

Shades of Grey: On the effectiveness of reputation based “blacklists”

- *Sushant Sinha, Michael Bailey and Farnam Jahanian*

University of Michigan

October 8, 2008.

Malware 2008



Today's security challenges

- Large number of hosts on the Internet are infected today with malware
 - 129.5 *million* malware detections during 2007/07~12, 66.7% more than 2007/01~06 (MSFT)
 - Infected through a variety of ways (email, web)
- Preventing malware infection is difficult
 - Malware is getting more sophisticated
 - 40% of malware exhibited fewer malicious behavior under debuggers (Xu Chen - DSN 08)
 - Anti virus software is lagging in detection
 - Average detection rate 52% (CloudAV - Usenix Sec 08)

One solution: Blacklists

- List of network identifiable entities known for bad behavior
 - Source IP known for spewing spam
 - TCP ports that are constantly probed (windows file sharing ports)
 - DNS domains known for hosting malware
- Getting popular and it is distributed by a number of organizations



Our paper

- Little is known about the effectiveness of current blacklists
- Investigation of spam tools and email characteristics on a large academic network
- Evaluation of the accuracy of four prominent spam blacklists on the academic network
 - SpamHaus, SpamCop, Sorbs and NJABL
- Preliminary study on the causes of inaccuracy
 - Why do blacklists exhibit false positives and false negatives?

Related work

- Ramachandran collected spam to an unused domain (Sigcomm 2006)
 - Showed that the spammers were clustered in few IP ranges
 - We perform a detailed investigation of all spam and legitimate mail (or ham) to an academic network
- A number of studies have suggested that blacklists may be incomplete
 - Ramachandran and Feamster (CEAS 2006, CCS 2007)
 - We quantify this inaccuracy

Related Work ... (contd.)

- Spam blacklist are generated by monitoring unused email addresses called *spamtraps*
 - Ramachandran suggests blacklisting based on sending patterns of the sources (CCS 2007)
 - Xie suggests including dynamic IP ranges in blacklists to catch more spam (Sigcomm 2007)
- Zhang suggests using your network similarity with other networks to reduce blacklist size (Usenix Security 2008)
- We evaluated only production spam blacklists

Approach

- Monitored email traffic through an academic network using span port to a gateway router
- SMTP sessions were reconstructed using libnids
- Added “Received” header to each mail
 - Mail server automatically adds this header
 - But we are not behind the mail server
- Each email was fed to SpamAssassin to check whether it is spam or not

Validating SpamAssassin

- SpamAssassin may itself be wrong and so we validated it
- Evaluated four email mailboxes that were hand classified as ham or spam
- For a threshold of 5.0, SpamAssassin
 - false positives from 0.02% to 0.56%
 - false negative from 4.02% to 5.6%
- It has very few false positives and reasonable false negatives

Evaluation

- Over a period of 10 days we observed 2 million SMTP connections
 - Half of these connections failed
- Average of 8,000 SMTP connections per hour
- 20% of successful SMTP messages were ham
 - 15% of the messages got a score of zero
- 53,579 unique email destinations
 - 64 within the university network
- 609,199 unique email sources
 - 111 within the university network

Characteristics of spam and ham

- Spam is distributed across a large number of sources but ham is concentrated in a few sources
 - Top 10 ham hosts covered 80% of ham
 - Top 10 spam hosts covered 10% of spam
- Spam is targeted to a few destinations but ham is distributed across destinations
 - Top 10 destinations covered 80% of spam
 - Top 10 destinations covered 50% of ham

