Finding and Analyzing Evil Cities on the Internet

Matthijs Polen, Giovane Moura, Aiko Pras

Finding and Analyzing Evil Cities on the Internet

Matthijs G.T. van Polen, Giovane C. M. Moura, and Aiko Pras

Centre for Telematics and Information Technology (CTIT)
Faculty of Electrical Engineering, Mathematics
and Computer Science (EEMCS)
Design and Analysis of Communications Systems (DACS)
Enschede, The Netherlands
matthijs@vanpolen.biz, {g.c.m.moura, a.pras} @utwente.nl

Abstract. IP Geolocation is used to determine the geographical location of Internet users based on their IP addresses. When it comes to security, most of the traditional geolocation analysis is performed at country level. Since countries usually have many cities/towns of different sizes, it is expected that they behave differently when performing malicious activities. Therefore, in this paper we refine geolocation analysis to the city level. The idea is to find the most dangerous cities on the Internet and observe how they behave. This information can then be used by security analysts to improve their methods and tools. To perform this analysis, we have obtained and evaluated data from a real-world honeypot network of 125 hosts and from production e-mail servers.

Keywords: Geographical Analysis, Bad Neighborhoods, Internet Geolocation, IP Geolocation, Spam, Network Attacks, Honeypots

1 Introduction

IP Geolocation aims to determine the Internet users' geographical location based on their IP address [1]. It has been used by industries and businesses for many purposes, including targeted advertisement (e.g., a global portal can deliver customized ads according to the user's location), fraud detection (e.g., online stores can check the physical location of a client against its billing address), media licensing (e.g, broadcasters, such as those on Hulu [2], only stream content to IPs belonging to certain countries) and even spam filtering [3].

In relation to security, most of the current Internet Geolocation analysis is done only at the country level. For example, the latest 'State of the Internet' report by Akamai shows only the top 10 countries from where attacks originated [4] [1]. Another example of country-level analysis is the daily generated map provided by Quarantainenet BV [5].

In this paper, we address the Internet Geolocation for security on a city level instead. The assumption is that countries are too big and heterogeneous so their cities/towns are expected to exhibit different behavior in relation to security. The motivation for doing so is that it would allow provide security developers with hints on how to better

[1] In this work, by originated we mean where the attack came from. We do not consider if there were other hosts controlling the attacking one.

tweak/improve their tools. Therefore, the main research question addressed in this paper is: "*Which cities in the world are responsible for most of the security incidents?*" Following the main research question, in this paper we address other sub questions:

- *Are there cities that are relatively more 'evil' than others?* Some cities might be more evil than others (that is, they generate more attacks than others), in some cases just because this city has more inhabitants, which leads to more computers and a higher probability of starting attacks. This sub research question addresses the evilness of cities taking into account its number of attackers per inhabitants. The idea is to observe if and how the evilness of a city changes according to its population.
- *Are the cities where the most attacks originated located in the countries where the most attacks originated?* If this would be the case, then filtering on a city level would not be very necessary, since filtering on a country level would yield the same results. However, if the answer to this question is 'no', then filtering on a city level might lead to more accurate results. If there are countries that cause a small number of attacks, but there is a city in that country that does cause a relatively large number of attacks, it might be a good idea to mistrust only the city and not the whole country.
- *Is there much change over time in the list of the most evil cities?* The list of countries where the most attacks come from doesn't change much, as can be observed for the map generated by Quarantainenet [5]. There is a mild variation during the day (probably due to day/night), but seen over the course of a week this list stays mostly the same. Would this behavior hold when cities are evaluated?
- *Do the evil cities change according to the type of attack?* In this question we verify whether the list of evil cities remains the same for different types of security attacks. If the list remains the same, it suggests, for example, that a list of cities where the most SSH attacks originate can be fed into a spam filter when scoring IPs based on their geolocation. If not, then a list of evil cities should be obtained per application.

