Improving our good old blacklisting

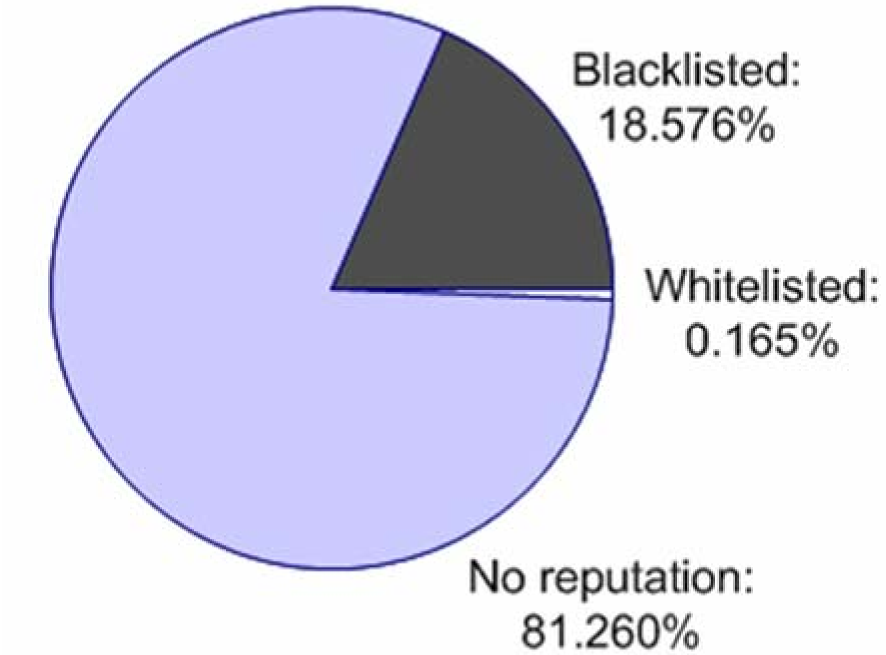
Christian Rossow
rossow[@]internet-sicherheit.de

Institute for Internet security
https://www.internet-sicherheit.de
University of Appl. Sciences Gelsenkirchen

Blacklists build up IP reputation

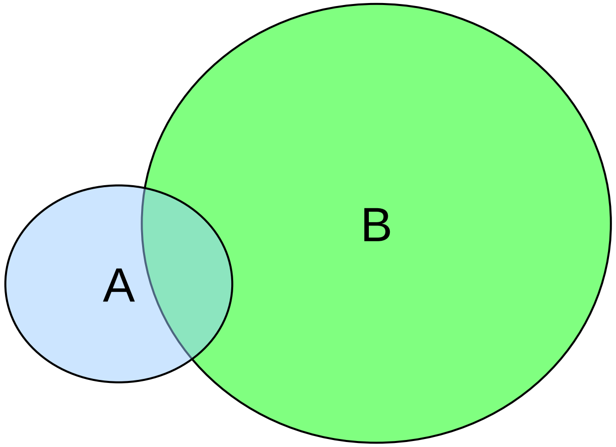
- Combining black- and whitelisting is probably the most effective antispam-mechanism
- However, a high dependency on black-/whitelist providers exists
- **Union of the most important blacklists => only 19% of advertised IPv4 addresses can be judged concerning email reputation!**

- What can be said about blacklists?
- Which blacklist(s) to choose?



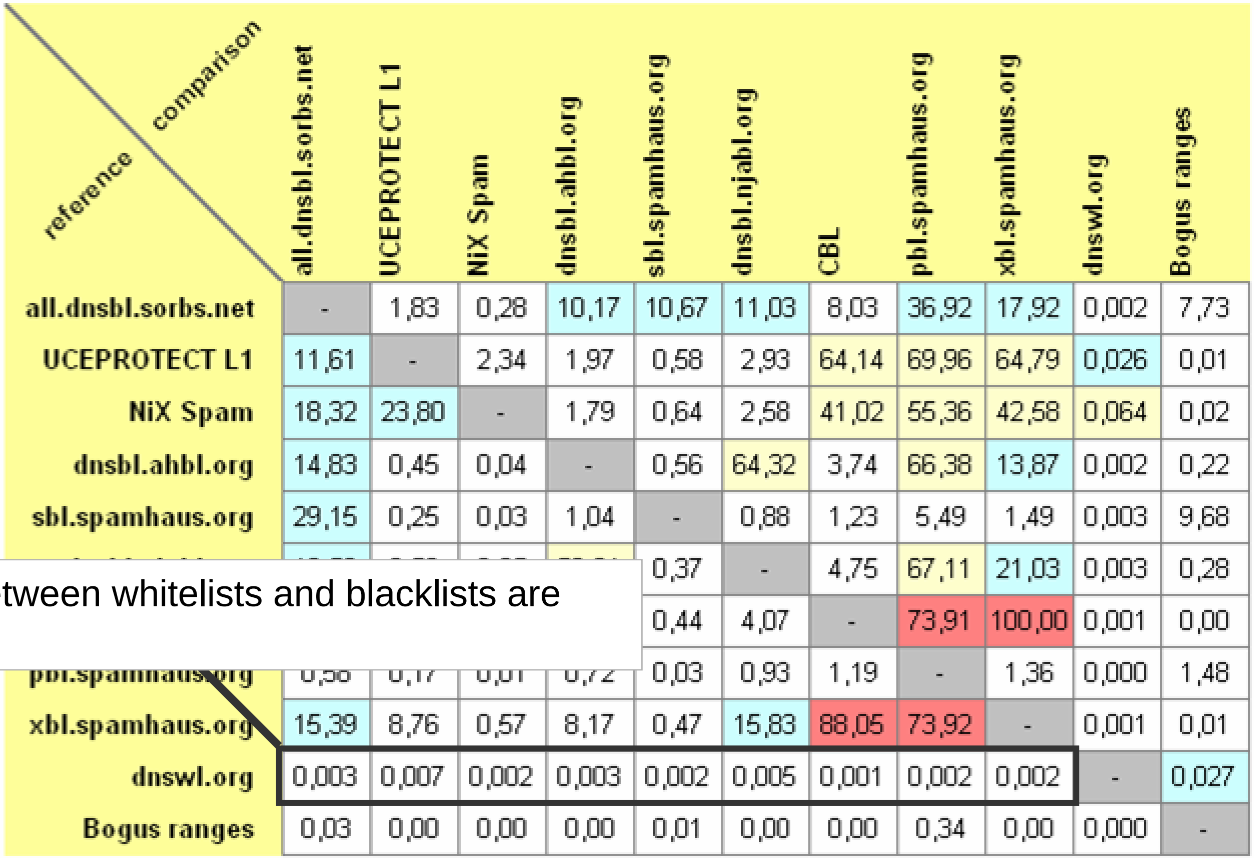
Intersections between blacklists (1/2)

