Economics of Cyber Security (WM0824TU)
Malware Domain List
Assignment Block 2

Clinton Cao, Jehan Da Camara, Sharif Jacobino and Sunwei Wang
Group - 9

September 25, 2017

1. Introduction

Malware is short for malicious software and it is used to cause harm to computer systems. A well known example of malware is a Trojan horse. This is a software that hides its true intention from the user of the system. Some other forms of malware include adware, ransomware, spyware, and much more.

All these type of malware has lead to different issues e.g. privacy breach. The privacy of the user can be compromised by having the malware read files that contain personal information of the user, or the adversary can gain access to the webcam without permission. For a Internet user it is hard to tell whether a website is malicious or not; the adversary can make their website look like the real one. The system of the Internet user can get infected with a malware without them even knowing.

A way for the user to mitigate this problem, is to install an anti-virus software. The anti-virus software can warn the user when they are visiting a malicious website. But the anti-virus software would only be able to help the user if it is updated with the latest list of malware domains. This means that the anti-virus company will have to release multiple updates for their software. If the anti-virus company fails to do so, then malware designers can use a new domain to spread their malwares or design new malwares that are unknown to the anti-virus software. This can cause the anti-virus company to lose a lot of their customers and even go bankrupt.

So from the perspective of an anti-virus company, they would have to deal with the following issues: keeping a list of domains that are considered to be malicious and being able to detect whether an arbitrary domain (that is not in the list) is malicious or not. Our main actor in this report is therefore the anti-virus companies.

The rest of the report is outlined as follows: first in section 2., we discuss on what would be ideal metrics for security decision makers. Then in section 3., we dive into some of the existing metrics that are used in practice. In section 4., we provide the metric that we have designed based on the dataset that we have received. Lastly in section 5., we evaluate the metric that we have designed.

2. Ideal Metrics

The ideal metrics for security decision metrics makers, in our case anti-virus companies, would be to cover the following aspects: the cost of security, benefit of security, and security level. The security level can be further divided into the following categories controls, vulnerabilities, incidents and prevented losses. For anti-virus companies, they want to increase the security level by being able to detect whether a domain is malicious or not. They would need metrics that would provide the following information:

- How likely a software is a malware?
- How likely a domain is a malicious domain?
- What kind of behaviours of the user is most likely to lead them getting their system infected?

- What kind of system are most likely to get infected?
- Does most malwares originate from a particular country?
- What do most people do when their system is infected?
- How hard it is to remove the malware?

3. Existing Metrics Used in Practice

For the state-of-the-art metrics that are being used in practice, we have taken a look at five different sources. Three are reports from anti-virus companies [1][5][2], one is an academic paper [4] and the last one is an article by a consultancy [3].

A way to find malware on Android is to use the software complexity metrics, as proposed by Protsenko and Müller [4]. An example of software complexity metric is Mcabe's Cyclomatic Complexity.

NoticeBored has also provided different metrics that can be used to measure and report malwares [3] An example is the awareness metrics. With awareness metrics, people are evaluated on how familiar they are with certain malwares and whether they understand the risks. Another example is the coverage of anti-virus software. This metrics evaluates how many systems are protected with an anti-virus software.

Kaspersky Lab has published a report containing different metrics [1]. A metric that they have used is by looking at the domain names where a attack has occurred and match them to the actual IP address of the domains. From here they derived in which country the attack has occurred. With this information they have compiled the top ten countries where online resources are seeded with malware. The top ten countries can be seen in Fig. 1.

Another metric that Kaspersky Lab has used is the number of attacks that is caused by a certain malware [1]. With the numbers, they derived the percentage of the attacks that is caused by a certain malware. With this information they derived the top 20 malware that were detected online. The list can be seen in Fig. 2

From the reports of the anti-virus companies that we have looked at, a common metric that is used by the anti-virus companies is the amount of new malwares that are created. The statistics can be seen in Fig. 3, Fig. 4 and Fig. 5.

4. Our Own Metric

From the dataset that was given to us, we have designed two metrics. One of the metrics is solely from the dataset and the other using external resources. These will be explained later in this section and some potential problems of these metrics will be discussed in section 5..