The remainder of this paper is organized as follows: Section 2 discusses the related work. Section 3 presents our approach on how to find evil cities and describes our data sets. Section 4 addresses the main research question, providing analysis on the most malicious cities. Section 5 aims at the first sub research questions, in which the evilness of cities is evaluated taking into account its population. Section 6 presents results on the second sub question, evaluating whether the most evil cities belong to the most evil countries. Next, Section 7 covers the third sub question, in which the list of evil cities changes over time. After that, we evaluate how the evil cities change according to the type of attack in Section 8. Finally, Section 9 presents our conclusions and remarks for future work.

2 Related Work

Most of the current research works focusses on geographical location at the country level. For example, Jiang *et al.* [6] propose a spam filtering technique that uses country-level geographical information, which lead to a reduction of 13.9% in their experiments.

Even though they were able to reduce the number of spam messages, the authors do not describe what could happen if city-level information would be used instead of country level for filtering spam.

Sobel *et al.* [3], on the other hand, hold a U.S. patent for use of geolocation data for spam detection. It is stated in the patent that "the geolocation data may be any type of geographical information such as city, country, state or presence within a pre-selected radius of a geographical point". As a patent, the method is only described while its effectiveness is not addressed. Other non-scientific reports on the number of attacks per country also exist. For example, the Internet hosting company Akamai provides a quartely report named 'The State of the Internet' [4], which is obtained from the analysis of users that access Akamai servers (many sites, such as Hulu, BBC iPlayer and the White House use the Akamai content distribution network). In their latest report, they have observed attacks from 209 countries/regions, with the U.S. being the first one, in terms of traffic (12%). However, only 10 countries are mentioned in the report, and they do not provide analysis at city-level. Quarantainenet also provides a daily map of the countries that have attacked their honeypot infrastructure [5].

Other work is also related to ours. In a previous piece of work, Van Wanrooij and Pras [7] employ the concept of 'Internet bad neighborhood' – that is, there are certain IP blocks on the Internet more evil than others – to filter mail messages. Using data from blacklists, the authors were able to filter detect 95% of the spam messages. However, in their work, no geographical information is used. The research questions in our work address the existence of malicious cities on the Internet, instead of network blocks. In another work , Koike *et al.* [8] perform data visualization on the origin of attacks at IP block level or country level. Finally, Muir *et al.* [1] present a survey on the current Internet geolocation methods.

In the next section we describe the dataset and the approach used in our study to find evil cities on the Internet.

3 Dataset and Approach

3.1 Quarantainenet Honeypot Data

In order to answer our research questions, the first step was to obtain reliable data from real world attacks. In our case, we have obtained full non-anonymized data from Quarantainenet B.V [9], a Dutch company that develops network management and security tools and provides admission control and malware detection for their customers, including more than half of Dutch universities. Quarantainenet has a honeypot infrastructure which is distributed mostly over the Netherlands. In total, 125 machines are used for this purpose. Each has multiple IP addresses assigned to it to increase the chance of it being targeted by attackers.

Quarantainenet collects information from each honeypot and combines it into one single database. By definition, every new attack is logged. However, if the same IP address attacks a same honeypot multiple times within 48 hours, only the first incident is recorded. This is employed in order to keep the database more concise. For example, a host performing SSH dictionary attacks could be logged many times if this would not

be employed. However, for our research, this does not represent a problem, since we are interested in the IP addresses of attackers, instead of the number of attacks.

Each honeypot is able to log many different type of attacks. Among them are *SSH*-attacks, *Conficker* [10], known exploits of Microsoft Windows and others. Attacks that are as of yet unknown, are forwarded to Qnetlabs, the sister company of Quarantainenet, for further analysis. The honeypots used are passive, which means they wait for incoming connections that are then analyzed to see whether they are malicious.

The data provided by Quarantainenet is not publicly available, as it, of course, contains IP addresses of the attackers. Under Dutch law IP addresses can be *'persoonsgegevens'*, personal data. It is illegal to make them public. Therefore all IP addresses were processed automatically. Then they were discarded, leaving only an internal ID and a location.