- Blacklists are similar to each other
  - Same data sources
  - Data exchange between blacklists
  - Same spammers are detected by many blacklists
- Analysis of intersections
  - How much does blacklist A cover blacklist B?
  - What can be concluded?



Intersections between blacklists (2 / 2)

- Array with intersections:



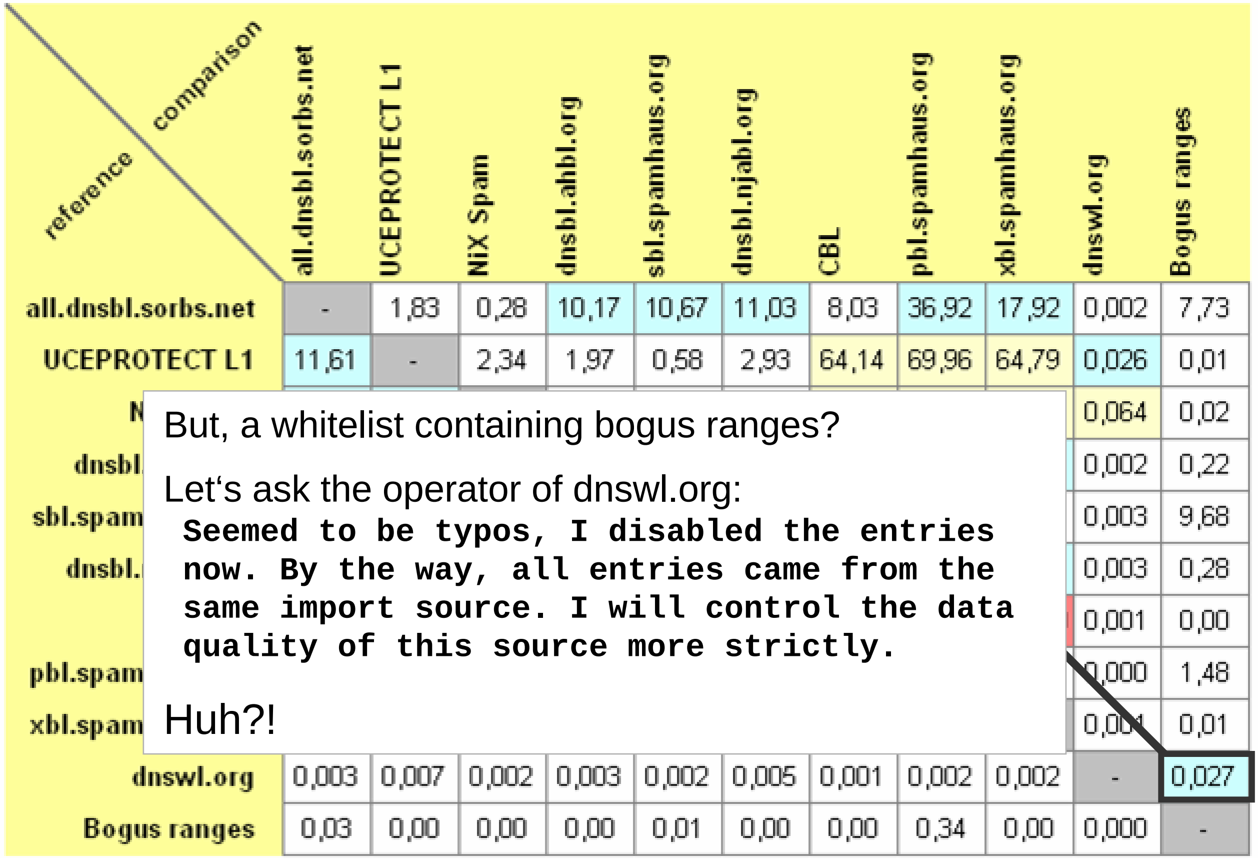
reference \ comparison	all.dnsbl.sorbs.net	UCEPROTECT L1	NiX Spam	dnsbl.ahbl.org	sbl.spamhaus.org	dnsbl.njabl.org	CBL	pbl.spamhaus.org	xbl.spamhaus.org	dnswl.org	Bogus ranges
all.dnsbl.sorbs.net	-	1,83	0,28	10,17	10,67	11,03	8,03	36,92	17,92	0,002	7,73
UCEPROTECT L1	11,61	-	2,34	1,97	0,58	2,93	64,14	69,96	64,79	0,026	0,01
NiX Spam	18,32	23,80	-	1,79	0,64	2,58	41,02	55,36	42,58	0,064	0,02
dnsbl.ahbl.org	14,83	0,45	0,04	-	0,56	64,32	3,74	66,38	13,87	0,002	0,22
sbl.spamhaus.org	29,15	0,25	0,03	1,04	-	0,88	1,23	5,49	1,49	0,003	9,68
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,37	-	4,75	67,11	21,03	0,003	0,28
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,44	4,07	-	73,91	100,00	0,001	0,00
pbl.spamhaus.org	0,58	0,17	0,01	0,72	0,03	0,93	1,19	-	1,36	0,000	1,48
xbl.spamhaus.org	15,39	8,76	0,57	8,17	0,47	15,83	88,05	73,92	-	0,001	0,01
dnswl.org	0,003	0,007	0,002	0,003	0,002	0,005	0,001	0,002	0,002	-	0,027
Bogus ranges	0,03	0,00	0,00	0,00	0,01	0,00	0,00	0,34	0,00	0,000	-

