Empirical research on IP blacklisting

Christian Dietrich
Institute for Internet Security
University of Gelsenkirchen
45877 Gelsenkirchen
Germany

Christian Rossow
Institute for Internet Security
University of Gelsenkirchen

Abstract

This paper describes two innovative analysis methods for IPv4 address sets such as antispam blacklists. First, the contents analysis provides means of measuring key properties of any set of IPv4 addresses as well as revealing relationships between such sets. Second, the behavior analysis defines behavioral attributes of querying addresses and requested addresses. Furthermore, the behavior analysis provides an insight into the global email communication. These two analysis methods are applied and the empirical results are presented as part of this paper.

1 Introduction

IP blacklisting in the context of anti-spam describes collecting IP addresses in a list and prohibit any email communication attempts initiated from these addresses. Usually IP blacklisting is the first level of spam protection at email servers. With the help of blacklisting in particular big Email Service Providers filter up to 80% of their incoming SMTP connections to email systems ([1], [2], [3]). In order to choose which IP blacklists are to be used, email operators either depend on hearsay and on their own experience or set up their own blacklist. The same applies to white- and bogonlists.

2 Motivation

To the best of our knowledge, key properties of well-known blacklists as well as the relationship such as intersections between different black-, white- and bogonlists have not yet been subject to research.

Furthermore, little is known about the behavior of blacklists. Only few research results concerning the behavior of IP blacklists exist ([4], [5], [6]). Ramachandran and Feamster analyze in [4] the network-level behavior of spammers by looking at a large spam sinkhole between August 2004 and December 2005 and correlating this with BGP routing information, blacklist lookups, traces from a known botnet and traces of legitimate email. DNS-based IP blacklists play a role in this study in that about 80% of the received spam was listed in at least one of eight blacklists. However, hardly any information is given concerning the eight blacklists that were used and their behavior or contents. Ramachandran et al. present in [5] so-called behavioral blacklisting, a technique used to classify email senders based on their sending behavior rather than on their IP address. The evaluation is based on email logs for over 115 domains. In this paper, the behavior of email from at least 1.3 million different senders to about 10,000 different receivers, corresponding to more than 10,000 target domains, has been analyzed.

Ramachandran et al. studied in [6] the behavior of a blacklist in order to detect botnet membership. They observed a mirror of a well-known blacklist for a 45-day period in November and December 2005. In this paper, we analyzed the behavior based on measurements of a total of 8 months during July 2007 and March 2008. Furthermore, we present for the first time an evaluation of blacklist usage statistics from the server's perspective as well as activity periods of email sources. Jung and Sit analyze DNS blacklist usage from the client's perspective ([7]) based on measurements in 2000 and 2004. Moreover, in this paper, key facts as well as the contents of 11 blacklists, one whitelist and a bogon list are analyzed for the first time.

2.1 Contribution

In this paper, we present for the first time analysis methods and empirical results that reveal key properties of and intersections between 13 IP black-, white and bogon lists. These facts can be used by researchers, black- and whitelist operators and email operators.

In Section 3, we present our analysis methods. Section 4 contains the evaluation results of applying the contents analysis method to 13 lists and the behavior analysis method to one blacklist. Section 5 provides a discussion of the results as well as possible future work before we conclude this paper in section 6 with a brief summary.

3 Analyses

3.1 Contents analysis

We developed and applied two analysis methods. Subject of the – as we define it – contents analysis is a full set of IP addresses at a given point in time, such as for example the contents of an IP blacklist. The contents analysis focuses on properties of one such set of IP addresses and the comparison of different sets of IP addresses. Moreover, it aims at monitoring the mutation of one or more sets over time.

- Properties of an IP list may be but are not limited to one of the following:
- Covered net range (the amount of IPv4 address space covered by the list)
- Number of entries (the number of entries of the list, single IPv4 addresses as well as IPv4 net ranges)
- Percentage of total IPv4 address space (percentage of the covered net range among the total theoretical IPv4 address space)
- Percentage of advertised IPv4 address space (percentage of the covered net range among the advertised IPv4 address space)

An overlapping entry of two sets means that a certain IP address or IP address range was listed in both of the two lists at a certain point in time.

Black-, white- and bogonlists change over time, i.e. new addresses get on the list, others may be removed.