For this research, we have evaluated a one week period from Quarantainet database – from October 29th to November 4th, 2010. During this period 25474 attacks were logged, from 23814 different IP addresses. Of these attacks, 20174 came from a form of the *Conficker* worm, a worm that targets the Windows operating system. The next largest number of attacks, namely 2052, were attacks trying to take advantage of the vulnerability in certain Windows-versions dubbed MS08-067 by Microsoft[2].

In the Sections 7 and 8 we have used different periods, which are detailed in the same sections. Next we describe the method we have employed to obtain Internet Geolocation information.

3.2 Method for obtaining IP Geolocation

After having obtained the IP addresses for the monitoring period, we have mapped them to their geographical location using GeoPlugin [11]. GeoPlugin is a free online API which uses Maxmind database [12] to resolve Internet Geolocation. They provide the following data for a particular address: city, region, area code, dma code , country name, country code, longitude, latitude, currency code, currency symbol and exchange rate. For our experiments, we needed only city and country code.

The main problem with using GeoPlugin that it relies on the accuracy of Maxmind database [12], of which numbers on accuracy are available [13]. Even though the database is not 100% precise, (Maxmind claims that their "GeoIP databases are 99.8% accurate on a country level, 90% accurate on a state level and 83% accurate for the US within a 25 mile radius"), we believe the results obtained would still hold, even though with some margin for errors.

For the sub-research question in which the number of inhabitants is taken into account, this was manually done using numbers obtained from Wikipedia. We intend as future work to develop a more automated and precise way to perform this using an online database, so that all cities can be checked, and not just the top 20. It would also be fitting to use the number of internet subscribers in a certain city, but at the time of writing this paper a database with information on the number of internet subscribers per city could not be found.

[2] Please see http://www.microsoft.com/technet/security/bulletin/ms08-067.mspx for details from Microsoft TechNet.

In the next sections we present the analysis for the research questions addressed in this paper.

4 Which cities are responsible for most of the security incidents?

To answer this question, we have evaluated the dataset described in Section 3. Table 1 shows the top 20 cities from which most of attackers were present, in absolute numbers.

#	City	Ctry	# of attacks	#	City	Ctry	# of attacks
1	Seoul	(KP)	735	11	Guangzhou	(CN)	219
2	Taipei	(TW)	618	12	Shanghai	(CN)	210
3	Beijing	(CN)	563	13	Ho Chi Minh City	(VN)	179
4	Jakarta	(ID)	362	14	Kuala Lumpur	(MY)	177
5	Buenos Aires	(AR)	351	15	Bogota	(CO)	162
6	Bangkok	(TH)	308	16	Saint Petersburg	(RU)	160
7	Moscow	(RU)	268	17	Rio De Janeiro	(BR)	152
8	Hanoi	(VN)	267	18	Caracas	(VE)	143
9	Santiago	(CL)	246	19	Bucharest	(RO)	139
10	Sao Paulo	(BR)	229	20	Chelyabinsk	(RU)	129

Table 1: Top 20 attacking cities

Analyzing this table, we can observe that, despite most of Quarantainenet's honeypot infrastructure being located in Europe, we have observed only one European city (Bucharest) among the top 20 evil cities. In addition, 9 of the top attacking cities are located in Asia and other 6 in South America, while no evil city from North America was found among the top 20. Three evil cities are located in Russia and other three in China. Figure 1 shows the attackers concentration in the world map, from the top 20 cities. As can be seen, most of attackers are located in Asia.

Taking the numbers into account, we observe that for that evaluated period most of attacks were originated in Seoul. The reasons for that might be due to Seoul has a huge number of inhabitants – Seoul metropolitan area has more than 24 million people (that is more than the population of the Netherlands, for example) combined with a high Internet penetration level [4]. Another reason could be the precision of the MaxMind database – they claim that 75% of all IP addresses that are located in South Korea are correctly placed within 25 miles of their true location [13]. However, even with such precision, Seoul would still be responsible for many attacks in absolute numbers.

On the next section we present the results for the most evil cities taking into account its number of inhabitants.