Intersections between whitelists and blacklists are quite common.

© Christian J Dietrich, Institute for internet security (ifis)

Intersections between blacklists (2 / 2)

- Array with intersections:



comparison reference	all.dnsbl.sorbs.net	UCEPROTECT L1	NiX Spam	dnsbl.ahbl.org	sbl.spamhaus.org	dnsbl.njabl.org	CBL	pbl.spamhaus.org	xbl.spamhaus.org	dnswl.org	Bogus ranges
all.dnsbl.sorbs.net	-	1,83	0,28	10,17	10,67	11,03	8,03	36,92	17,92	0,002	7,73
UCEPROTECT L1	11,61	-	2,34	1,97	0,58	2,93	64,14	69,96	64,79	0,026	0,01
N										0,064	0,02
dnsbl.										0,002	0,22
sbl.spam										0,003	9,68
dnsbl.										0,003	0,28
										0,001	0,00
pbl.spam										0,000	1,48
xbl.spam										0,001	0,01
dnswl.org	0,003	0,007	0,002	0,003	0,002	0,005	0,001	0,002	0,002	-	0,027
Bogus ranges	0,03	0,00	0,00	0,00	0,01	0,00	0,00	0,34	0,00	0,000	-

But, a whitelist containing bogus ranges?

Let‘s ask the operator of dnswl.org:

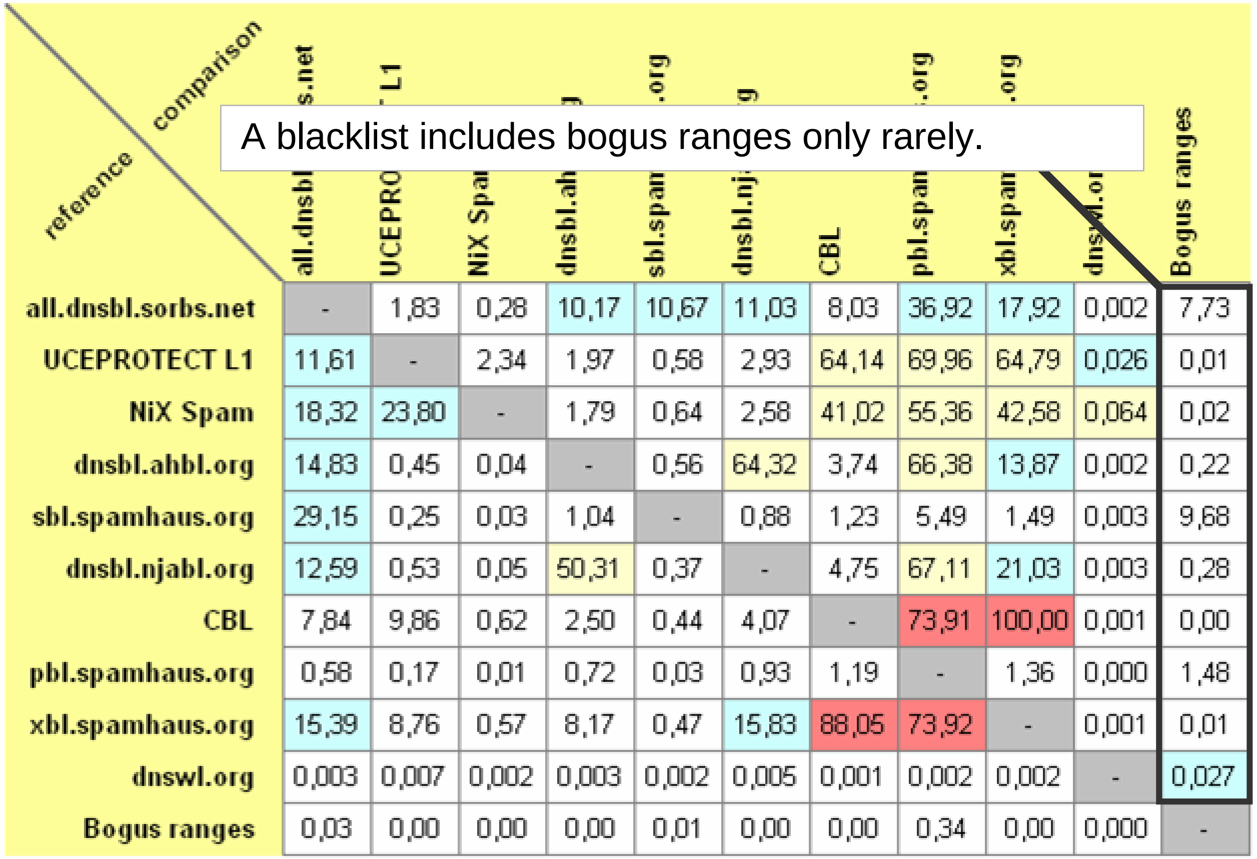
Seemed to be typos, I disabled the entries now. By the way, all entries came from the same import source. I will control the data quality of this source more strictly.