3.2 Behavior analysis

Not only contents of blacklists are of interest. We define the term behavior analysis as looking into how clients request information of a list from a server. Most lists make use of the DNS protocol in order to query a black- or whitelist. Thus, our behavior analysis refers to DNS-based IP lists. A client can be any host that looks up a certain IP address in the list. Usually email servers lookup the source IP addresses of incoming SMTP connections (see figure 1). Apart from that, bots may look up their own addresses in order to find out whether they are listed or not as described in [6].

Two main sets of IP addresses result from the behavior analysis: The set of IP addresses that are requested, so-called requested IP addresses, and the set of source IP addresses of clients that perform the queries.

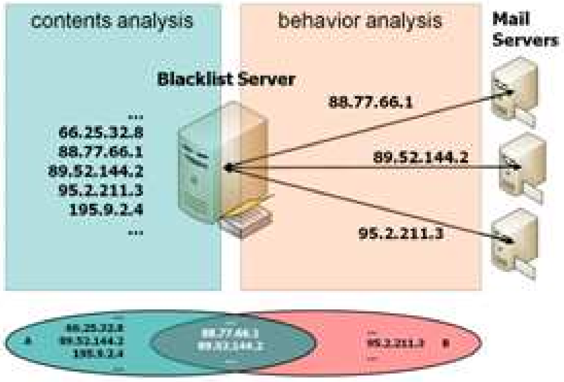


Figure 1: Contents And Behavior Analysis

For each request from a client to the blacklist server, the source IP address can be extracted. This leads to a set of source IP addresses of the users of the blacklist.

Figure 1 displays two different underlying sets between contents and behavior analysis. On the left side, set A represents the set of listed IP addresses, whereas on the right, set B is the set of requested IP addresses. The intersection of the two sets, A ∩ B, is the set of requested IP addresses that were on the list at the time they were queried. A \ B is the set of listed IP addresses that were never queried. B \ A is the set of requested IP addresses that were not listed at the time they were queried.

The behavior analysis reveals at least the following properties:

- Total number of requests to a list over a period of time
- Total number of positive and negative responses over a period of time
- The ratio between total requests and positive responses, so-called hit rate, over a period of time
- The number of distinct requested IP addresses over a period of time
- The number of distinct source IP addresses over a period of time

4 Evaluation

Both contents analysis and behavior analysis have been applied. We present the results in this section.

4.1 IP list contents analysis

In March 2008, contents analysis was applied to 11 blacklists, one whitelist and one bogon list. Results are grouped in list properties and intersection between the lists.

4.1.1 List properties

Very abstract information on IP lists, such as the number of entries, can reveal first insights into the concepts and policies of blacklists. Especially the amount of IP address space covered is of interest, since a single entry can be a net range of multiple IP addresses. Furthermore the amount of listed addresses among the whole theoretical IPv4 space can be computed. In practice, the amount of assigned IP addresses, the so-called advertised IP address space, plays an important role. Building a ratio between advertised space and list sizes shows how much is actually known about the IP address space. Table 1 displays this information for all lists taken into account during our research.

Table 1: List properties of 11 blacklists, 1 whitelist and bogon ranges

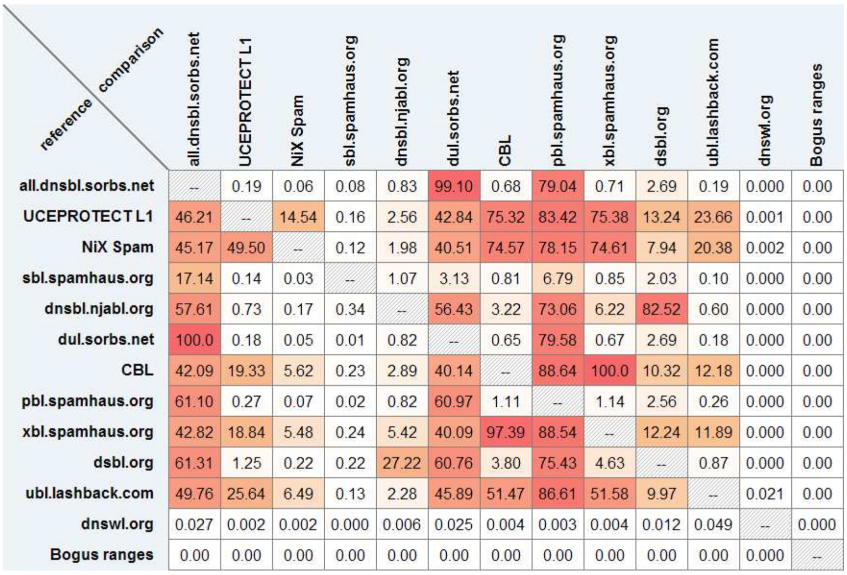
LIST	NET RANGE	ENTRIES	COVERAGE	COVERAGE ADV.
all.dnsbl.sorbs.net	313609137	1099179	7.302%	16.979%
UCEPROTECT L1	1300216	1300216	0.030%	0.070%
NiX Spam	382085	382085	0.009%	0.021%
sbl.spamhaus.org	1456104	5091	0.034%	0.079%
dnsbl.njabl.org	4537328	4537328	0.106%	0.246%
dul.sorbs.net	310801329	472915	7.236%	16.827%
CBL	5066714	5066714	0.118%	0.274%
pbl.spamhaus.org	405706490	938807	9.446%	21.965%
xbl.spamhaus.org	5202469	5202469	0.121%	0.282%
dsbl.org	13755714	13755714	0.320%	0.745%
ubl.lashback.com	1199454	1199454	0.028%	0.065%
dnswl.org	521323	22322	0.012%	0.028%
Bogus ranges	1276379392	29	29.718%	-