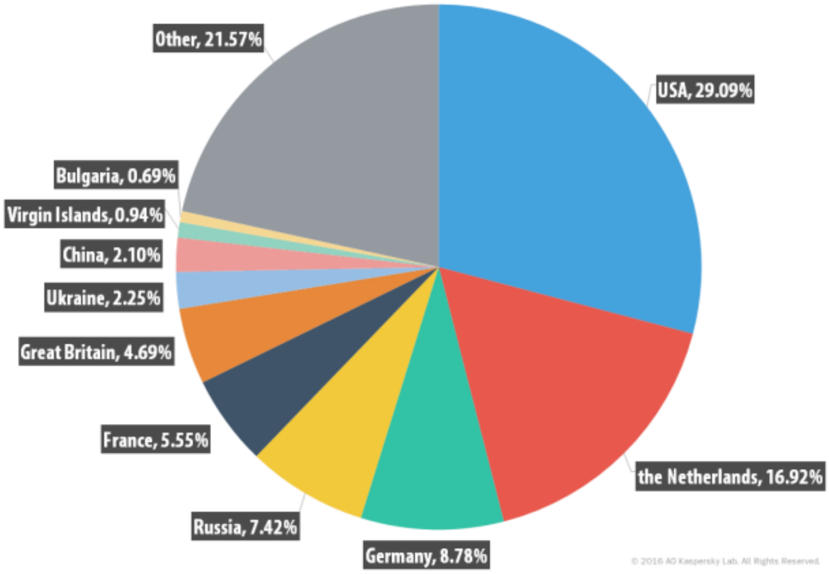
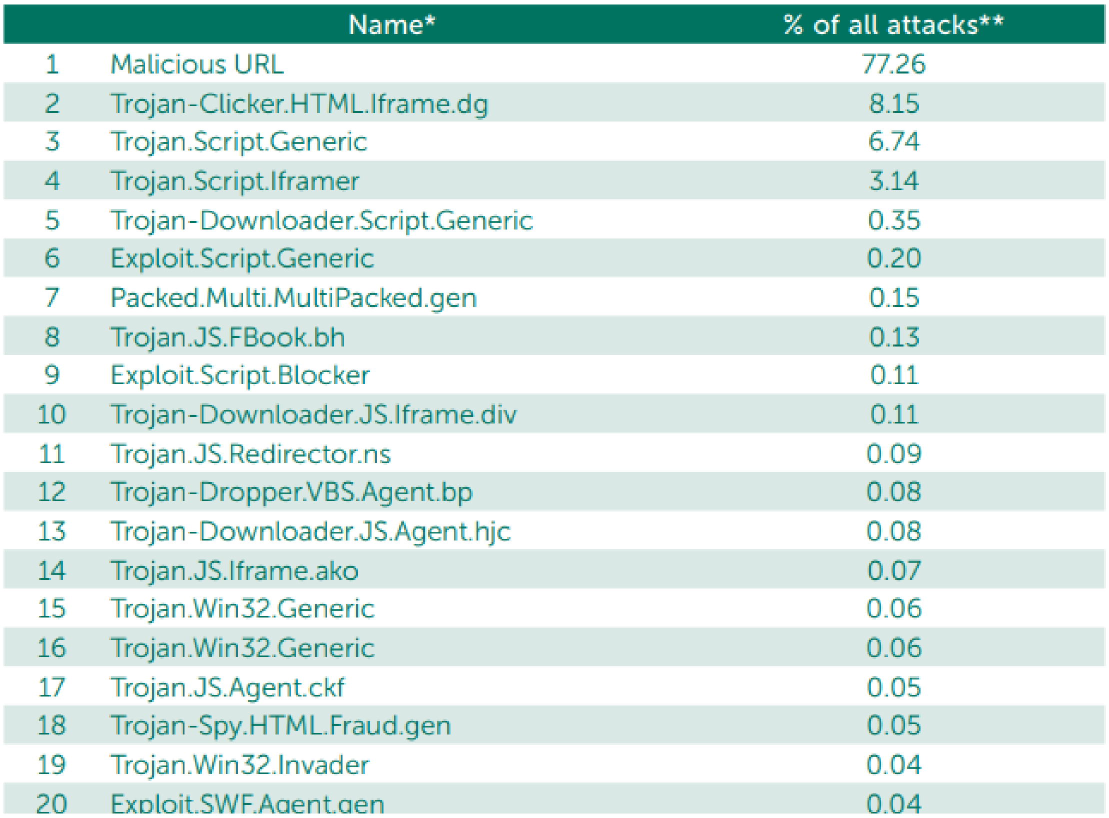


Figure 1: Top ten countries where online resources are seeded with malware [1]

4.1 Location based (Our Initial Design)

This metric was our initial design of our own metric. We found out that this is already implemented by a anti-virus company as seen in [1]. But we still include it for discussion. After looking at the dataset, the country which a domain is connected to, can be found. So given that each row has this attribute we decided to count the amount of times a country is encountered to see which location is most likely to be bounded with malware infested domains. The result are depicted in Fig. 6. It's clear that US is most likely to have malware infested domains compared to the rest of the other countries. Furthermore the sites without locations were classified as "–" or as "NaN".