Huh?!

© Christian J Dietrich, Institute for internet security (ifis)

Intersections between blacklists (2 / 2)

- Array with intersections:

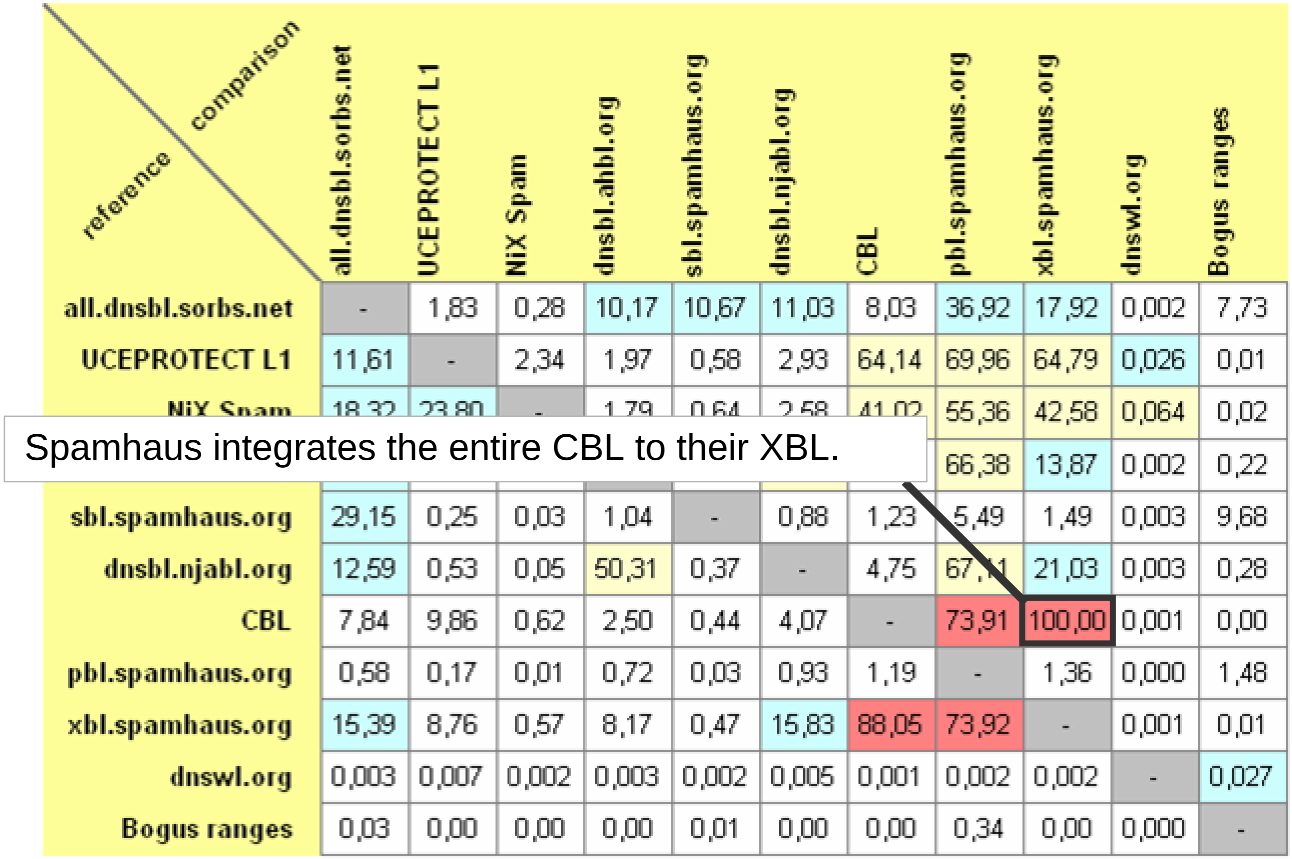


reference \ comparison	all.dnsbl s.net	UCEPRO L1	NiX Spa	dnsbl.ah	sbl.span .org	dnsbl.nj g	CBL	pbl.span .org	xbl.span .org	dnswl.or	Bogus ranges
all.dnsbl.sorbs.net	-	1,83	0,28	10,17	10,67	11,03	8,03	36,92	17,92	0,002	7,73
UCEPROTECT L1	11,61	-	2,34	1,97	0,58	2,93	64,14	69,96	64,79	0,026	0,01
NiX Spam	18,32	23,80	-	1,79	0,64	2,58	41,02	55,36	42,58	0,064	0,02
dnsbl.ahbl.org	14,83	0,45	0,04	-	0,56	64,32	3,74	66,38	13,87	0,002	0,22
sbl.spamhaus.org	29,15	0,25	0,03	1,04	-	0,88	1,23	5,49	1,49	0,003	9,68
dnsbl.njabl.org	12,59	0,53	0,05	50,31	0,37	-	4,75	67,11	21,03	0,003	0,28
CBL	7,84	9,86	0,62	2,50	0,44	4,07	-	73,91	100,00	0,001	0,00
pbl.spamhaus.org	0,58	0,17	0,01	0,72	0,03	0,93	1,19	-	1,36	0,000	1,48
xbl.spamhaus.org	15,39	8,76	0,57	8,17	0,47	15,83	88,05	73,92	-	0,001	0,01
dnswl.org	0,003	0,007	0,002	0,003	0,002	0,005	0,001	0,002	0,002	-	0,027
Bogus ranges	0,03	0,00	0,00	0,00	0,01	0,00	0,00	0,34	0,00	0,000	-

A blacklist includes bogus ranges only rarely.

Intersections between blacklists (2 / 2)

- Array with intersections:

