Université du Québec
École de technologie supérieure

Analysis of massive backscatter of email spam

Christopher P. Fuhrman, PhD
Department of Software and IT Engineering
ETS - École de technologie supérieure
Montreal, Quebec, Canada
Christopher.Fuhrman@etsmtl.ca

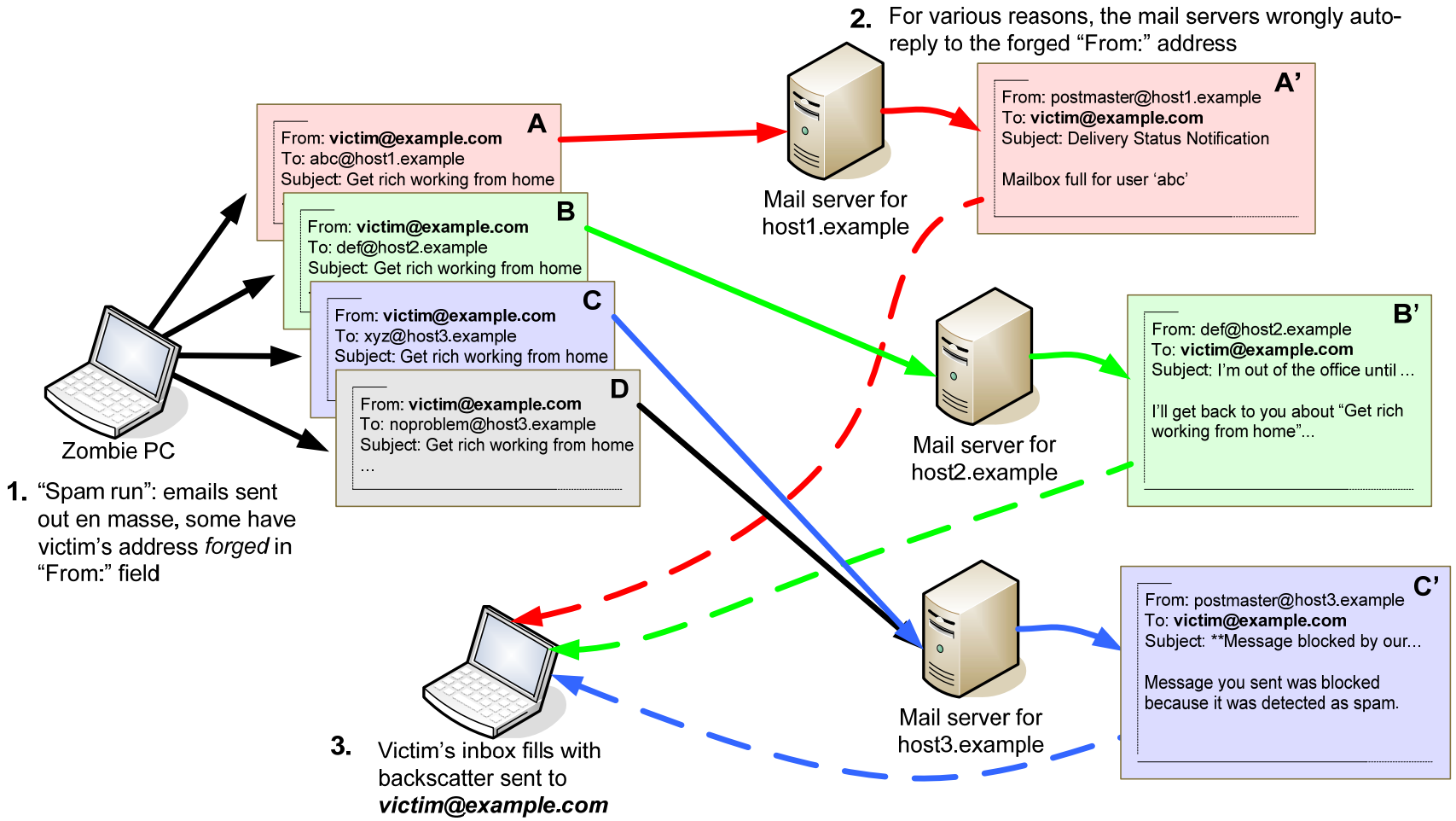
Overview

- What is backscatter of email spam?
- Potential value of analyzing backscatter
- Analysis tool
- Analysis results
- Conclusion and future work