	Name*	% of all attacks**
1	Malicious URL	77.26
2	Trojan-Clicker.HTML.Iframe.dg	8.15
3	Trojan.Script.Generic	6.74
4	Trojan.Script.Iframer	3.14
5	Trojan-Downloader.Script.Generic	0.35
6	Exploit.Script.Generic	0.20
7	Packed.Multi.MultiPacked.gen	0.15
8	Trojan.JS.FBook.bh	0.13
9	Exploit.Script.Blocker	0.11
10	Trojan-Downloader.JS.Iframe.div	0.11
11	Trojan.JS.Redirector.ns	0.09
12	Trojan-Dropper.VBS.Agent.bp	0.08
13	Trojan-Downloader.JS.Agent.hjc	0.08
14	Trojan.JS.Iframe.ako	0.07
15	Trojan.Win32.Generic	0.06
16	Trojan.Win32.Generic	0.06
17	Trojan.JS.Agent.ckf	0.05
18	Trojan-Spy.HTML.Fraud.gen	0.05
19	Trojan.Win32.Invader	0.04
20	Exploit.SWF.Agent.gen	0.04

Figure 2: Top 20 malwares detected online [1]

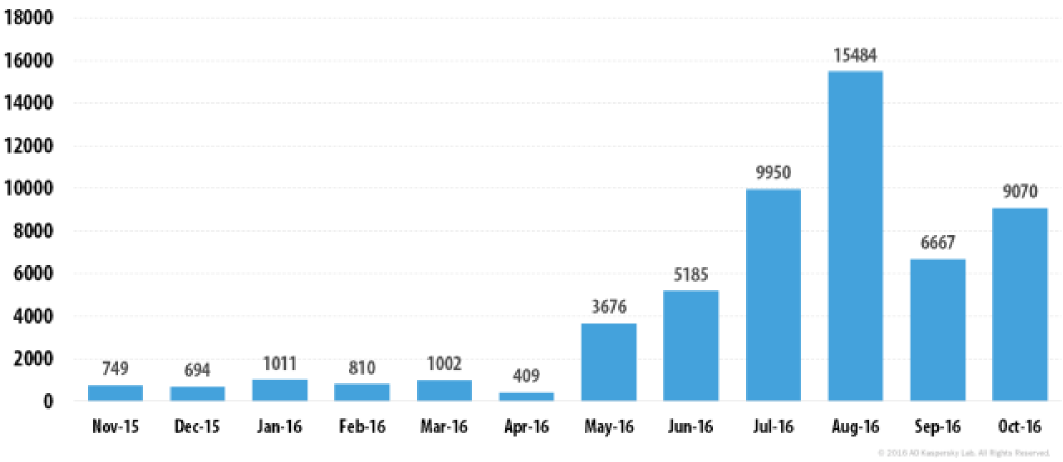
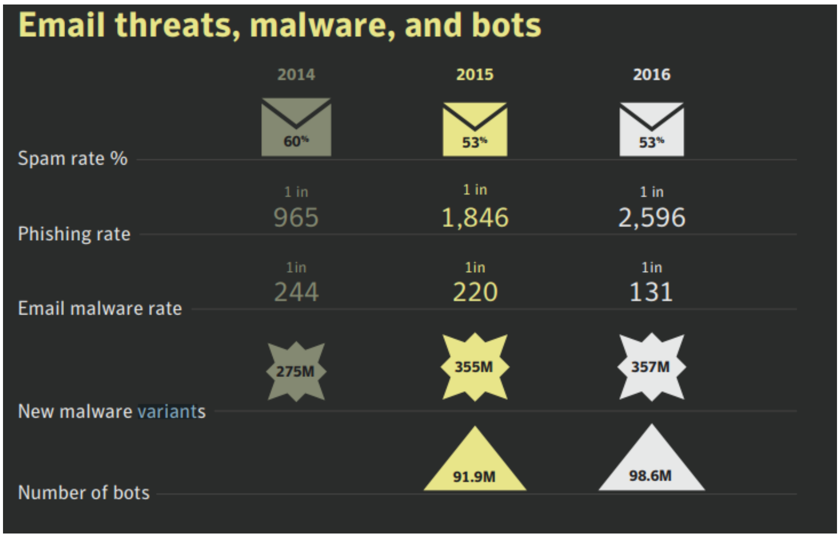