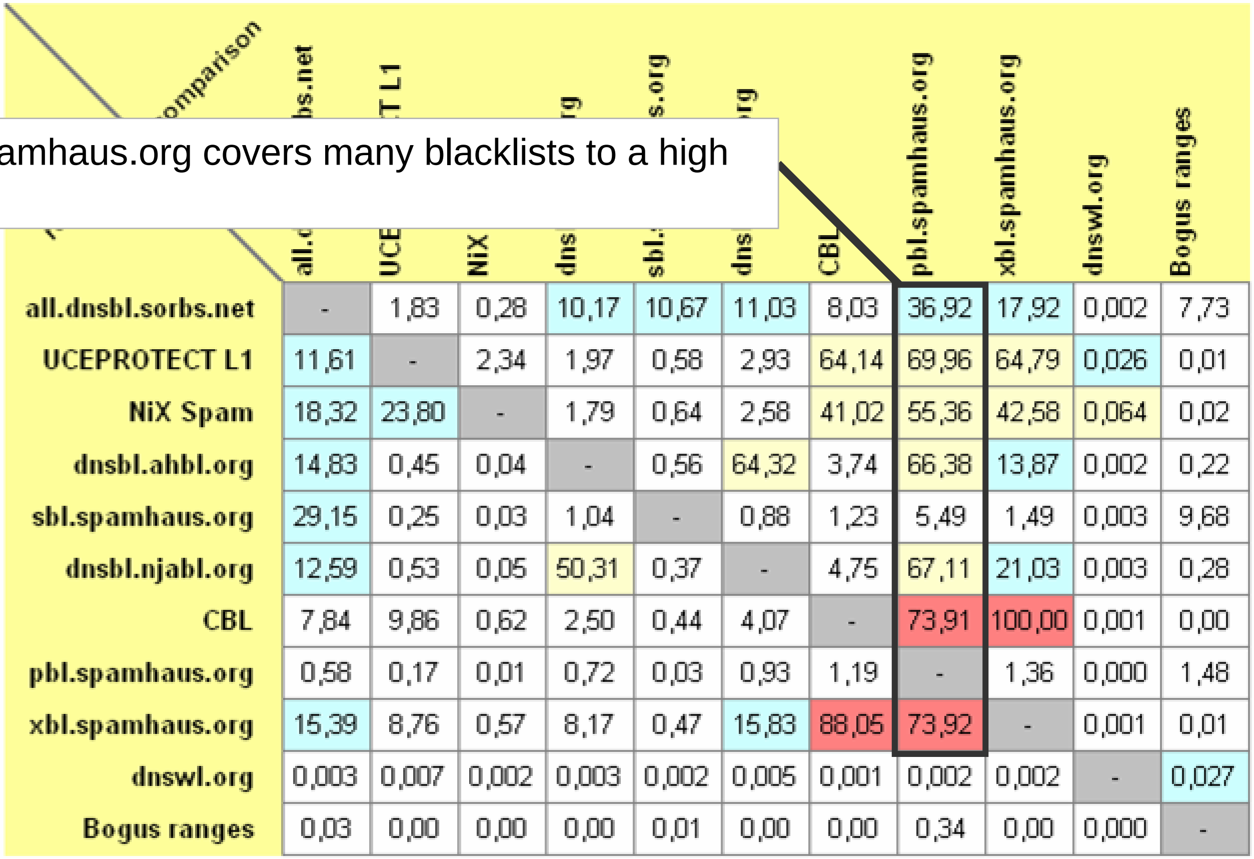


reference \ comparison	all.dnsbl.sorbs.net	UCEPROTECT L1	NiX Spam	dnsbl.ahbl.org	sbl.spamhaus.org	dnsbl.njabl.org	CBL	pbl.spamhaus.org	xbl.spamhaus.org	dnswl.org	Bogus ranges
all.dnsbl.sorbs.net	-	1,83	0,28	10,17	10,67	11,03	8,03	36,92	17,92	0,002	7,73
UCEPROTECT L1	11,61	-	2,34	1,97	0,58	2,93	64,14	69,96	64,79	0,026	0,01
NiX Spam	18,32	23,80	-	1,79	0,64	2,58	41,02	55,36	42,58	0,064	0,02
								66,38	13,87	0,002	0,22
sbl.spamhaus.org	29,15	0,25	0,03	1,04	-	0,88	1,23	5,49	1,49	0,003	9,68
dnsbl.njabl.org	12,59	0,53	0,05	50,31	0,37	-	4,75	67,41	21,03	0,003	0,28
CBL	7,84	9,86	0,62	2,50	0,44	4,07	-	73,91	100,00	0,001	0,00
pbl.spamhaus.org	0,58	0,17	0,01	0,72	0,03	0,93	1,19	-	1,36	0,000	1,48
xbl.spamhaus.org	15,39	8,76	0,57	8,17	0,47	15,83	88,05	73,92	-	0,001	0,01
dnswl.org	0,003	0,007	0,002	0,003	0,002	0,005	0,001	0,002	0,002	-	0,027
Bogus ranges	0,03	0,00	0,00	0,00	0,01	0,00	0,00	0,34	0,00	0,000	-

Spamhaus integrates the entire CBL to their XBL.

Intersections between blacklists (2 / 2)