Next to the plain number of entries, the covered net range of those lists is of interest. 7 out of those 13 lists list single IP addresses only, whereas the remaining 6 lists also list entire net ranges in a single entry. This design decision highly affects the eventual size of the set containing all IP addresses listed. An outlier is given with a list of bogus ranges, where only 29 entries cover more than 1.27 billion of single IP addresses. On the other hand, NiX Spam and others define in their policy to take into account single addresses only.

Another important aspect of the listing policies is the type of addresses that enter the list. As such, Spamhaus' PBL and SORBS' DUL both try to list addresses of home users, that should not run their own mailserver and are threatened to be used as spamming bots. Those listings usually cover big net ranges used by providers to assign to their dial-up customers. A union of all IP addresses of Spamhaus' PBL reveals that up to 22% of the advertised IP address space is listed in this list. On the other hand, for example CBL analyses email traffic and uses spamtraps to build up metrics based on single IP addresses. Building the biggest set of single IP addresses, dsbl.org lists ~0.7% of the advertised IP address space. However, a high coverage is certainly not per se an indicator for the quality of a list.

4.1.2 Intersections

Furthermore, a comparison matrix shows the amount of intersections between different lists. Figure 2 gives the percentages about which amount of IP addresses listed in list A (row) is covered by list B (column). The higher



comparison reference	all.dnsbl.sorbs.net	UCEPROTECT L1	NiX Spam	sbl.spamhaus.org	dnsbl.njabl.org	dul.sorbs.net	CBL	pbl.spamhaus.org	xbl.spamhaus.org	dsbl.org	ubl.lashback.com	dnswl.org	Bogus ranges
all.dnsbl.sorbs.net	--	0.19	0.06	0.08	0.83	99.10	0.68	79.04	0.71	2.69	0.19	0.000	0.00
UCEPROTECT L1	46.21	--	14.54	0.16	2.56	42.84	75.32	83.42	75.38	13.24	23.66	0.001	0.00
NiX Spam	45.17	49.50	--	0.12	1.98	40.51	74.57	78.15	74.61	7.94	20.38	0.002	0.00
sbl.spamhaus.org	17.14	0.14	0.03	--	1.07	3.13	0.81	6.79	0.85	2.03	0.10	0.000	0.00
dnsbl.njabl.org	57.61	0.73	0.17	0.34	--	56.43	3.22	73.06	6.22	82.52	0.60	0.000	0.00
dul.sorbs.net	100.0	0.18	0.05	0.01	0.82	--	0.65	79.58	0.67	2.69	0.18	0.000	0.00
CBL	42.09	19.33	5.62	0.23	2.89	40.14	--	88.64	100.0	10.32	12.18	0.000	0.00
pbl.spamhaus.org	61.10	0.27	0.07	0.02	0.82	60.97	1.11	--	1.14	2.56	0.26	0.000	0.00
xbl.spamhaus.org	42.82	18.84	5.48	0.24	5.42	40.09	97.39	88.54	--	12.24	11.89	0.000	0.00
dsbl.org	61.31	1.25	0.22	0.22	27.22	60.76	3.80	75.43	4.63	--	0.87	0.000	0.00
ubl.lashback.com	49.76	25.64	6.49	0.13	2.28	45.89	51.47	86.61	51.58	9.97	--	0.021	0.00
dnswl.org	0.027	0.002	0.002	0.000	0.006	0.025	0.004	0.003	0.004	0.012	0.049	--	0.000
Bogus ranges	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000	--