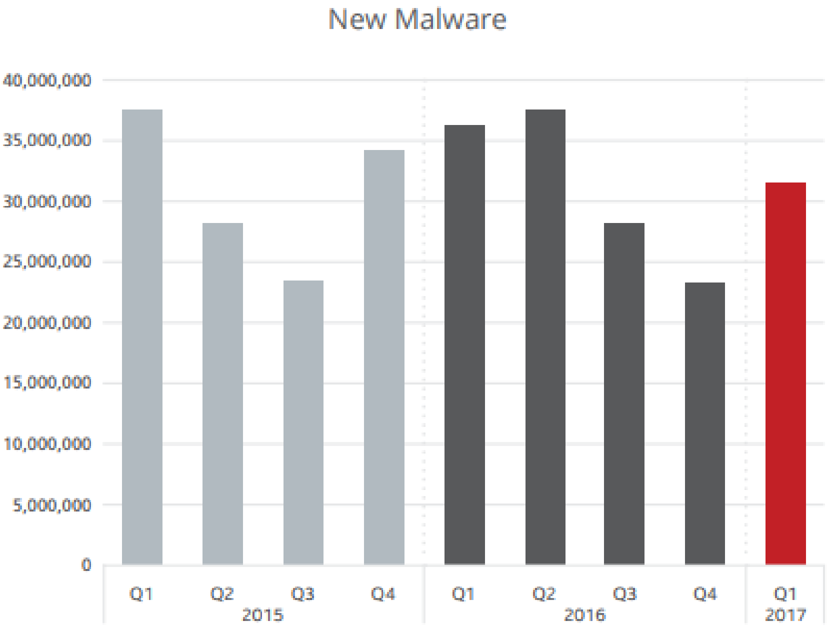
Figure 3: New malwares statistics from Kaspersky Lab [1]



Email threats, malware, and bots

	2014	2015	2016
Spam rate %	60%	53%	53%
Phishing rate	1 in 965	1 in 1,846	1 in 2,596
Email malware rate	1in 244	1in 220	1in 131
New malware variants	275M	355M	357M
Number of bots		91.9M	98.6M

Figure 4: New malwares statistics from Symantec [5]



Source: McAfee Labs, 2017.

Figure 5: New malwares statistics from McAfee Labs [2]