- Array with intersections:

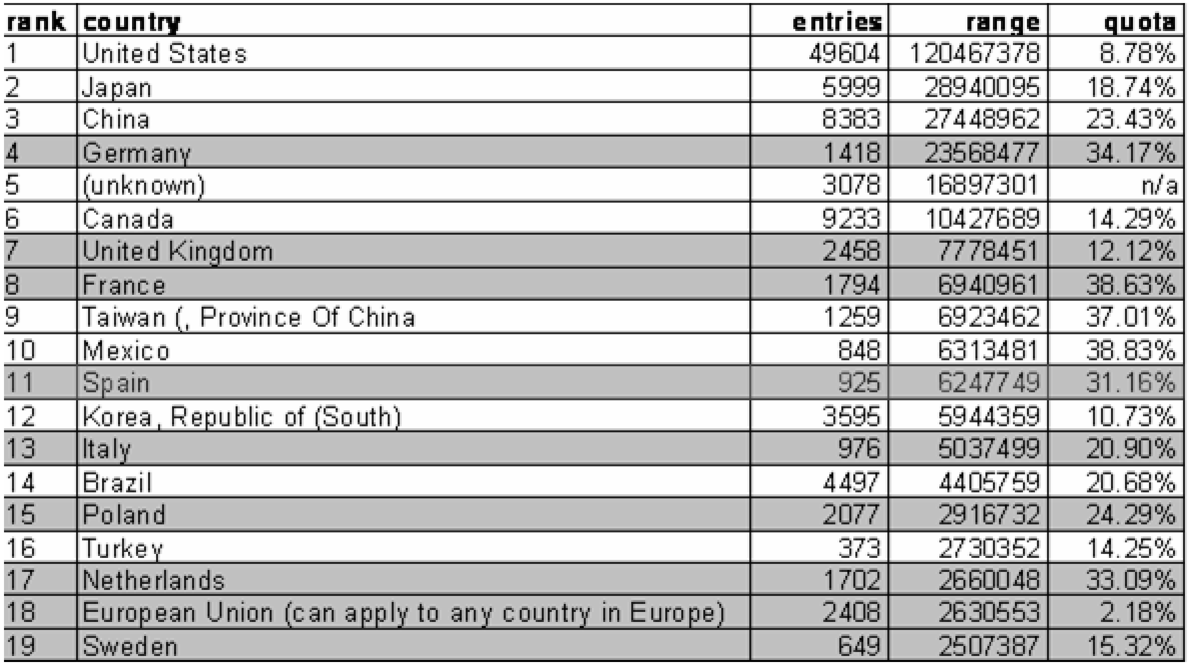


PBL.spamhaus.org covers many blacklists to a high degree.

omparison	all.c bs.net	UCE T L1	NiX	dns rg	sbl. s.org	dns rg	CBL	pbl.spamhaus.org	xbl.spamhaus.org	dnswl.org	Bogus ranges
all.dnsbl.sorbs.net	-	1,83	0,28	10,17	10,67	11,03	8,03	36,92	17,92	0,002	7,73
UCEPROTECT L1	11,61	-	2,34	1,97	0,58	2,93	64,14	69,96	64,79	0,026	0,01
NiX Spam	18,32	23,80	-	1,79	0,64	2,58	41,02	55,36	42,58	0,064	0,02
dnsbl.ahbl.org	14,83	0,45	0,04	-	0,56	64,32	3,74	66,38	13,87	0,002	0,22
sbl.spamhaus.org	29,15	0,25	0,03	1,04	-	0,88	1,23	5,49	1,49	0,003	9,68
dnsbl.njabl.org	12,59	0,53	0,05	50,31	0,37	-	4,75	67,11	21,03	0,003	0,28
CBL	7,84	9,86	0,62	2,50	0,44	4,07	-	73,91	100,00	0,001	0,00
pbl.spamhaus.org	0,58	0,17	0,01	0,72	0,03	0,93	1,19	-	1,36	0,000	1,48
xbl.spamhaus.org	15,39	8,76	0,57	8,17	0,47	15,83	88,05	73,92	-	0,001	0,01
dnswl.org	0,003	0,007	0,002	0,003	0,002	0,005	0,001	0,002	0,002	-	0,027
Bogus ranges	0,03	0,00	0,00	0,00	0,01	0,00	0,00	0,34	0,00	0,000	-