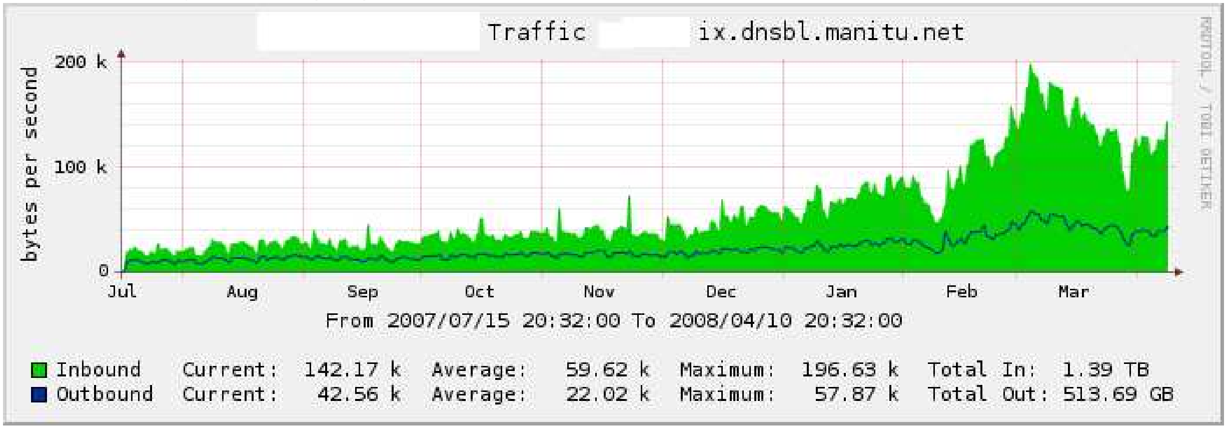
Figure 2: IP Address List Intersection Matrix

the value, the darker is the background of the table cell.

The red-coloured cells clearly reveal the relationship between blacklists of high intersection such as Spamhaus blacklists. It is obvious that Spamhaus' XBL covers CBL completely (100%). A huge part of the CBL is contained in Spamhaus' PBL (86%). On the contrary, the blacklist NiX Spam does not cover much of other blacklists due to its small size (~ 400,000 entries).

Combining two blacklists in order to fight spam is much more efficient if the two lists have a low intersection value. Thus, in our eyes, it does not make sense using the CBL in addition to Spamhaus' XBL. On the other hand, it is sensible to use NiX Spam in combination with dsbl.org, because they hardly overlap.

The last two columns and rows of the matrix play a special role. Being the only public whitelist considered in our research, dnwsl.org shows minor intersections with given blacklists. As one considers the goal of whitelists, namely preventing legitimate mails from getting blocked based on blacklist decisions, the values are reasonable. On average each blacklist is covered to ~0.001% by the whitelist, which in theory is one wrong entry out of 10,000. Practically one can neither assume the completeness of the whitelist, nor its correctness.



Finally considering bogus net ranges does not show any noteworthy intersection with blacklists. Only SORBS has little, but negligible intersections with not routable IP addresses. It can be considered to also block any SMTP or even IP traffic coming from those bogus net ranges. Other research deals with detecing spam based on further network-level properties ([4]).

4.2 Behavior analysis

The analysis of the blacklist behavior was performed between July 2007 and March 2008 on one DNS slave for the NiX Spam blacklist. The NiX Spam blacklist cluster consists of ten DNS servers. DNS requests to the blacklist are served in a round robin fashion. The NiX Spam has never been subject to any kind of empirical analysis before. NiX Spam was invented by Bert Ungerer, an editor of the German computer magazine iX.