5 Are there cities that are relatively more 'evil' than others?

In the previous section we have presented the most evil cities ranked according to the number of attackers observed. In this section we present how evil they are, taking into account the number of inhabitants of each city. Table 2 shows the results.

Fig. 1: Attacks on Dutch networks from 2010-10-29 until 2010-11-04

In this table, each city is ranked according to the number of attackers per million inhabitants. In the third column, the number behind each city name shows the position the city was in Table 1, which only takes into account the absolute number of attackers. One can notice that results change when the population is taken into account. Chelyabinsk, in Russia, is the city with more attackers per inhabitants, while it was number 20 in absolute numbers. Seoul, which was the first, is ranked as 5th in this table. Chelyabinsk has ten times more attacks per inhabitant than Shanghai does, which was not clear looking only at the previous section.

However, we do see that the top four cities don't differ that much; at least not enough to call Chelyabinsk a more evil city than Buenos Aires, for example. Lower down the list the results change. For example, Shanghai, while being on the twelfth position for absolute number of attacks, is on the twentieth position on the relative table. This might be a relatively small drop, however this is because only the absolute top twenty of cities was evaluated.

Finally, looking at these results, one can conclude that there are also significant differences when evaluating cities' evilness according to the number of inhabitants.

6 Are the cities where the most attacks originated located in the countries where the most attacks originated?

Table 3 shows the 20 countries that have started more attacks to the Quarantainenet honeypots. As one can notice, China is the leading country, followed by Brazil, U.S. and Russia. This table can be compared with the the list provided by Akamai in their report [4]. In this report, they present the top 10 countries originating attacks, using a different metric: volume traffic, instead of number of incidents. In their report, 6 of

#	Ctry	City	# of attackers	# inhabitants	Att/10^6inhab.
1	(RU)	Chelyabinsk (20)	129	1078300	119.63
2	(AR)	Buenos Aires (5)	351	3050728	115.05
3	(MY)	Kuala Lumpur (14)	177	1809699	97.81
4	(TW)	Taipei (2)	618	6776264	91.20
5	(KP)	Seoul (1)	735	10208302	72.00
6	(RO)	Bucharest (6)	139	2151880	64.59
7	(CL)	Santiago (9)	246	5278044	46.61
8	(VN)	Hanoi (8)	267	6500000	41.08
9	(ID)	Jakarta (4)	362	9580000	37.79
10	(RU)	Saint Petersburg (16)	160	4661219	34.33
11	(TH)	Bangkok (6)	308	9100000	33.85
12	(CN)	Guangzhou (11)	219	7841695	27.93
13	(VE)	Caracas (18)	143	5196514	27.52
14	(RU)	Moscow (7)	268	10126424	26.47
15	(CN)	Beijing (3)	563	22000000	25.59
16	(VN)	Ho Chi Minh City (13)	179	7162864	24.99
17	(BR)	Rio De Janeiro (17)	152	6186710	24.57
18	(CO)	Bogota (15)	162	7392241	24.91
19	(BR)	Sao Paulo (10)	229	11037593	20.75
20	(CN)	Shanghai (12)	210	19210000	10.93

Table 2: Top 20 evil cities taking into account the population

the top 10 countries match our results (China, Brazil, U.S., Russia, Taiwan and Italy), despite different ordering, being the U.S. the first country in their results.

To answer the research question addressed in this section, we should compare Table 3 to Table 1. In Table 1, the most evil city is Seoul. However, The Republic of Korea (South Korea) ranks at the 7th position when we aggregate the number of attackers per country. On the other hand, China tops the list as the most evil country, while it has 3 cities among the top 20 (Beijin, Guangzhou and Shanghai). While only one European city is among the top 20 cities, 8 countries are among the top 20 most evil countries. Finally, there are countries among the top 20 that have no city among the top 20 (e.g., Italy and Spain).

This results shows that there is little correlation between the most evil cities and the most evil countries. There are countries (e.g, Italy, U.S.) that have no cities among the most evil, but when attacks are aggregated at the country level their evilness is revealed. Therefore, filtering traffic taking into account the originating country is a risky approach, and a much more precise solution is to use cities instead.

