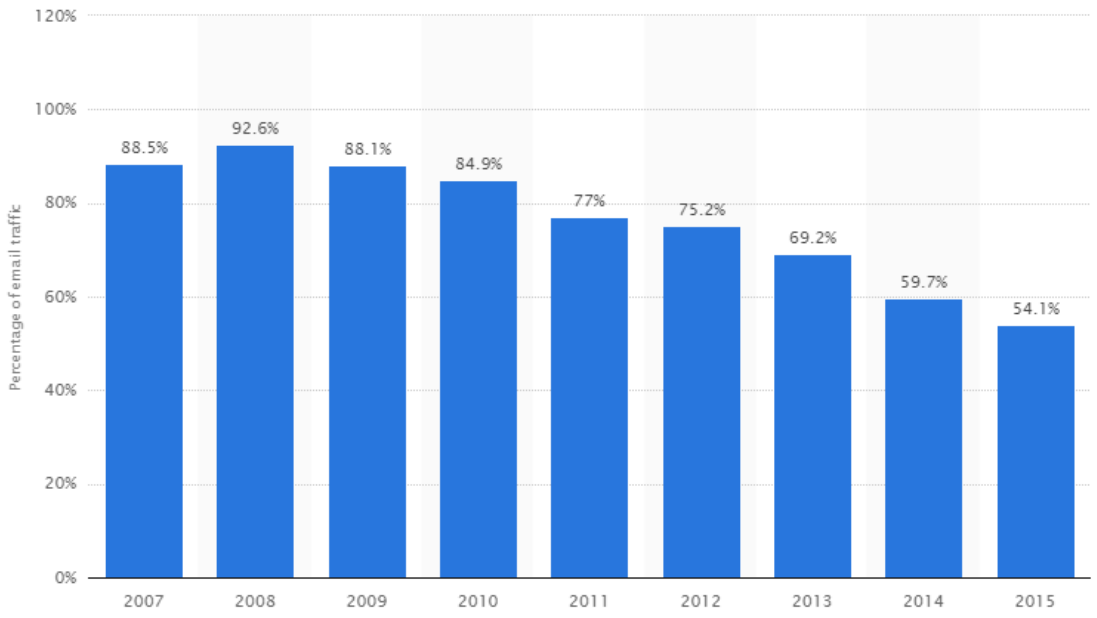


Figure 1: Global spam volume as a percentage of total e-mail traffic from 2007 to 2015, source: https://www.statista.com/statistics/420400/spam-email-traffic-share-annual/

Email is considered spam if it is unsolicited and sent in bulk. Besides junk mail from businesses to advertise goods, emails containing viruses are also considered spam. Another category of spam mail sends the receiver to websites that contain scripts to collect information for the purpose of identity theft and other criminal operations. The mail could also contain links that claim to take you of the mailing list, but in fact the intention is to verify whether the email is actively used.

2. Ideal metrics for security decision makers

The metrics for security decision makers should consist of metrics based on all four types. In practice metrics are usually based on control and a bit on vulnerability. This is because controls are closely related to the cost and are put in place to mitigate risk. Metrics based on vulnerabilities evaluate how these controls preform under threat. These metrics are deterministic in contrast to the metrics based on incidents and the prevision of loss, whom are based on events driven by attackers and are therefore stochastic. They map the losses whenever a curtain events occurs. Implementing these metrics is resource consuming and are used less.

There are a lot of decision makers dealing with the issue of spam. In the next section the most important decision makers will be mentioned with specific metrics that are useful to them.

First of all there are the users who receive the spam. The users can be separated into private and corporate users. For both the matric: ratio of spam received to average amount of spam, can help them understand if they are targeted. The matric; income loss by employees engaging in spam and income loss by false negatives is also of great value to decision makers from companies. The internet service providers can use the following metric to help the make security decisions. The amount of spam received by their clients. They can compare this matric with other Internet services providers to indicate if their security measures are adequate. A Metric that can be used by the criminals who use botnets to send spam is; success rate of spam bombs. With this metric criminals can pinpoint weaknesses in the security and exploit it. The amount of competing botnets with their relative size can be used for economic purposes. Governments are interested in metrics that calculate economic losses and the amount of damage done by engaging in spam. Last there are the email software developers, they would use metrics to indicate if their platform is targeted and has weak spots. Metrics like the amount of false positives and amount of spam send to their domain could help the software developers.

In conclusion the ideal metrics for security decision makers consist of a mix of already existing metrics from the four types and specific metrics suited for the activities different security decision makers have.

3. Existing metrics

4. Metrics from dataset

Using the dataset, a few metrics are defined. More information follows.