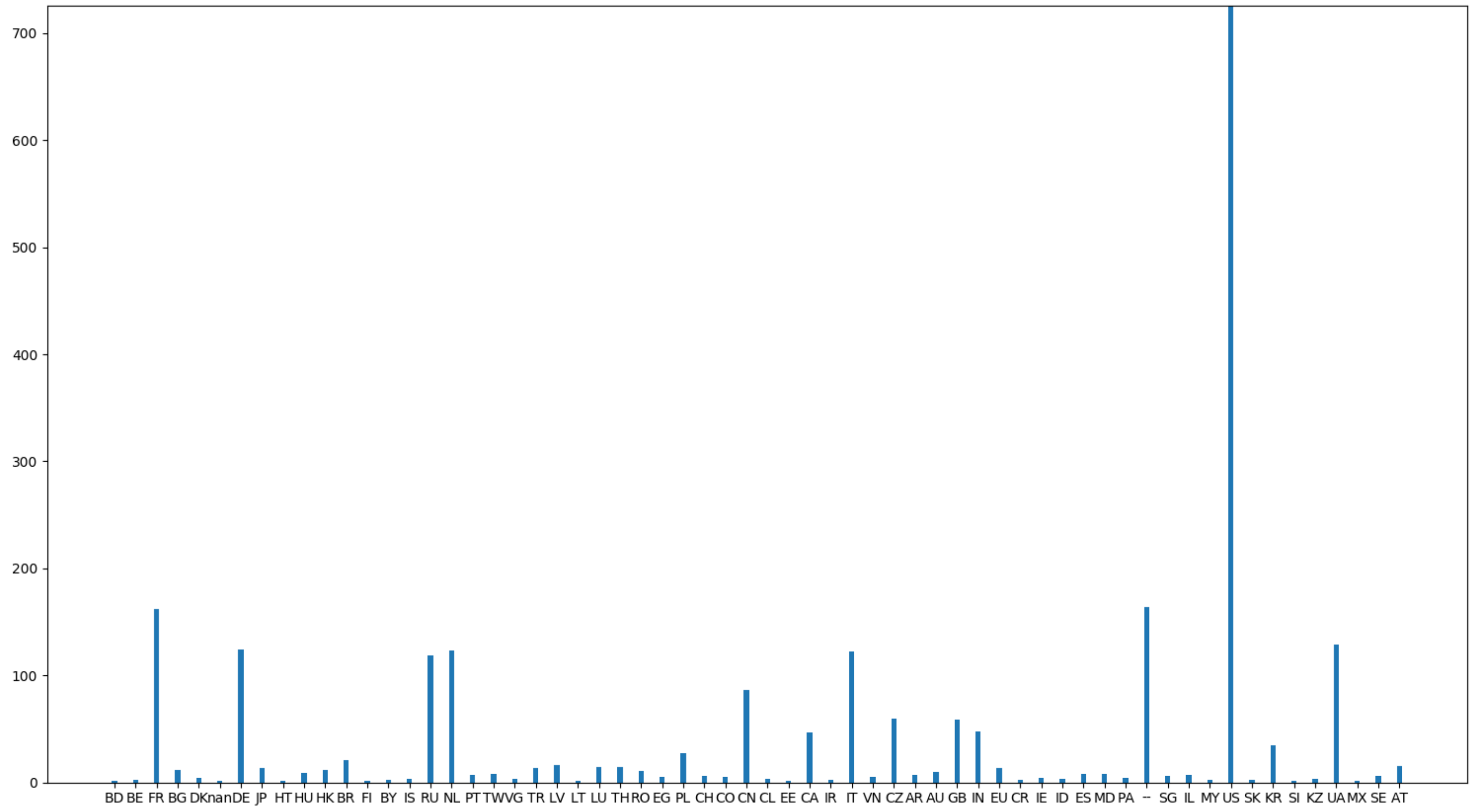
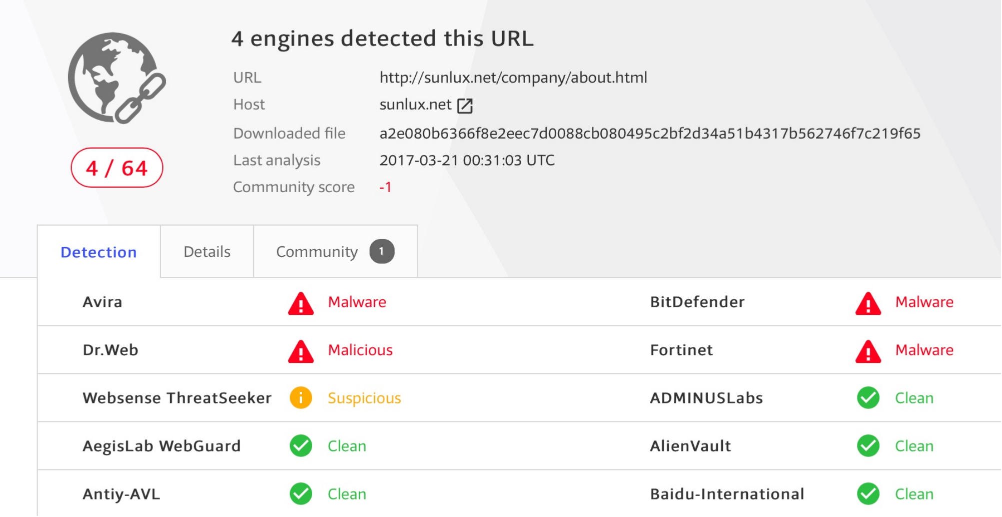


Figure 6: Amount of malware for a given domain.

4.2 VirusTotal Score

We came up with another metric that makes use of an external resource. In this case we used Virus Total. Using VirusTotal we can find a report for each URL in the dataset. The report states for different anti-virus and anti-malware software if the URL is considered malicious. We can set a threshold on the count of positive results in the report to classify a URL as malicious or not.



4 / 64

4 engines detected this URL

URL	http://sunlux.net/company/about.html
Host	sunlux.net
Downloaded file	a2e080b6366f8e2eec7d0088cb080495c2bf2d34a51b4317b562746f7c219f65
Last analysis	2017-03-21 00:31:03 UTC
Community score	-1

Avira	Malware	BitDefender	Malware
Dr.Web	Malicious	Fortinet	Malware
Websense ThreatSeeker	Suspicious	ADMINUSLabs	Clean
AegisLab WebGuard	Clean	AlienVault	Clean
Antiy-AVL	Clean	Baidu-International	Clean

Figure 7: Snapshot of a VirusTotal report (with a score of 4/64)

Fig. 8 shows that the vast majority of the malicious URLs found in our dataset is thus considered to be malicious by at least 2 anti-virus or anti-malware checks. The VirusTotal score of a URL can also be used to determine the likelihood of a URL being malicious. The higher the score the more confident you can be that the URL contains malware.