Backscatter of email spam

- June 2007
  - Began receiving 300+ bounce messages/day on a private email address
  - Bounces of spam messages with my private address forged as the "From:" (sender)
  - Contained RFC822 headers (Received:) indicating spams originated from Zombie PCs
- January 2008
  - Receiving 900+ bounces/day

Possible anatomy of email backscatter from spam



History of email backscatter

- 1997: infamous joes.com spam attack
  - Angry users send out spams with forged "From:" headers to try to harm reputation of joes.com [8], resulting in a DDoS
- Spam nearly always has forged sender, choice was likely arbitrary
- 2003: postfix mail software implements Sender Address Verification (SAV) – detects bogus forged sender address by attempting to send it a probe via SMTP
  - *Some spammers possibly started using "real" addresses as the forged sender to fool SAV.*
  - Note: SAV is considered harmful by some for this reason as the probes can be a form of DDoS.
- 2007: Backscatter messages are classified as spam by Gordon & Thomas, who quantified it as 1% of total spam [7]

Types of email backscatter

1. Delivery Status Notification (DSN)
2. Out-of-office (OOO) auto-reply*
3. Verification challenge message*
4. Anti-virus notification
5. Anti-spam notification

*don't typically contain info about the original spam

All are automatic replies sent to a forged address.

Potential value in analyzing backscatter

- Identification of IP addresses of Zombie PCs
- Identification of spam subjects used in "fresh" spam runs
- Timestamps on auto-replies could indicate start/end times of spam runs

Problem: spammers (and not the researchers) choose the forged address. Repeating studies on email backscatter is not easy.

The stench of a Zombie PC used in a spam run…

- IP address doesn't have reverse DNS
- IP address is part of a dynamic IP block and therefore should not be the origin of SMTP (email)
- IP address is on a DNS block list
  - bl.spamcop.net
  - cbl.abuseat.org
  - etc.

http://www.pressebox.de/attachment/65739/Zombie-PC.jpg

A few words about forged "Received:" headers

- Common technique used in the past to fool spam analyzers.
- Most (probably all) of the spams seen in our backscatter emails are direct Zombie→MX spams
  - Majority had only one "Received:" line (no forgery)
  - A few had a signature forged Received line, but the IP of the zombie was not hidden. The forgery appeared to make the spam look like it was truly sent from the forged Sender's domain.

Relationship between forged sender and "Received:" line.

RFC822 headers

```
From sa...@telesensventures.com Mon Nov 26 00:08:28 2007
Received: from mx0.public.com (mx0.public.com [66.112.160.20])
        by public.com (8.12.10/8.12.10) with ESMTP id
        lAQ58SST093564 for <x...@public.com>; Mon, 26 Nov 2007 00:08:28
        -0500 (EST)
Received: from 121.88.184.97 ([121.88.184.97]) by mx0.public.com
        (8.11.6/8.11.6) with ESMTP id lAQ58Rs29724 for <m...@fw.merk.com>;
        Mon, 26 Nov 2007 00:08:28 -0500
Received: from [121.88.184.97] by a.ns.joker.com; Mon, 26 Nov 2007
        05:08:11 +0000
Message-ID: <000701c82fea$052ea66a$5e6137b7@dinoh>
From: "Replica Watches" <sa...@telesensventures.com>
To: "Exquisite Replica" <m...@fw.merk.com>
Subject: Exquisite Replica
Date: Mon, 26 Nov 2007 03:20:49 +0000
```