- Unique no. of IP addresses per botnet
- Top 10 countries per botnet
- Top 10 ISP hosting botnets
- Top 10 spam sending countries
- Botnet activity over time
- Number of botnets active per country
- Amount of countries active per botnet

5. Evaluation of the defined metrics

Methodology

Dataset was built of 10.554.552 rows and 8 columns. Before analysing the dataset, some cleaning had to be done. Firstly a random row in the middle of the dataset, containing the names of the columns, was removed. Next, all the records which did not contain a timestamp or an ASN number were removed. At the end 15.636 records were removed (which is 0.14% of all the data in the data set). The final dataset contains 10.538.915 to work with. SPSS was used to clean the data, R was used to analyse the data.

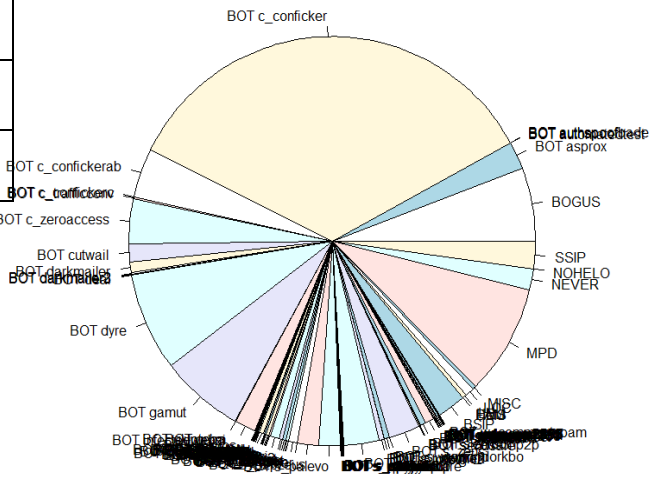
Metrics

Unique number of IP addresses per botnet

Since every IP address is unique in this dataset, this is a very straightforward question to answer. Using this R command, we were able to check which botnets contained the most IP addresses:

```
summary(spamdata$Diagnostic)
```

Rank	Botnet	#IP addresses	% of total
1	BOT c_conficker	3654641	35%
2	MPD	920926	9%
3	BOT dyre	818051	8%
4	BOT gamut	705992	7%
5	BOGUS	610905	6%
6	BOT c_confickerab	404182	4%
7	BOT c_zeroaccess	369457	4%
8	BOT s_tinba	301139	3%
9	BOT s_zeus	258807	2%
10	BSIP	242669	2%



Top 10 countries hosting Botnets

Investigating the ASN codes present in the dataset, we were able to see which providers hosted the most infected computers who were sending spam. The ASN codes are resolved using the RIPE Database at https://apps.db.ripe.net/search/query.html#resultsAnchor

```
summary(spamdata$ASN)
```

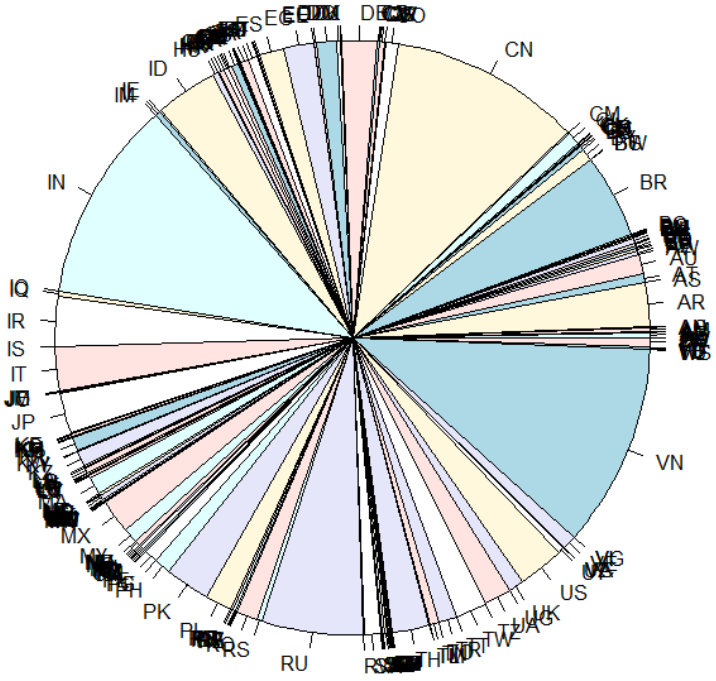
Rank	ASN Number	Name	Country	#records
1	AS4134	CHINANET-BACKBONE	CHINA	701.205
2	AS45899	VNPT-AS-VN	VIETNAM	693.424
3	AS9829	BSNL-NIB	INDIA	501.901
4	AS17974	TELKOMNET-AS2-AP	INDONESIA	286.744
5	AS7552	VIETEL-AS-AP	VIETNAM	200.668
6	AS45595	PKTELECOM-AS-PK	PAKISTAN	198.701
7	AS18403	FPT-AS-AP	VIETNAM	177.056
8	AS4837	CHINA169-Backbone	CHINA	167.542
9	AS3462	HINET	TAIWAN	143.430
10	AS8151	Uninet S.A. de C.V.	MEXICO	133.791