4.2.1 Usage

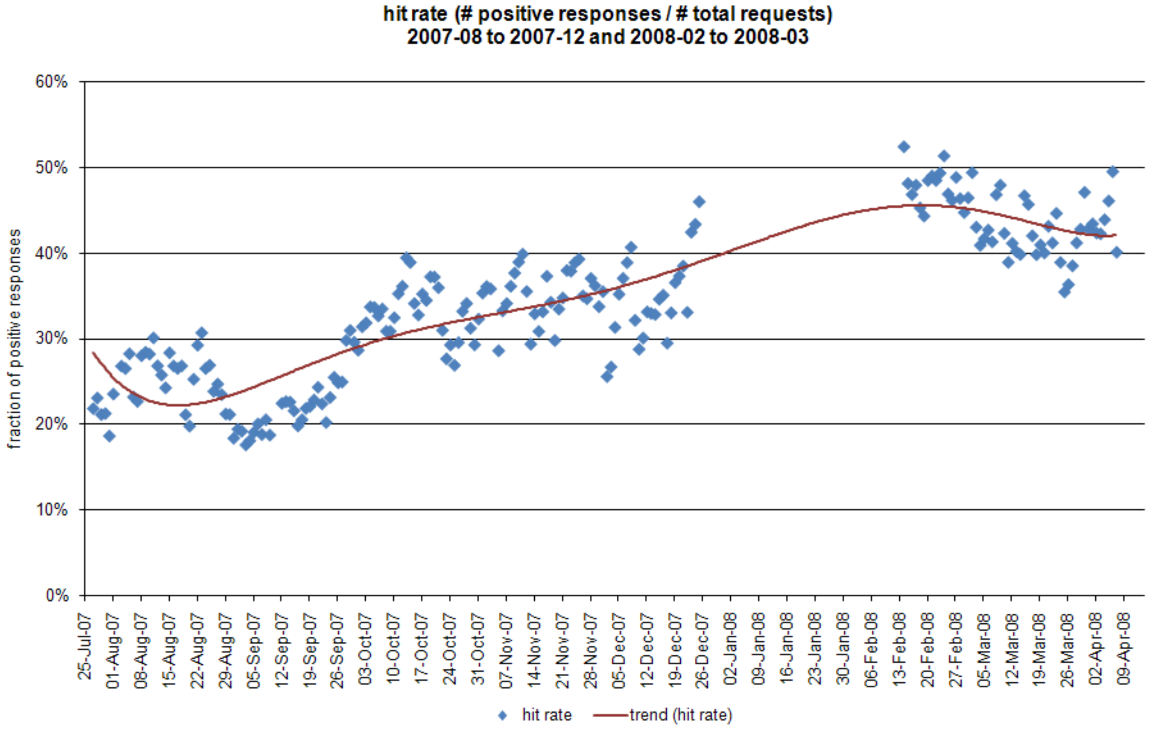
Total requests

As part of the behavior analysis we measured the total number of requests as well as the total number of responses that were processed. In July 2007, our NiX Spam blacklist mirror analyzed on average 5.5 million requests per day. 5 months later, in December 2007 the number has increased to 9.5 million requests per day. Due to the immense increase in requests, our analysis setup had to be changed. Thus, the behavior analysis was interrupted between December 25th 2007 and February 14th 2008. Since February 2008 we analyze on average 16.6 million requests per day. Meanwhile the traffic to our blacklist slave has been measured without interruption. It show a significant increase.

Positive responses and hit rate

In the context of behavior analysis the number of positive responses can be compared to the total number of requests. The ratio of total requests to positive responses is called hit rate.

The hit rate of NiX Spam nearly doubled during the 8-month-period. In July 2007, it started at about 23% and finally reached 44% on average in February and March 2008.



A high hit rate means that many addresses requested from the list are actually listed and – in case of the NiX Spam blacklist – known as spam sources. This information is only reliable if the list provides a low false positive rate. In order to assess a blacklist, not only the hit rate, but also the false positive rate should be considered.

Interpreting source and requested IP addresses

Two further values for a blacklist are the number of requesting IP addresses (i.e. the source IP addresses of the query) as well as the number of requested IP addresses. The number of source IP addresses gives an impression of the number of users of a blacklist. In case of the NiX Spam, we detected on average about 11,000 different source IP addresses per day throughout the whole 8-month-period. It is important to note that this number is a lower bound due to DNS response caching. The number of different source IP addresses increases slightly towards the end of the analysis period. Surprinsingly, however, it does not increase as much as the total number of requests or the traffic.

During February and March 2008, on average 1,529,054 distinct IP addresses have been requested on NiX Spam per day. Among these, 226,771 distinct addresses (14.83%) were on the list at the time they were queried. The NiX Spam blacklist had a total of 440,662 addresses listed on average. This shows that about 51.5% of the listed addresses were queried per day.

