

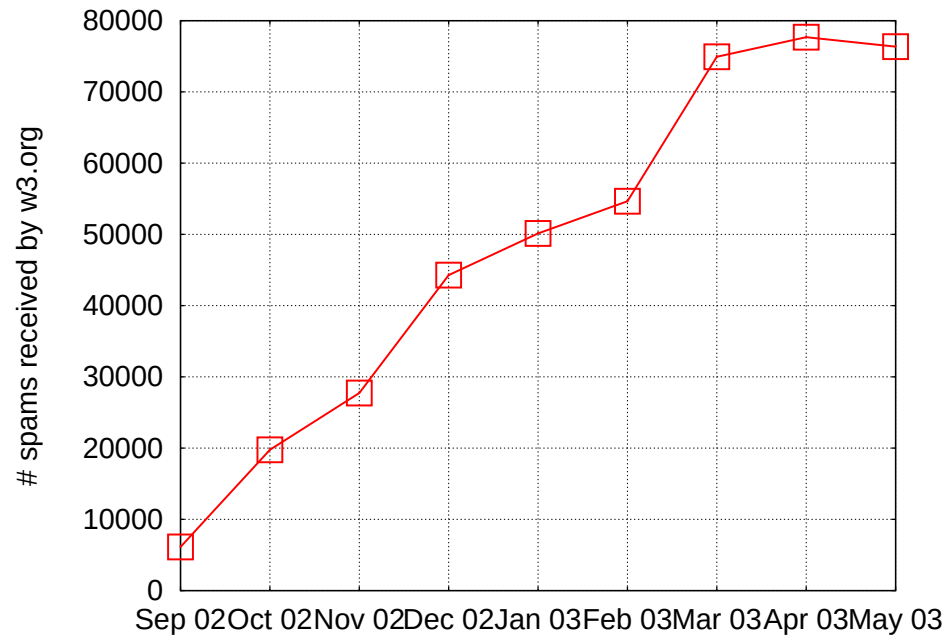
Spam Analysis - Re-examining SMTP

Jaeyeon Jung (jyjung@lcs.mit.edu)

MIT CSAIL

September 25, 2003

Spam keeps coming



	Oct 02	Feb 03	May 03
Total spams	19752	54642	76344
MTAs involved	28346	93243	105110
MTAs / spam	1.44	1.71	1.38

Spam detection relying on DNS broke

	Oct 02	Feb 03	May 03
MTAs involved	28346	93243	105110
NXDOMAIN	44.8%	43.9%	47.4%
Forged	33.2%	25.1%	33.6%
Broken	13.4%	20.3%	9.9%
Legit	8.5%	10.7%	9.2%

- ✓ Recent Verisign problem made it worse.

Problems with SMTP

- ✓ Hard to trace back to the original sender
 - IP address of the sending MTA recorded by receiving MTA
 - Compromised middleman makes it harder

	Oct 02	Feb 03	May 03
MTAs involved	28346	93243	105110
Broken MTAs	13.4%	20.3%	9.9%

Problems with SMTP (contd.)

- ✓ Lack of defense against spoofing
 - Sender name
 - Sender domain name
 - Header

	Oct 02	Feb 03	May 03
Total spams	19752	54642	76344
From Error	1.9%	1.5%	2.8%
From w3.org	3.8%	3.6%	4.3%
MTAs involved	28346	93243	105110
Forged	33.2%	25.1%	33.6%

Our approach

Push more work over a sender side by delaying a full email reception from a *suspicious* MTA

Our approach

Push more work over a sender side by delaying a full email reception from a *suspicious* MTA

- ✓ How to identify *suspicious* MTAs?
 - On-line MTA fingerprinting
 - MTA trace-back

Our approach

Push more work over a sender side by delaying a full email reception from a *suspicious* MTA

- ✓ How to identify *suspicious* MTAs?
 - On-line MTA fingerprinting
 - MTA tracing-back
- ✓ How much delay?
 - Long enough for spammers to use up their storage/network bandwidth
 - Short enough for the false positives not to complain

Implementation details

- ✓ Enhance SMTP to accept only header in the first try and ask for a full body after N second
- ✓ Characterize MTA behavior by OSes and programs
- ✓ Develop on-line algorithm to detect a suspicious MTA

Plans

- ✓ More data analysis and implementation