7 Is there much change over time in the list of top offenders?

In order to answer this research question, we have obtained a list of the top 20 cities for each day of the evaluated week. Table 4 shows the obtained results. Due to space constraints, we do not show the number of attackers.

#	Country	# of attackers	#	Country	# of attackers
1	China	2532	11	France	772
2	Brazil	1943	12	Germany	746
3	United States	1815	13	Ukraine	658
4	Russia	1733	14	Vietnam	622
5	Italy	1690	15	Malaysia	590
6	Spain	955	16	Japan	577
7	Republic of Korea	936	17	Thailand	454
8	Argentina	907	18	United Kingdom	448
9	Indonesia	870	19	Romania	431
10	Taiwan	832	20	Poland	413

Table 3: Top 20 attacking countries

#	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
1	Seoul	Seoul	Seoul	Taipei	Seoul	Seoul	Seoul
2	Taipei	Taipei	Taipei	Beijing	Beijing	Taipei	Taipei
3	Beijing	Beijing	Beijing	Seoul	Taipei	Beijing	Beijing
4	Jakarta	B. Aires	Jakarta	B. Aires	Bangkok	Jakarta	Jakarta
5	Bangkok	Bangkok	S. Paulo	Bangkok	B. Aires	B. Aires	B. Aires
6	Moscow	Hanoi	B. Aires	Jakarta	Jakarta	Moscow	Hanoi
7	B. Aires	Jakarta	Santiago	S. Paulo	Hanoi	Bangkok	Santiago
8	Santiago	Guangzhou	Bangkok	Hanoi	Moscow	HoChiMinh	Guangzhou
9	Hanoi	Santiago	R.deJaneiro	Santiago	Shanghai	Shanghai	Moscow
10	Shanghai	Bogota	Hanoi	Cairo	Guangzhou	Hanoi	Bangkok
11	S.Petersburg	S. Paulo	Moscow	Guangzhou	K.Lumpur	Santiago	S.Paulo
12	K.Lumpur	Moscow	Caracas	Moscow	Chelyabinsk	S. Paulo	Shanghai
13	S. Paulo	S.Petersburg	Chelyabinsk	Shanghai	HoChiMinh	K.Lumpur	Bogota
14	Guangzhou	Caracas	Bucharest	Chelyabinsk	S. Paulo	Guangzhou	S.Petersburg
15	R.deJaneiro	Shanghai	HoChiMinh	Rome	Madrid	Bucharest	Rome
16	Bucharest	K.Lumpur	S.Petersburg	Madrid	S.Petersburg	Caracas	K.Lumpur
17	Rome	R.deJaneiro	Bogota	Bogota	Bogota	R.deJaneiro	R.deJaneiro
18	Caracas	HoChiMinh	Madrid	S.Petersburg	Bucharest	Bogota	Bucharest
19	HoChiMinh	Chelyabinsk	Brasilia	Caracas	Tokyo	Cairo	HoChiMinh
20	Shenzhen	Mexico	Guangzhou	K.Lumpur	Santiago	Madrid	Caracas

Table 4: Top 20 evil cities over one week

As can be seen, Seoul is the most malicious city for 6 of the 7 days. In addition, the top 3 cities are always Seoul, Taipei or Beijing for each day. The one time Taipei was number one, the difference between the two was only two attackers. However, the mid-section was also quite stable. There were changes between cities already on the list, but only rarely did a city make the list for a day that wasn't already on the top twenty list of the entire week.

So, all in all, the top twenty list of cities is quite stable. This means that it can be used as the foundation of a set of rules for day-to-day use. There were no cases of a

city making the overall top twenty list because there was a one-day spike of traffic. This suggests that is not necessary to update the list of malicious cities on a daily basis.

8 Do the evil cities change according to the type of attack?

To investigate if the evil cities differ or not according to type of attack, we have analyzed data from two different datasets: (*i*) Quarantainenet database and (*ii*) log files from e-mail servers of the Electrical Engineering, Mathematics and Computer Science Department at University of Twente (EWI/UT). The difference between the datasets is that the first one lists IP addresses of hosts performing different types of brute force/break-in attempts (as described in Section 3), while the second one lists spamming hosts.