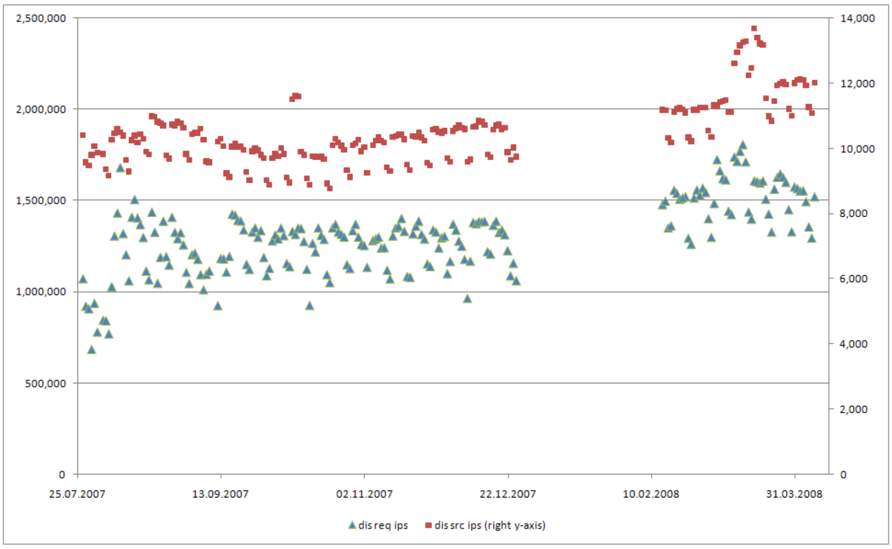
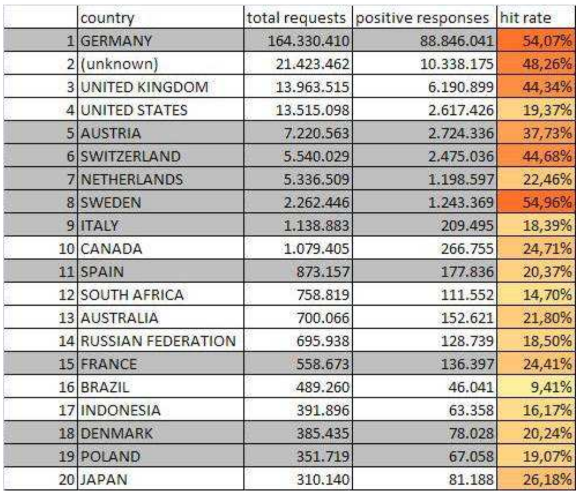


Figure 3: IP Address List Intersection Matrix

4.2.2 Mail receivers



	country	total requests	positive responses	hit rate
1	GERMANY	164.330.410	88.846.041	54,07%
2	(unknown)	21.423.462	10.338.175	48,26%
3	UNITED KINGDOM	13.963.515	6.190.899	44,34%
4	UNITED STATES	13.515.098	2.617.426	19,37%
5	AUSTRIA	7.220.563	2.724.336	37,73%
6	SWITZERLAND	5.540.029	2.475.036	44,68%
7	NETHERLANDS	5.336.509	1.198.597	22,46%
8	SWEDEN	2.262.446	1.243.369	54,96%
9	ITALY	1.138.883	209.495	18,39%
10	CANADA	1.079.405	266.755	24,71%
11	SPAIN	873.157	177.836	20,37%
12	SOUTH AFRICA	758.819	111.552	14,70%
13	AUSTRALIA	700.066	152.621	21,80%
14	RUSSIAN FEDERATION	695.938	128.739	18,50%
15	FRANCE	558.673	136.397	24,41%
16	BRAZIL	489.260	46.041	9,41%
17	INDONESIA	391.896	63.358	16,17%
18	DENMARK	385.435	78.028	20,24%
19	POLAND	351.719	67.058	19,07%
20	JAPAN	310.140	81.188	26,18%

Figure 4: Mail Receivers By Country

When looking at regional characteristics of a blacklist, NiX Spam reveals that more than two thirds of all requests to the blacklist originate from Germany. More than a half of all requests issued from Germany have a positive response, which equals to a hit rate of 54% for German users. Only NiX Spam users from Sweden have the highest hit rate of 55%. On the other side, the US also uses the NiX Spam list ranked 4th, but the hit rate for US American users is surprisingly low with 15%. Rows colored in grey show European countries (as will also be the case in the following tables).