Regional views of blacklists

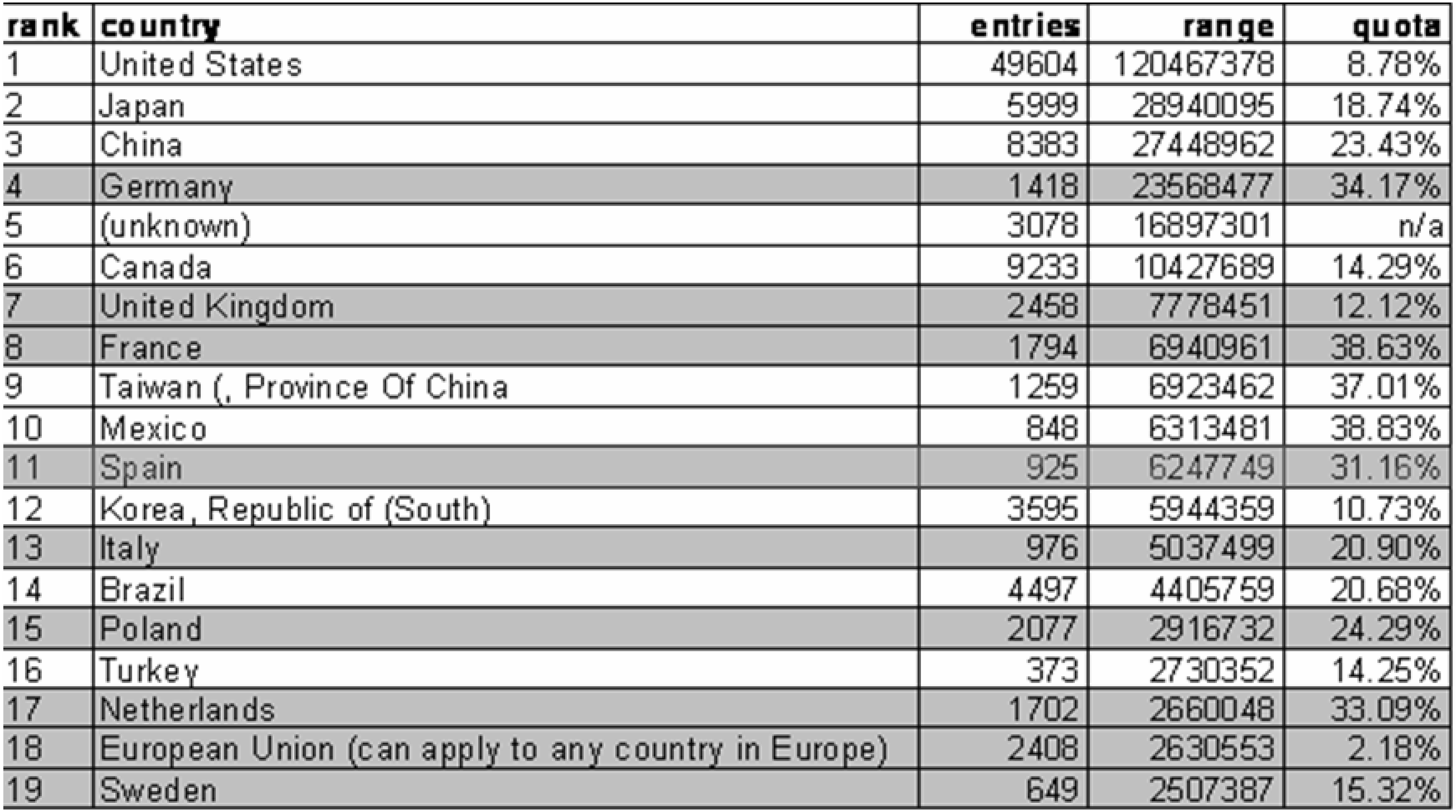
- Assign blacklist entries to regional attributes, e.g.
  - Country
  - RIR
  - Autonomous System
  - …



rank	country	entries	range	quota
1	United States	49604	120467378	8.78%
2	Japan	5999	28940095	18.74%
3	China	8383	27448962	23.43%
4	Germany	1418	23568477	34.17%
5	(unknown)	3078	16897301	n/a
6	Canada	9233	10427689	14.29%
7	United Kingdom	2458	7778451	12.12%
8	France	1794	6940961	38.63%
9	Taiwan (, Province Of China	1259	6923462	37.01%
10	Mexico	848	6313481	38.83%
11	Spain	925	6247749	31.16%
12	Korea, Republic of (South)	3595	5944359	10.73%
13	Italy	976	5037499	20.90%
14	Brazil	4497	4405759	20.68%
15	Poland	2077	2916732	24.29%
16	Turkey	373	2730352	14.25%
17	Netherlands	1702	2660048	33.09%
18	European Union (can apply to any country in Europe)	2408	2630553	2.18%
19	Sweden	649	2507387	15.32%