Blacklists effectiveness

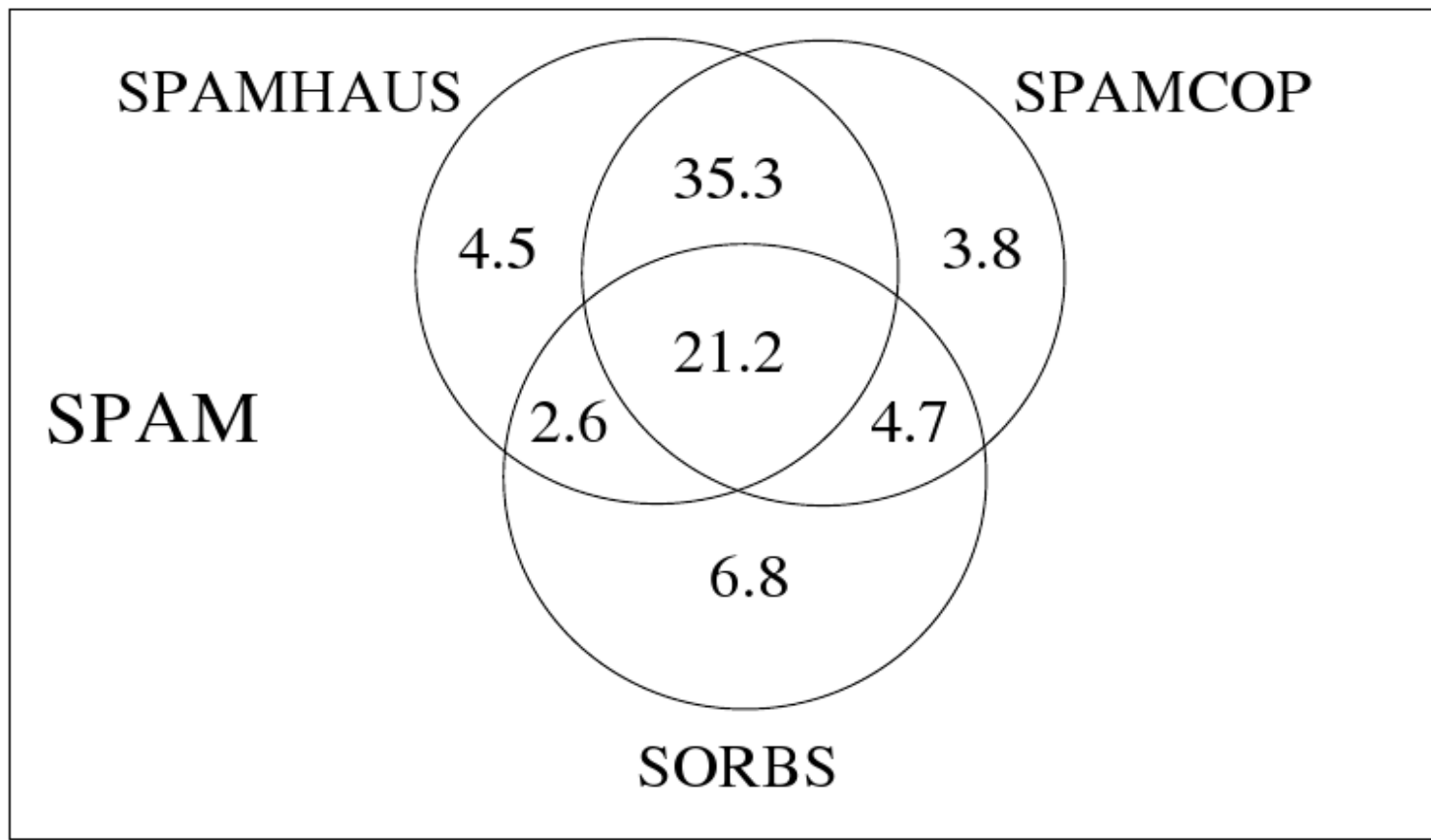
- For a SpamAssassin threshold of 5

	False positive rate		False negative rate	
	total	uniq src	total	uniq src
SpamHaus	0.6	5.2	36.3	41
SpamCop	2.3	13.6	34.9	40.2
SORBS	9.5	26.5	64.8	59.2
NJABL	0.2	0.5	98.4	98.1

- . Blacklists have few false positives but high false negatives
- . Sorbs has high false positives