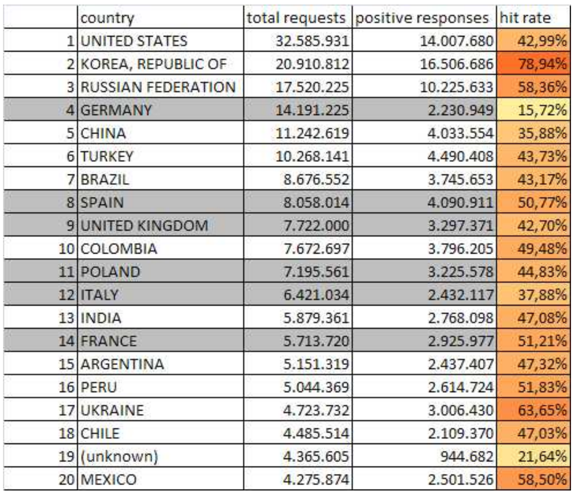
4.2.3 Mail sources (requested IP addresses)

By Autonomous System

All NiX Spam users together form a representative group of email receivers. In order to analyze the email sources, we assigned the source Autonomous System to each requested IP address in a 2-week-period in February 2008. This results in a view on the email sending activity of certain Autonomous Systems. Whereas the hit rate for requested IPs from the AS 4766 (Korea Telecom) is considerable high with 80%, IP addresses of AS 3320 (Deutsche Telekom) are listed in only 23% of all requests.

By Country

Grouping requested IP addresses by countries provides another point of view. Interestingly, IP addresses from the US form the group of most active email sources during the 2-week-period in February 2008. With a hit rate of 43%, around 14 million requests caused a blacklist hit and signaled a bad sender reputation. Still, countries as the Republic of Korea follow nearby with a considerably higher hit rate of 79%. Although IP addresses from Germany were requested very often (14,191,225), the number of positive responses (2,230,949) is quite low (hit rate 15.72%). Correlating this with the geographical distribution of users (most NiX Spam users come from Germany, see 4.2.2) this result does not surprise. Many legitimate email senders are leading to high request rates and keep the positive responses low. On the other hand, mass email senders from other countries will certainly not that often contact German users, and thus cannot decrease the hit rate of countries they are located in.



	country	total requests	positive responses	hit rate
1	UNITED STATES	32.585.931	14.007.680	42,99%
2	KOREA, REPUBLIC OF	20.910.812	16.506.686	78,94%
3	RUSSIAN FEDERATION	17.520.225	10.225.633	58,36%
4	GERMANY	14.191.225	2.230.949	15,72%
5	CHINA	11.242.619	4.033.554	35,88%
6	TURKEY	10.268.141	4.490.408	43,73%
7	BRAZIL	8.676.552	3.745.653	43,17%
8	SPAIN	8.058.014	4.090.911	50,77%
9	UNITED KINGDOM	7.722.000	3.297.371	42,70%
10	COLOMBIA	7.672.697	3.796.205	49,48%
11	POLAND	7.195.561	3.225.578	44,83%
12	ITALY	6.421.034	2.432.117	37,88%
13	INDIA	5.879.361	2.768.098	47,08%
14	FRANCE	5.713.720	2.925.977	51,21%
15	ARGENTINA	5.151.319	2.437.407	47,32%
16	PERU	5.044.369	2.614.724	51,83%
17	UKRAINE	4.723.732	3.006.430	63,65%
18	CHILE	4.485.514	2.109.370	47,03%
19	(unknown)	4.365.605	944.682	21,64%
20	MEXICO	4.275.874	2.501.526	58,50%

Figure 5: Mail Sources By Country

Distribution of activity periods of requested IPs

The distribution of activity periods of requested IP addresses reveals that for NiX Spam only 8% of all requested IP addresses last longer than 3 days. In other words: 92% of all requested IP addresses are requested during a period of 3 days only. The numbers become even more obvious when looking at a single day: 3 of 4 requested IP addresses only last for one day.

Hit rate as a function of activity

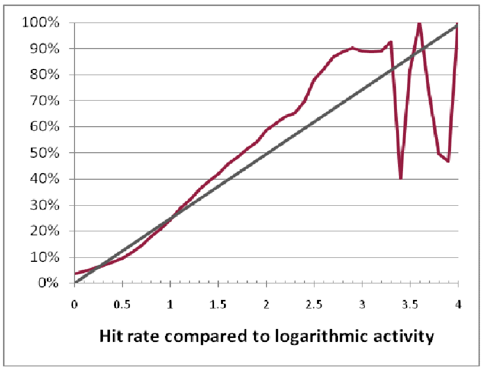


Figure 6: Hit rate of IP addresses compared to their mail activity

Figure 6 shows the mean hit rate of IP addresses at the NiX Spam blacklist with an ascending activity. More specific, IP addresses that are requested often from the list have a high activity. The activity was measured in DNS requests per IP address coming in at the blacklist slave as a logarithm to base 10. In other words, 10 to the power of activity is the number of DNS requests for an IP address.