Forged information

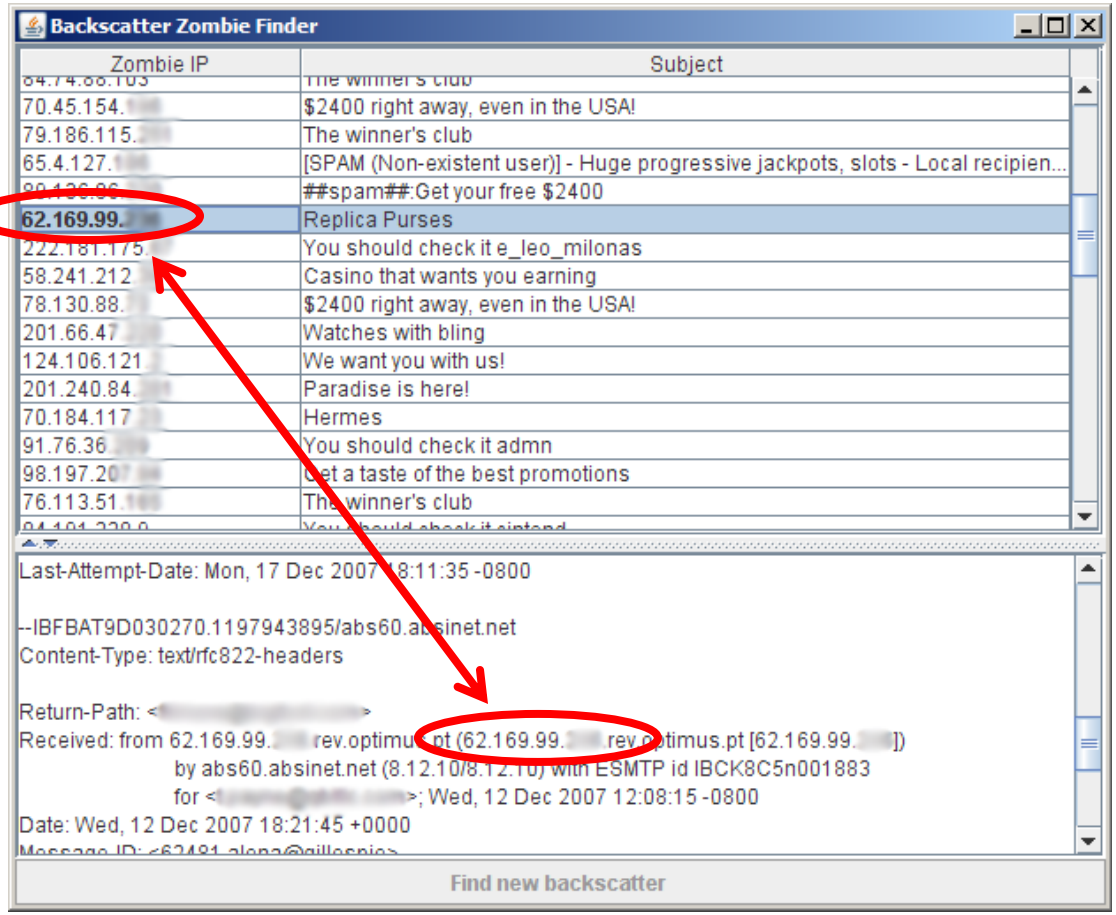
DNS Lookup

Domain	Type	Class	Result
telesensventures.com.	MX	IN	10 mx0.telesensventures.net.
telesensventures.com.	MX	IN	10 mx10.telesensventures.net.
telesensventures.com.	NS	IN	b.ns.joker.com.
telesensventures.com.	NS	IN	c.ns.joker.com.
telesensventures.com.	NS	IN	a.ns.joker.com.