In order to have a more fair comparison, we have evaluated the IP addresses of malicious hosts for the same day: April 22nd, 2010. In this very day, Quarantainenet dataset had 6,797 IPs as malicious. The mail log files from EWI/UT, on the other hand, contained 240,733 spam messages from 70,546 different IP addresses. The IP addresses of both datasets were resolved to city level and then analyzed.

Table 5 presents the Top 20 evil cities for both datasets. At a first glance, one could notice that 9 out of 20 cities are found in both cases (highlighted in boldface). In fact, out of top 100 evil cities, 50 are found for both datasets, and 105 cities are present in both cases when comparing the top 200 evil cities. Even though the position in tables might change for each city, around 50% of the cities remains the same. This could be used, for example, to application level filters (such as mail filters, http proxies), in which cities would get lower scores levels for a certain type of application just by being evil for other applications. However, further research is need to investigated the feasibility of this proposal.

In the table, we can also observe that, for both cases, Seoul is the city where most of attackers and spammers come from. This is an interesting fact that shows us a different side of Seoul (as in Section 4): South Korea is usually regarded as the country with the highest level of broadband adoption, including its capital Seoul[3]. However, as shown by our results, more broadband penetration does not mean higher security levels. Seoul network administrators should be aware of this fact in order to improve security levels in their networks.

Finally, we can conclude that around 50% of malicious cities remains the same for different types of attacks, even when analyzing data from different domains.

9 Conclusions and Future Work

In this paper we have employed Internet Geolocation in order to find what are the most evil cities on the Internet. To achieve this, we have obtained IP addresses from malicious hosts from 125 honeypots maintained by Quarantainenet [5], over a period of one week.

[3] Seoul was ranked the 9th city in the world with the highest average measured connection speed by Akamai Networks [4] – an average of 14.4 Mpbs (all top 11 cities were in South Korea for the reporting period).

#	QNET-Attacks Cities	# of attackers	#	Spamming Cities	# of spammers
1	**Seoul**	190	1	**Seoul**	1759
2	Beijing	176	2	Mumbai	1488
3	**Taipei**	147	3	**Hanoi**	1364
4	Buenos Aires	107	4	New Delhi	797
5	**Jakarta**	106	5	**Ho Chi Mihn**	790
6	Santiago	87	6	Delhi	752
7	Guangzhou	75	7	Riyadh	731
8	**Sao Paulo**	75	8	**Bogota**	717
9	**Bogota**	75	9	Jiddah	682
10	Moscow	73	10	**Sao Paulo**	677
11	Saint Petersburg	70	11	**Bangkok**	677
12	**Bangkok**	56	12	Bangalore	676
13	**Hanoi**	50	13	**Taipei**	604
14	**Bucharest**	50	14	**Bucharest**	593
15	Shanghai	49	15	Madras	576
16	Rio de Janeiro	49	16	Hyderabad	525
17	**Ho Chi Mihn**	40	17	Santiago	516
18	Rome	37	18	Kiev	467
19	Caracas	35	19	**Jakarta**	429
20	Shenzhen	32	20	Cairo	428

Table 5: Top 20 evil cities for differents type of attacks

Then, we have used Geoplugin [11] which, in turn, employs the Maxmind database to obtain the geographical information associated to a particular IP address.

The main research question addressed in this paper is: "*Which cities in the world are responsible for most of the security incidents?*". As detailed in Section 4, Seoul is the most dangerous city on the Internet, having 735 malicious hosts attacking Quarantainenet infrastructure. In addition, the results have shown that just one European city is among the top 20 most evil cities on the Internet, while 9 of the top 20 evil cities are located in Asia.

The main research question was followed by four sub-questions. The first one was if "*Are there cities that are relatively more 'evil' than others?*". In this sub-question we take into account the number of inhabitants per city to determine how evil they are. We can conclude that there are indeed cities that are relatively more evil than other cities. For example, Seoul caused the highest absolute number of attacks. When the number of inhabitants is taken into account, Seoul ends up on the fifth position. The number of attacks per inhabitant is lower than for example in Taipei.