Obviously, the tendency shows that IP addresses that are more frequently requested are more likely to be listed. However, a few legitimate mailers with high activities disturb this tendency. Sources of their solicited emails are usually not contained in the blacklist. It is left open to discussion, whether activities of IP addresses can be used as criteria for building a reputation, as discussed later in chapter 5.2.

5 Discussion

5.1 Conclusions from list intersections

In section 4.1.2 intersections between different lists were shown. Out of question it does not make sense to use two blacklists that show an intersection of 100%, since then one of the lists is contained in the other list. Implying the reverse might not be true. In other words, if no intersection consists, an anti-spam appliance may perform better. However, it is not guaranteed that spam will actually be sent from the net ranges that were gained by combining two lists with low intersections. As our experience has shown, some combinations do not increase the hit rate, due to the fact that most spam comes from addresses of the intersection of both lists. Still, the intersection matrix shows relations between lists and gives good indications whether specific combinations might be sensible. In addition it can be of help for blacklist operators to check their intersections with white- and bogonlists.

5.2 Machine learning based on behaviour

As it was shown in chapter 4.2.4.2, more active senders are more likely to be listed on a blacklist. Only a few outliers with a very high activity, namely legitimate email servers, prevent us from generally giving an IP address a reputation based on its behavior observed by email receivers all over the world. It is open to discussion, whether automating this frequency analysis is doable and a concept of a fully-automated blacklist with such data as input would be successful.

5.3 Future work

Interpreting results we gained so far gave us ideas for further research possibilities. To name but a few:

- Currently we set the hit rate in relation to AS, countries or activity of IP addresses. However, it is also interesting to correlate it with other possible parameters, such as the source port of requests
- As it was discussed before, some ISPs tend to automatically disconnect their clients after a specific time. Thus it would be of great interest to link e.g. countries with the activity periods of IP addresses requested from it. Assuming that usually bots send out emails, this would show deviating period lengths for those providers forcing a reconnect.
- Until now, the behavior analysis is only based on requests that were correctly sent to our DNS slave of the NiX Spam blacklist. However, we also see a number of malformed requests. We are planning to analyze malformed DNS requests and are curious about further possible attributes that can be of interest on the UDP or even IP level.

6 Conclusion

The results show important facts of blacklists such as the sizes and intersections among each other. The short activity periods of an IP address of less than or equal to one day prove that blacklists must react quickly and suggest that it might not be worth leaving addresses on the list forever. On the other hand the regional analysis reveals weaknesses such as low hit rates for users from certain countries. When applying blacklisting, these results can help to optimize blacklisting as a means to protect from spam.

Acknowledgements

We thank Bert Ungerer, Sebastian Ganschow, Prof. Dr. Norbert Pohlmann, Manuel Schmitt and Dr. Bülow-Masiak GmbH for their support.

References

[1]: C. Rossow (2007): Anti-spam measure of European ISPs/ESPs. http://www.internet-sicherheit.de/fileadmin/docs/publikationen/anti-spam-measures-of-european-isps-esps.pdf

[2]: Spamhaus (2008): Effective Spam Filtering. http://www.spamhaus.org/effective_filtering.html

[3]: P. Manzano, C. Rossow (2007): Provider Security Measures. Deliverable 2.1.6 of ENISA's Work Programme 2007. http://www.enisa.europa.eu/pages/spam/doc/enisa_spam_study_2007.pdf

[4]: A. Ramachandran, N. Feamster (2006): Understanding the network-level behavior of spammers. http://www.cc.gatech.edu/~feamster/papers/p396-ramachandran.pdf

[5]: A. Ramachandran, N. Feamster, S. Vempala (2007): Filtering spam with behavioral blacklisting. http://www.cc.gatech.edu/~feamster/papers/bb-ccs2007.pdf

[6]: A. Ramachandran, N. Feamster, D. Dagon (2006): Revealing botnet membership using DNSBL counter-intelligence. http://www.cc.gatech.edu/~feamster/publications/dnsbl.pdf

[7]: J. Jung, E. Sit (2004): An empirical study of spam traffic and the use of DNS black lists, http://www.imconf.net/imc-2004/papers/p370-jung.pdf