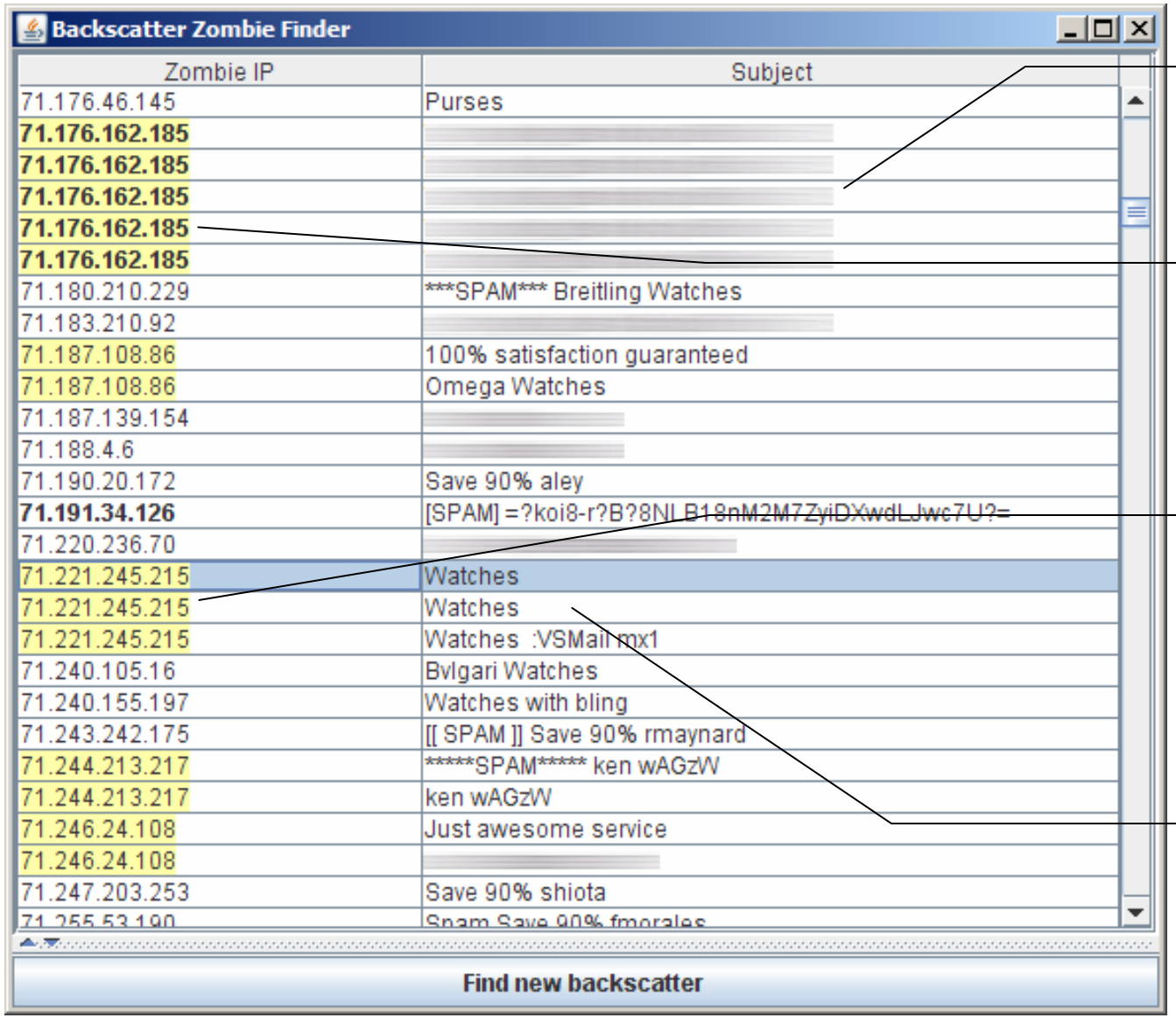
Automated analysis of email backscatter

- Tool written using JavaMail API
- Receives new backscatter in real-time with Gmail over IMAP (Gmail accurately classifies backscatter as spam)
- Isolates and extracts IP address of the Zombie PC for backscatter email that contains RFC822 headers
- Looks up the IP on four popular block lists using Domain Name Service (DNS)
- Facilitates observation of trends in backscatter (e.g., repeated IPs, subjects)
- Saves data in CSV format for later analysis with Excel

Analysis tool GUI (find the Zombie IP…)



Analysis tool GUI



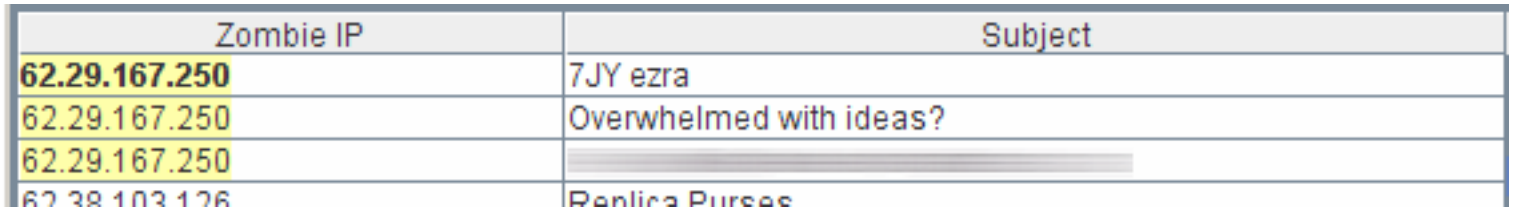
GENERAL AUDIENCES
All Ages Admitted

IP absent on the DNSBLs (in bold)

Repeated IP addresses (yellow highlight)