The next addressed subquestion was "*Are the cities where the most attacks originated located in the countries where the most attacks originated?*". In our results, we have observed that most of the attackers are in China, Brazil and then in the U.S.. While only one European city is among the top 20 evil cities, 8 countries are among the top 20 most evil contries. The answer to this question is that there is little correlation between the most evil cities and most evil countries. This means that using a list of evil cities to finetune firewalls or filters would yield better results than using a list of countries.

The next sub-question investigated was "*Is there much change over time in the list of top offenders?*". The top twenty of evil cities is pretty invariable. While lower on the list changes do occur over time (e.g. looking at different weeks or different days within a week), Seoul is (almost) always on top, followed by Taipei, Bejing, etc. This makes using the data easier, as there is no need to gather new data on a daily basis.

Finally, the last sub-question was if "*Do the evil cities change according to the type of attack?*". To answer this question, we have compared the Quarantainenet database against the spam log files from two mail servers from the University of Twente for a period of one day. The answer to this question is that around 50% of the cities remain the same, independently from the type of attack. This suggests that geographical information from one type of attack might be used as input to other types of attacks.

As future work, we intend to improve our approach by using a online database for the number of inhabitants per city. In addition, we intend to conduct a evaluation over a longer period of data (a year) to observe how evil cities change according to time, if there is any sort of pattern. We also intend to perform the same analysis on different datasets. Finally, the next step is to to find out if spam filters and/or firewalls can indeed be made more precise by utilizing information about evil cities. One way might be to automate the process of calculating the most dangerous cities over, for example, the last week. The data gathered from this could be incorporated into automated generating of spam rules.

Acknowledgments The authors would like to thank Quarantainenet B.V. for granting us access to their honeypot data, in special Casper Joost Eyckelhof. Also, many thanks to Marc Berenschot for his suggestions and contribution to this work.

References

1. James A. Muir and Paul C. Van Oorschot. Internet geolocation: Evasion and counterevasion. *ACM Comput. Surv.*, 42:4:1–4:23, December 2009.
2. Hulu. Hulu - What your favorites. Anytime. For free. `http://www.hulu.com`, accessed on February 2011.
3. William E. Sobel and Bruce McCorkendale. Use of Geo-Location Data for Spam Detection. U.S. Patent #7,366,919 issued Apr. 29, 2008, filed 2003.
4. Akamai. The State of the Internet, 3rd Quarter, 2010. Technical report, Akamai. Available online at: `http://www.akamai.com/stateoftheinternet/`, accessed on February 2011.
5. Quarantainenet B.V. Virus attacks. `http://quarantainenet.com/?language=en;page=infections`, accessed on February 2011.
6. Yu Jiang, Ni Zhang, and Binxing Fang. An email geographic Path-Based technique for spam filtering. In *2007 International Conference on Computational Intelligence and Security*, pages 750–753, 2007.
7. W. van Wanrooij and A. Pras. Filtering spam from bad neighborhoods. *International Journal of Network Management*, 20(6):433–444, November 2010.
8. Hideki Koike, Kazuhiro Ohno, and Kanba Koizumi. Visualizing cyber attacks using IP matrix. In *IEEE Workshops on Visualization for Computer Security*, volume 0, page 11, Los Alamitos, CA, USA, 2005. IEEE Computer Society.
9. Quarantainenet B.V. Quarantainenet. `http://quarantainenet.com/`, accessed on February 2011.

10. Microsoft. Computer Worms - Conficker — Microsoft Security. http://www.microsoft.com/security/pc-security/conficker.aspx, accessed on February 2011.
11. Geoplugin. Geoplugin. http://www.geoplugin.com, accessed on February 2011.
12. Maxmind. Maxmind. http://www.maxmind.com/, accessed on February 2011.
13. Maxmind. Geolite city accuracy. http://www.maxmind.com/app/geolite_city_accuracy, accessed on February 2011.