# Router-Level Spam Filtering Using TCP Fingerprints: Architecture and Measurement-Based Evaluation

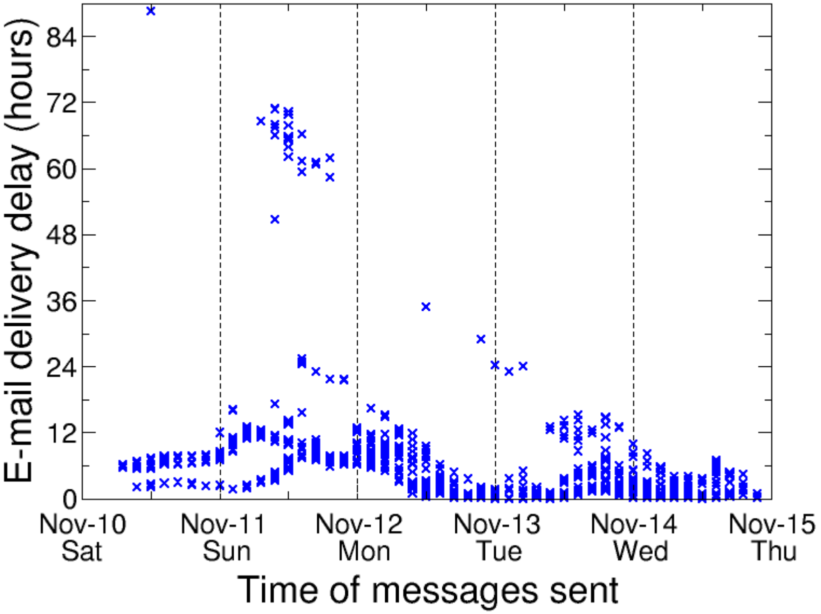
Holly Esquivel
Aditya Akella
Univ. Of Wisconsin-Madison

Tatsuya Mori
NTT

CEAS 2009
Mountain View, CA
July 15-16, 2009

# Why Router-Level Filtering?

- Scalability is a problem when it comes to spam filtering
- Large email services have 100s of millions of email accounts
- Email delivery delays can cause significant problems for businesses
- A  light-weight technique is needed to help ease these problems – we use TCP fingerprints for this filtering mechanism



Example of email delivery delay seen by a Corporation in Japan in early Nov. 2007 because of an increase in spam messages.

# Current Spam Filtering Techniques

- SMTP/End-Host
  - Blacklisting/Whitelisting
  - Greylisting
  - Authentication based
  - Content-based Filtering
- Router-level Mechanisms
  - Behavior based filters
    - Bayesian Classifiers applied to Bulk Email streams
    - Progressive Email Classifier
  - Commercial products
    - DPI-Based filters
    - Barracuda Spam Firewall