Repeated subjects for same IP

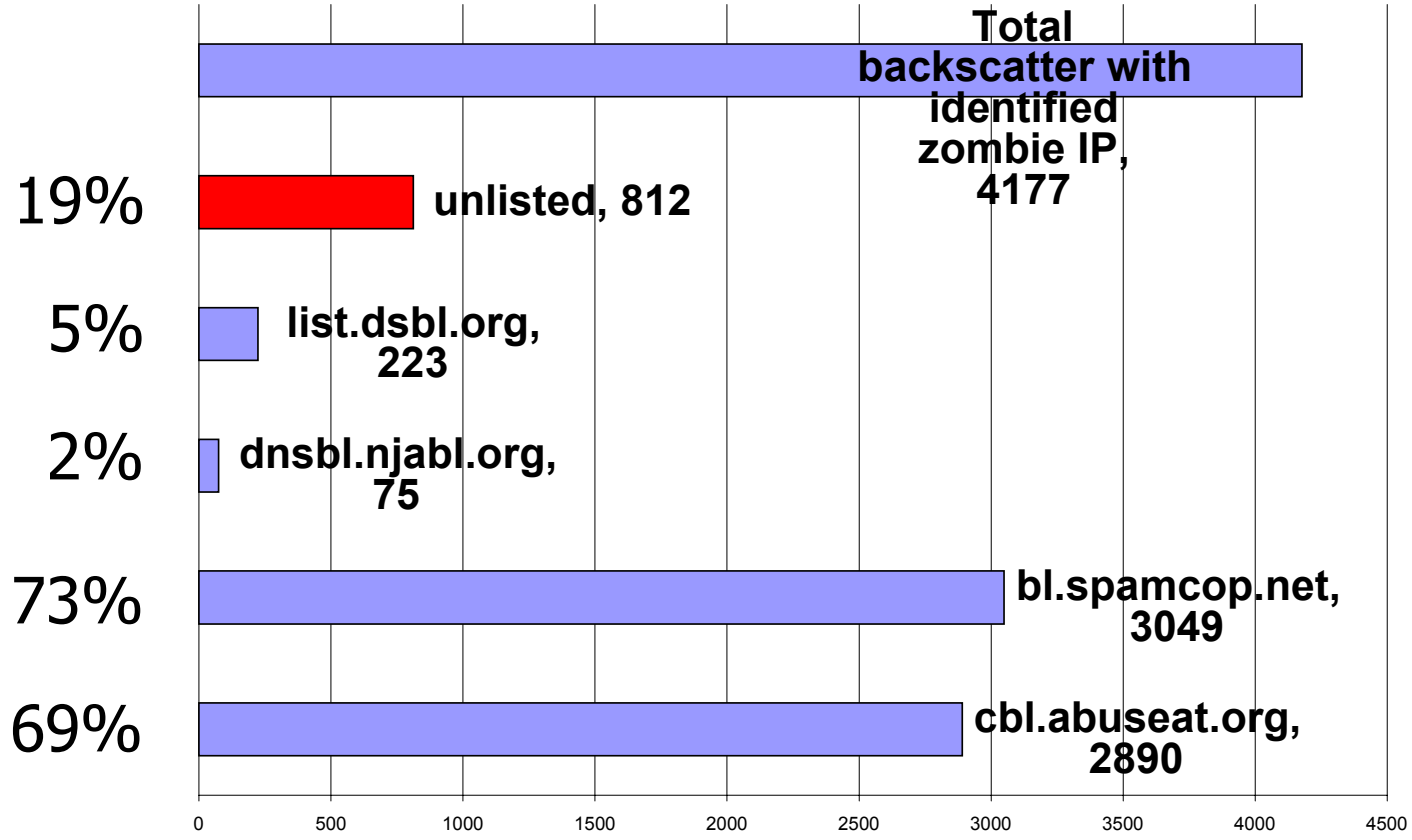
Analysis tool GUI



Zombie IP	Subject
62.29.167.250	7JY ezra
62.29.167.250	Overwhelmed with ideas?
62.29.167.250	
62.38.103.126	Replica Purses