Figure: Spamhaus' PBL by country

Regional views of blacklists



rank	country	entries	range	quota
1	United States	49604	120467378	8.78%
2	Japan	5999	28940095	18.74%
3	China	8383	27448962	23.43%
4	Germany	1418	23568477	34.17%
5	(unknown)	3078	16897301	n/a
6	Canada	9233	10427689	14.29%
7	United Kingdom	2458	7778451	12.12%
8	France	1794	6940961	38.63%
9	Taiwan (, Province Of China	1259	6923462	37.01%
10	Mexico	848	6313481	38.83%
11	Spain	925	6247749	31.16%
12	Korea, Republic of (South)	3595	5944359	10.73%
13	Italy	976	5037499	20.90%
14	Brazil	4497	4405759	20.68%
15	Poland	2077	2916732	24.29%
16	Turkey	373	2730352	14.25%
17	Netherlands	1702	2660048	33.09%
18	European Union (can apply to any country in Europe)	2408	2630553	2.18%
19	Sweden	649	2507387	15.32%

Figure: Spamhaus' PBL by country

Quality of a blacklist (1/2)

- Two basic quality features can help to choose a good blacklist:
  - **True Positive Rate (TPR)**

    => How many emails were correctly tagged as spam?

    => TPR should be high, ideally 100%

    => Measure by the help of spamtraps (= dedicated spam)

  - **False Positive Rate (FPR)**

    => How many emails were falsely tagged as spam?

    => FPR should be quite low, ideally 0%

    => How can we get dedicated legitimate emails to measure FPRs?

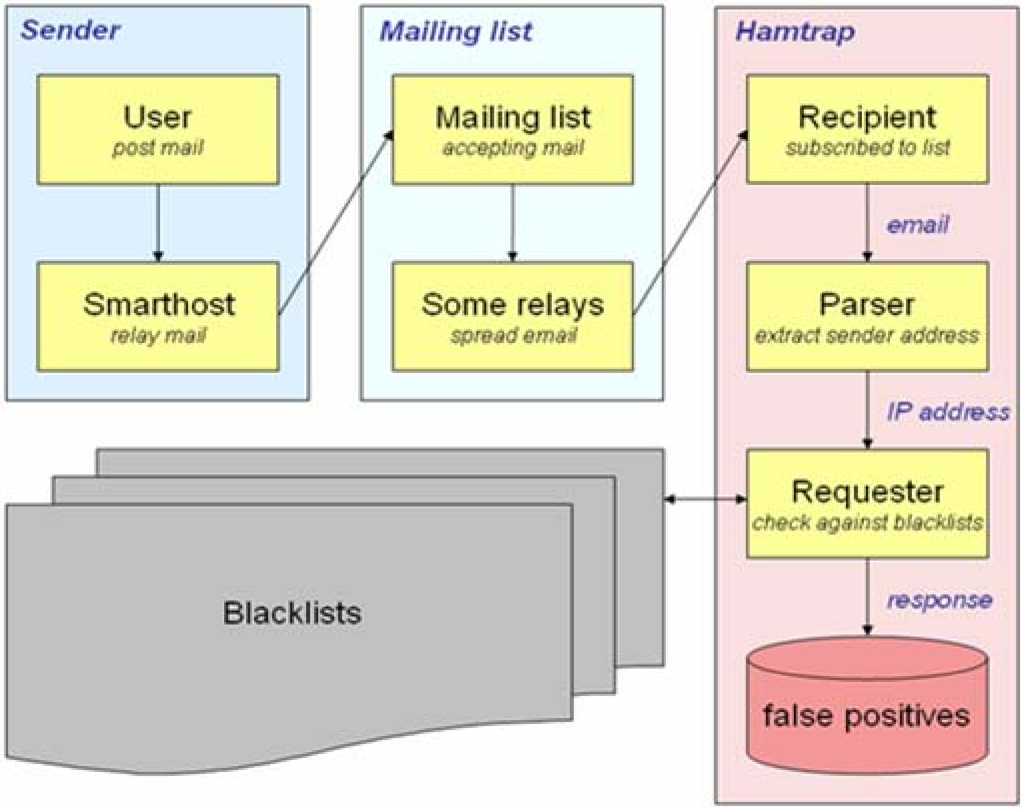
    => Development of hamtrap (= dedicated ham)

Quality of a blacklist (2 / 2)

- Draft of a hamtrap

- Moderated mailing lists serve as data sources

- In this way, blacklists can be checked against false positives, using the emails coming in from the lists.



Conclusion

- Which blacklist(s) should I use?
  - Do my clients accept some false positives?
  - Do my clients tolerate many false negatives?
  - Check our website for performance indicators of blacklists

- What can we conclude from regional views?
  - Europe does not have an entire clean slate
  - Providers should follow best practices to mitigate problems

For more information...

- Rely on our team of 8 members working in this area
- Planning to create a website dedicated to blacklist research
- We are looking for partners, so please contact us!
- Extensive information and statistics will be available at:

http://dnsbl.if-is.net

Improving our good old blacklisting

Thank you for your attention!

Any questions?

Christian Rossow
rossow[@]internet-sicherheit.de

Institute for Internet security
https://www.internet-sicherheit.de
University of Appl. Sciences Gelsenkirchen