Top 10 countries sending SPAM

Investigating the country field in our dataset gives the following results. Achieved by using this query:

```
summary(spamdata$Country)
```

Rank	Country Code	Country	#records	% of total
1	VN	Vietnam	1162444	11%
2	IN	India	1152149	10%
3	CN	China	1098155	10%
4	RU	Russia	579619	5%
5	BR	Brazil	480780	5%
6	ID	Indonesia	357036	3%
7	IR	Iran	285827	3%
8	US	United-States	268296	3%
9	IT	Italy	252711	2%
10	PK	Pakistan	250058	2%



Top 10 countries per botnet

The 10 most popular botnets are operated from the same countries. In this table, you'll see an overview. This means that the biggest botnets are using bots located in India, but more smaller botnets use Vietnamese infected computers to perform spam attacks.

```
aggregate(Country ~ Diagnostic, summary, data=spamdata)
```

	BOT c_conficker	**MPD**	**BOT dyre**	**BOT gamut**	**BOGUS**
1	Country.IN	Country.IN	Country.IN	Country.IN	Country.IN
2	Country.CN	Country.CN	Country.CN	Country.CN	Country.CN
3	Country.PK	Country.PK	Country.PK	Country.PK	Country.PK
4	Country.RU	Country.RU	Country.RU	Country.RU	Country.RU
5	Country.IR	Country.IR	Country.IR	Country.IR	Country.IR
6	Country.US	Country.US	Country.US	Country.US	Country.US
7	Country.VN	Country.VN	Country.VN	Country.VN	Country.VN
8	Country.MX	Country.MX	Country.MX	Country.MX	Country.MX
9	Country.PE	Country.PE	Country.PE	Country.PE	Country.PE
10	Country.BR	Country.BR	Country.BR	Country.BR	Country.BR

	BOT c_confickerab	**BOT c_zeroaccess**	**BOT s_tinba**	**BOT s_zeus**	**BSIP**
1	Country.IN	Country.IN	Country.IN	Country.IN	Country.IN
2	Country.CN	Country.CN	Country.CN	Country.CN	Country.CN
3	Country.PK	Country.PK	Country.PK	Country.PK	Country.PK
4	Country.RU	Country.RU	Country.RU	Country.RU	Country.RU
5	Country.IR	Country.IR	Country.IR	Country.IR	Country.IR
6	Country.US	Country.US	Country.US	Country.US	Country.US
7	Country.VN	Country.VN	Country.VN	Country.VN	Country.VN
8	Country.MX	Country.MX	Country.MX	Country.MX	Country.MX
9	Country.PE	Country.PE	Country.PE	Country.PE	Country.PE
10	Country.BR	Country.BR	Country.BR	Country.BR	Country.BR

Botnets active per country

Different botnets are active in different countries. As the following tables show, the conficker botnet is present in every country, but some botnets are more present in one country than in every other. The p2pzeus bot is significantly more active in Italy than in the rest of the world.

```
aggregate(Country ~ Diagnostic, summary, data=spamdata)
```

	Vietnam	**India**	**China**	**Russia**	**Brazil**
1	c_conficker	gamut	c_conficker	c_conficker	c_conficker
2	BOGUS	c_conficker	dyre	MPD	MPD
3	dyre	MPD	MPD	dyre	BOGUS
4	MPD	dyre	s_tinba	BOGUS	dyre
5	c_zeroaccess	s_zeus	BOGUS	c_confickerab	c_zeroaccess
6	c_confickerab	BOGUS	SSIP	c_zeroaccess	c_confickerab
7	kelihos	BSIP	gamut	asprox	s_zeus
8	asprox	s_tinba	c_confickerab	NEVER	asprox
9	s_zeus	SSIP	NEVER	kelihos	gamut
10	gamut	asprox	c_zeroaccess	s_tinba	kelihos

	Indonesia	**Iran**	**United States**	**Italy**	**Pakistan**
1	c_conficker	c_conficker	c_conficker	c_conficker	c_conficker
2	dyre	MPD	MPD	dyre	MPD
3	MPD	dyre	BOGUS	MPD	dyre
4	BOGUS	BOGUS	dyre	c_zeroaccess	s_tinba
5	c_zeroaccess	c_confickerab	c_confickerab	c_confickerab	BOGUS
6	c_confickerab	c_zeroaccess	c_zeroaccess	BOGUS	c_confickerab
7	s_zeus	asprox	kelihos	gamut	asprox
8	gamut	BSIP	asprox	s_p2pzeus	NEVER
9	kelihos	kelihos	gamut	asprox	c_zeroaccess
10	asprox	NEVER	s_tinba	kelihos	cutwail

6. References

Not yet here, will be in the final version!