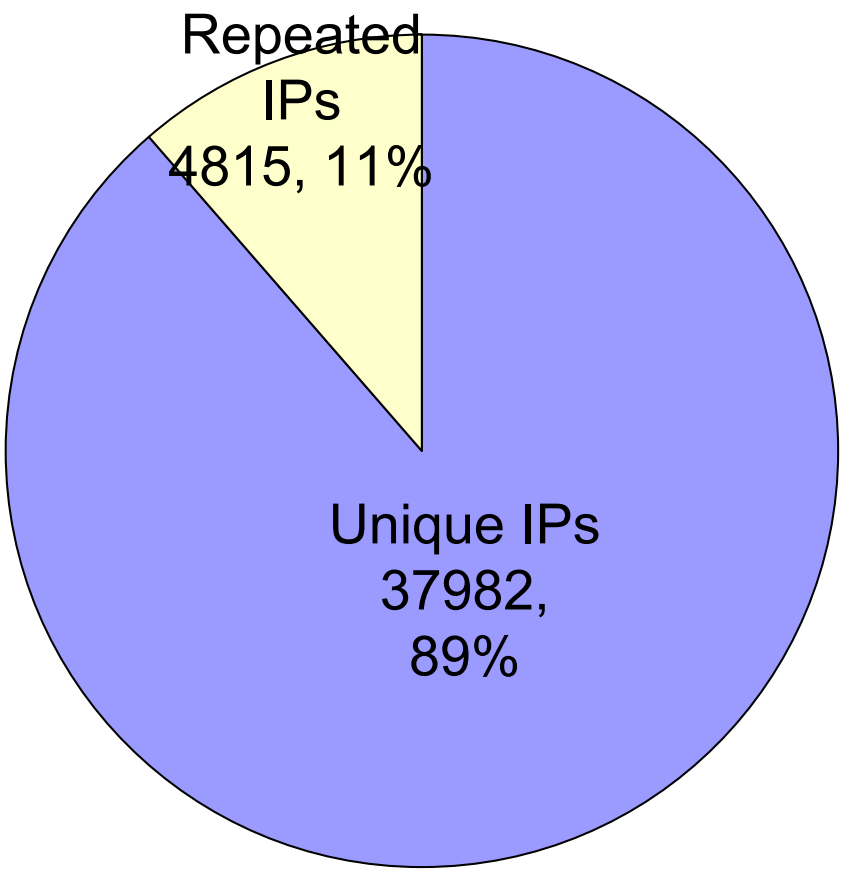
- Changes in a Zombie IP's listing
  - IP not listed at first (in bold)
  - Then appears on block lists afterwards

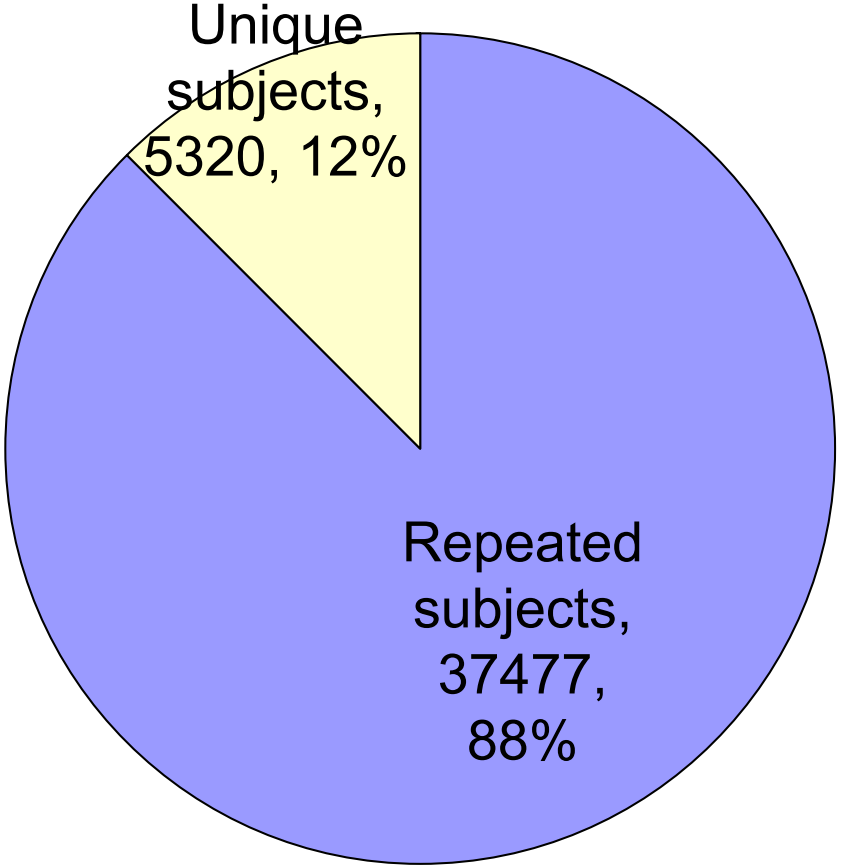
Results of a 5-day, real-time analysis using DNS Block Lists