Exploring false negatives - I

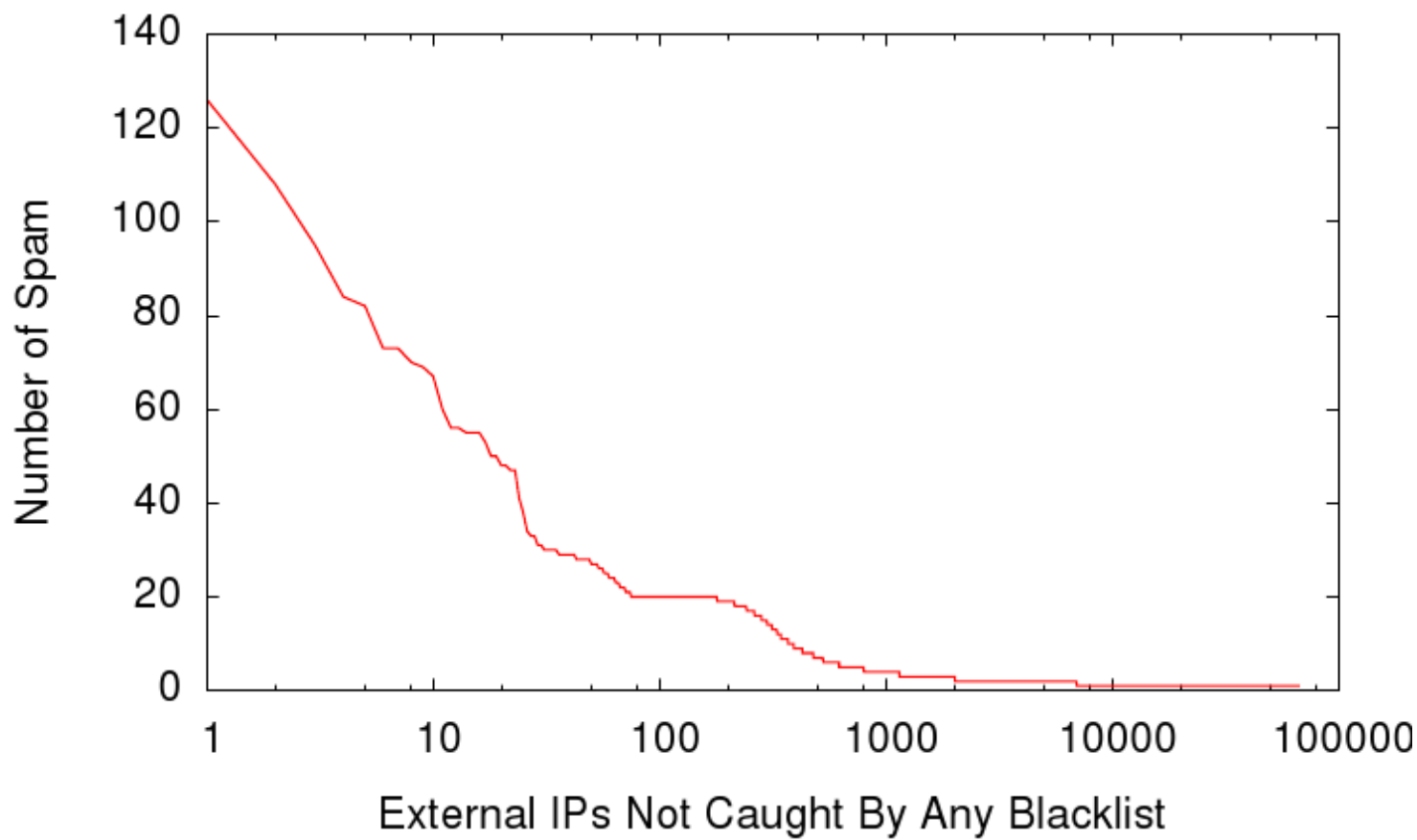
- Blacklists overlap in detection



- . Blacklists may not have wide visibility into spam ¹²

Exploring false negatives - II

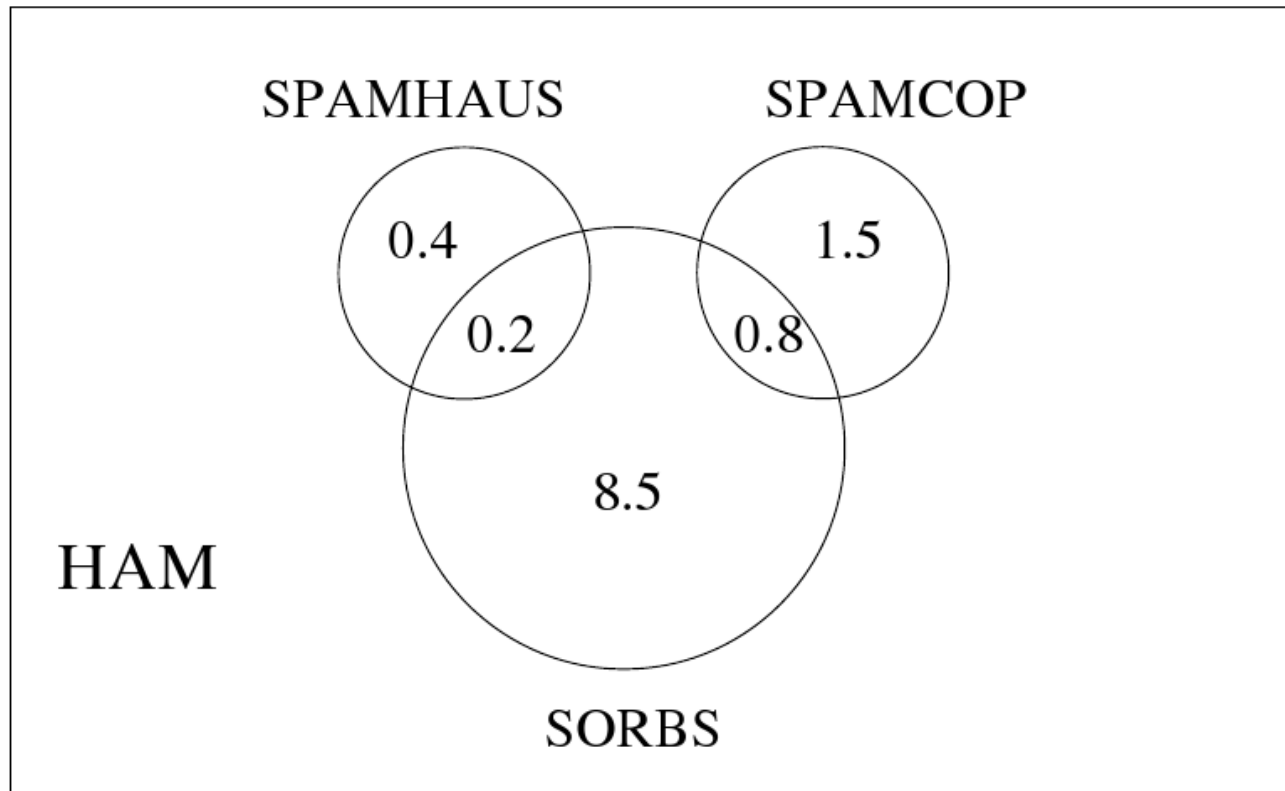
- Number of spam sent by sources that the blacklists missed



. Most of these sources seem to be low volume spammers

Exploring false positives - I

- Blacklists agreement on false positives



. They disagree on false positives indicating it may not be an incorrect classification of SpamAssassin¹⁴

Exploring false positives - II

- Sorbs blacklisted 6 Gmail servers that sent significant ham to our network
- Some blacklists are aggressive and may blacklist prominent mail servers
- Gmail does not add “Received” header when a mail is sent through the web interface
 - AOL and Yahoo add the received header
 - Gmail adds the header if sent through the IMAP interface
 - Gmail server may get blacklisted when used by spammers

Conclusions

- Blacklists have become popular among operational community to detect threats
- Presented an evaluation of current blacklists by deploying in a large academic network
- Blacklists have non-trivial false positives and false negatives
- Explored possible causes behind their inaccuracy

Future ideas on improving accuracy

- Mail servers need to be in whitelist
 - Currently who is in whitelist is political
 - Can be resolved if blacklist generation is moved within a network
 - Mail servers may have significant good activity on a particular network
- Observing network traffic to broaden blacklists
- Targeted attacks vs global attacks
 - Spam that targets universities may not be globally visible
 - Deploying local spam traps may be a solution (eg, fake identities within umich.edu)



Thank you!

What SA rules were fired for messages that missed four bls?

Rule name	Percentage
URIBL_BLACK	80
URIBL_JP_SURBL	77
RAZOR2_CHECK	60
RAZOR2_CF_RANGE_01_100	60
RAZOR2_CF_RANGE_E8_51_100	58
HTML_MESSAGE	58
URIBL_OB_SURBL	56
PYZOR_CHECK	53
URIBL_SC_SURBL	52
URIBL_AB_SURBL	49
MISSING_MID	38