5. Evaluation of the Proposed Metric

In this section we evaluate the metrics we have designed in section 4.. For each metric, we state the problems that one could encounter using the metric, and also the limitations that we have encountered or foresee using these metrics and dataset.

5.1 Location Based Evaluation

When using the location based metric, the bar chart (see Fig. 6)is not normalized to the amount of domains or traffic that the each country generates. This is due to the fact that we did not know the amount of traffic based on the information in the dataset. Also we found that some entries in the dataset do not have a domain and/or location. This can change the result of the bar chart, some information will be missing. Furthermore just

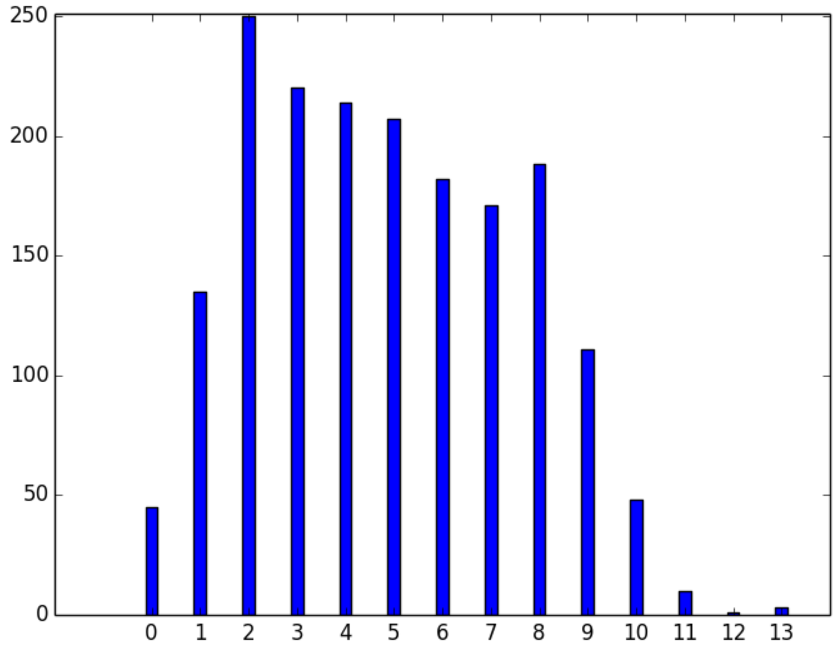


Figure 8: VirusTotal score distribution

because their is high number of malware infested site in a country, does not mean that all sites should be banned. Imagine banning all the domain in the US, we would assume that this action would be undesirable. Moreover, an adversary does not need to be in the same country to buy a top-level domain. So by banning this domain, it will just cause the adversary to buy a domain in another country.

5.2 VirusTotal Score Evaluation

If a URL has a high score on VirusTotal you can be fairly confident that the URL contains malware. However, a low a score or no result for a URL does not necessarily mean that a URL is safe. VirusTotal is dependent on the capabilities of external parties correctly classifying URLs as malicious or users reporting URLs for it to be able to construct correct reports. If a malicious URL simply has not been reported yet and it is unknown to the external parties, the VirusTotal score would be a false negative.

References

[1] Kaspersky Lab. Kaspersky Security Bulletin: Overall Statistics for 2016. https://kasperskycontenthub.com/securelist/files/2016/12/Kaspersky_Security_Bulletin_2016_Statistics_ENG.pdf.

[2] McAfee Labs. McAfee Labs Threats Report. https://www.mcafee.com/us/resources/reports/rp-quarterly-threats-jun-2017.pdf.

[3] NoticeBored. Management briefing on malware metrics. http://www.securitymetametrics.com/46_NB_management_briefing_on_malware_metrics.pdf.

[4] M. Protsenko and T. Müller. Android malware detection based on software complexity metrics. In *International Conference on Trust, Privacy and Security in Digital Business*, pages 24–35. Springer, 2014.

[5] Symantec. Internet Security Threat Report Government. https://www.symantec.com/content/dam/symantec/docs/reports/gistr22-government-report.pdf.