DNSBL success for zombie IPs found via backscatter analysis

Results of analysis of 49,000 backscatter emails over 4-month period





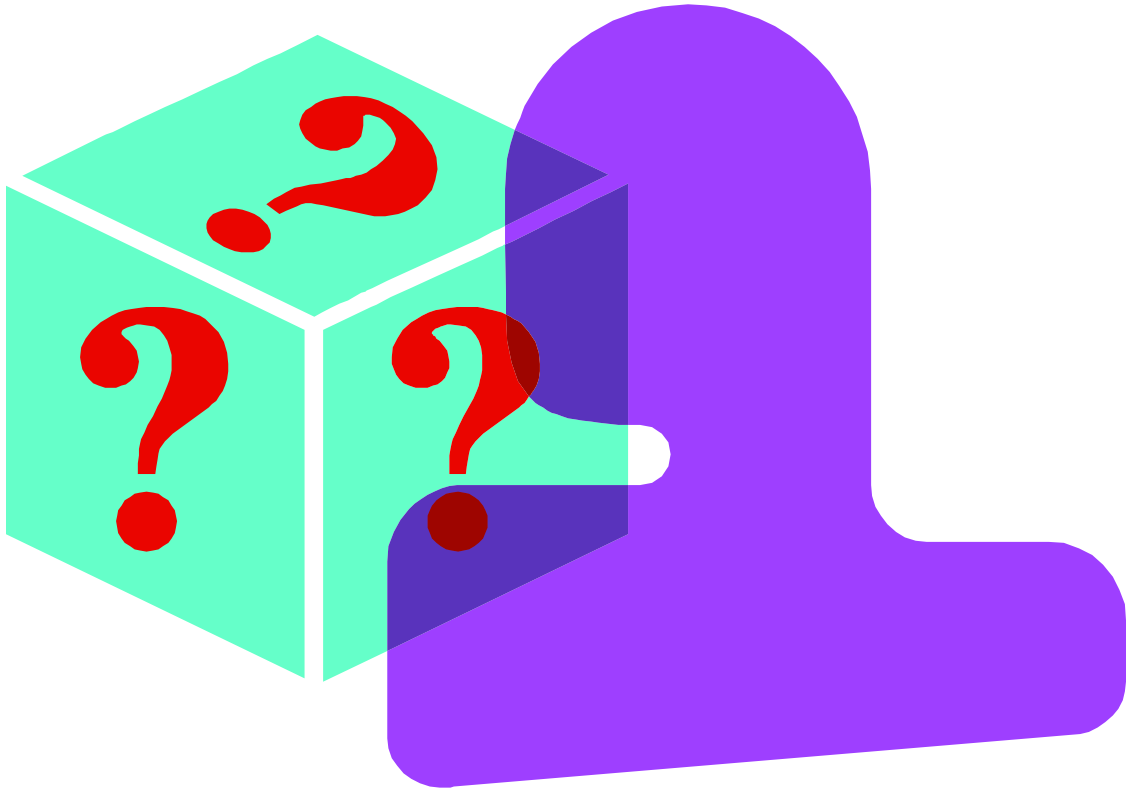
Conclusion

- Our analysis approach finds a significant number (19%) of IPs that are not (yet) block-listed
- The increasing numbers of backscatter emails received support the argument that spammers are defeating SAV by forging real email addresses
- It is difficult to repeat our analyses, since spammers decide which email addresses to forge.

Future directions

- Contribute the Zombie IPs identified in our real-time method to some DNSBLs
- Contribute the spam subjects/bodies to community Bayesian filter systems.
- Consider similar real-time analyses of email backscatter, but at the Mail Transfer Agent level.

Questions?



Background

- June 2007
  - Began receiving 300+ bounce messages/day on a private email address
  - Bounces of spam messages with my private address forged as the "From:" (sender)
  - Contained RFC822 headers (Received:) indicating spams originated from Zombie PCs
- July 2007
  - Discovered a "signature" forged Received: header in the spams (inside bounces)
- August 2007
  - Considered potential forensic value in analyzing the bounces (but no funding)
- September 2007
  - Discovered other "victims" of related backscatter in news.admin.net-abuse.sightings
- November 2007
  - Gmail provided IMAP access
  - Began using the javamail API to develop a tool to analyze backscatter data
- January 2008
  - Receiving 900+ bounce messages/day