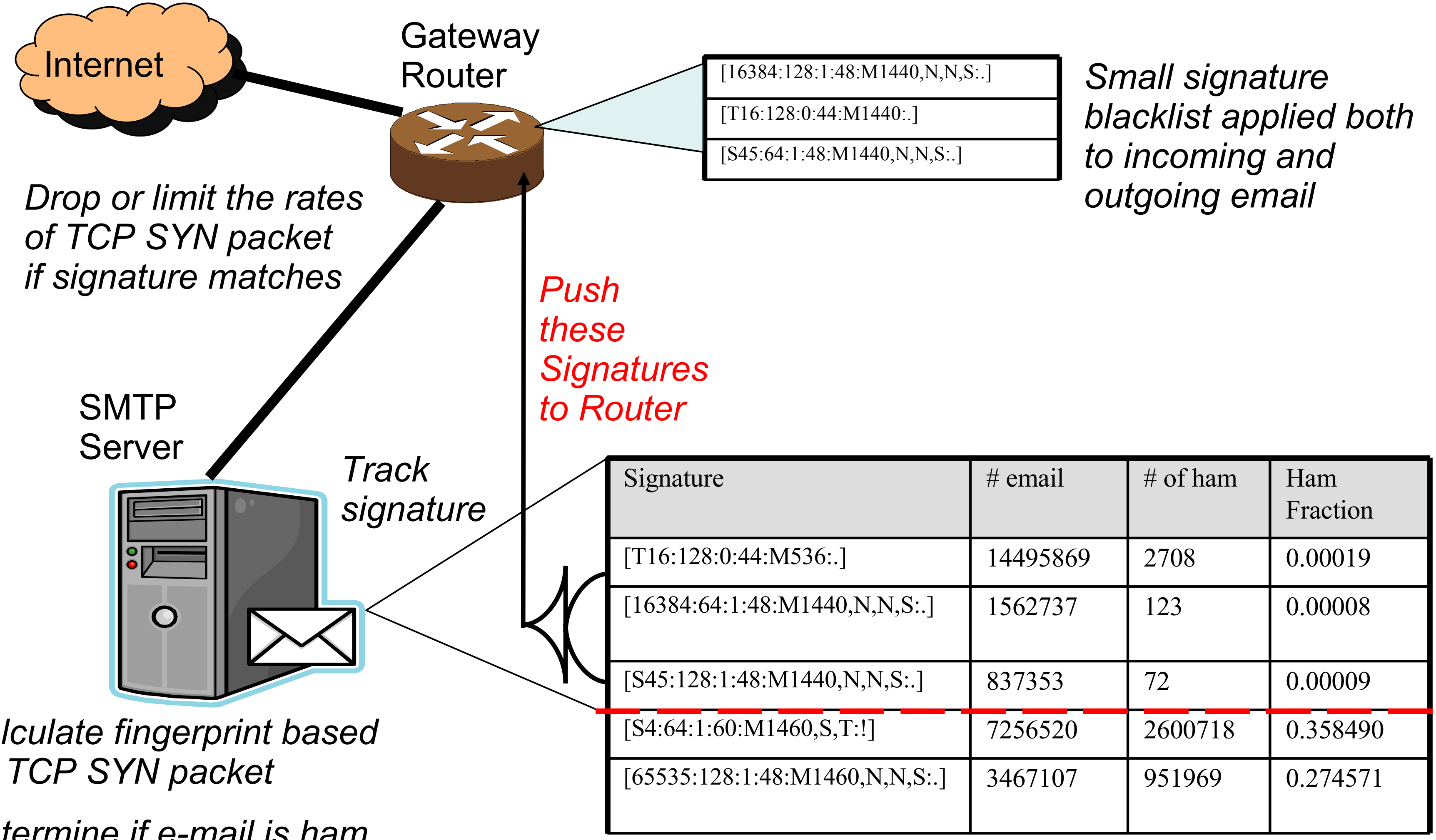
# TCP Fingerprinting

- Also known as Passive OS fingerprinting
- A single TCP SYN packet is all that is needed
- Done without a suspect’s knowledge
- Fingerprints can identify the OS genre and version
- Small number of legitimate fingerprints
- Tools: p0f, Ettercap, Siphon
  - Signature format: [W:T:D:S:O…:Q]
  - Example: [S4:64:1:60:M*,S,T,N,W5:.:] - Linux 2.6

# Our Approach

- Build a router-level *architecture* for spam filtering using TCP fingerprints
- Look beyond operating system genres
  - Use fine grain fingerprints
- Goals:
  - Light-weight and stateless in nature
  - Feedback based approach
  - Small amount of required memory
  - Supplement existing filters

# Architecture



| |
|---|
| [16384:128:1:48:M1440,N,N,S:.] |
| [T16:128:0:44:M1440:.] |
| [S45:64:1:48:M1440,N,N,S:.] |

*Small signature blacklist applied both to incoming and outgoing email*

*Drop or limit the rates of TCP SYN packet if signature matches*

*Push these Signatures to Router*

*Track signature*

| Signature | # email | # of ham | Ham Fraction |
|---|---|---|---|
| [T16:128:0:44:M536:.] | 14495869 | 2708 | 0.00019 |
| [16384:64:1:48:M1440,N,N,S:.] | 1562737 | 123 | 0.00008 |
| [S45:128:1:48:M1440,N,N,S:.] | 837353 | 72 | 0.00009 |
| [S4:64:1:60:M1460,S,T:!] | 7256520 | 2600718 | 0.358490 |
| [65535:128:1:48:M1460,N,N,S:.] | 3467107 | 951969 | 0.274571 |

Signature Tracking Table at SMTP server

*Calculate fingerprint based on TCP SYN packet*

*Determine if e-mail is ham, or spam*

# Pushing Signatures To Router

| Week | Action |
|---|---|
| 1 | Turn off spam filters (first 2 days), gather signature history, push to router on day 3 |
| 2 | Track Signatures |
| 3 | Track Signatures |
| 4 | Track Signatures |
| 5 | Push offending signatures to router – add to existing signatures |
| 6 | Track Signatures |
| 7 | Track Signatures |
| 8 | Track Signatures |
| 9 | Push offending signatures to router – add to existing signatures |
| 10 | Track Signatures |
| 11 | Track Signatures |
| 12 | Track Signatures |
| 13 | Turn off spam filters (first 2 days), add to tracked signature history, clear signatures from router, and push new list to router on day 3 |
|  | Repeat Week 2-13 |

# Data

- Two sites: University of Wisconsin- Madison and a corporation in Toyko, Japan
- Tcpdump – tracks all incoming TCP SYN connection packets
- SMTP logs – tracks on packets which pass greylisting and SpamAssassin scores them
- Logs are correlated across time

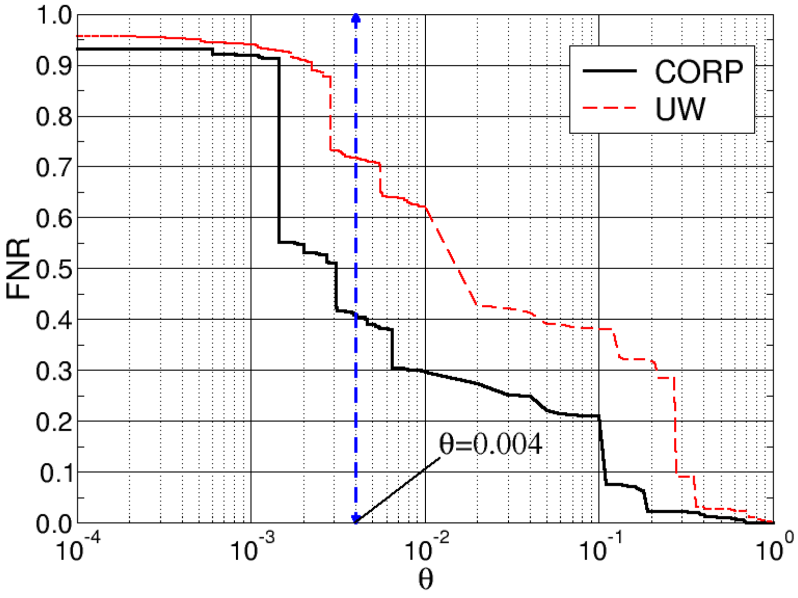
| Dataset | # senders | #delivered emails | #delivered spam | #delivered ham | #greylisted |
|---|---|---|---|---|---|
| UW | 7.4 Million | 26.2 Million | 13.3 Million | 12.3 Million | 87.8 Million |
| CORP | 3.1 Million | 2.0 Million | 1.3 Million | .5 Million | 18.8 Million |

# Extracting Signatures

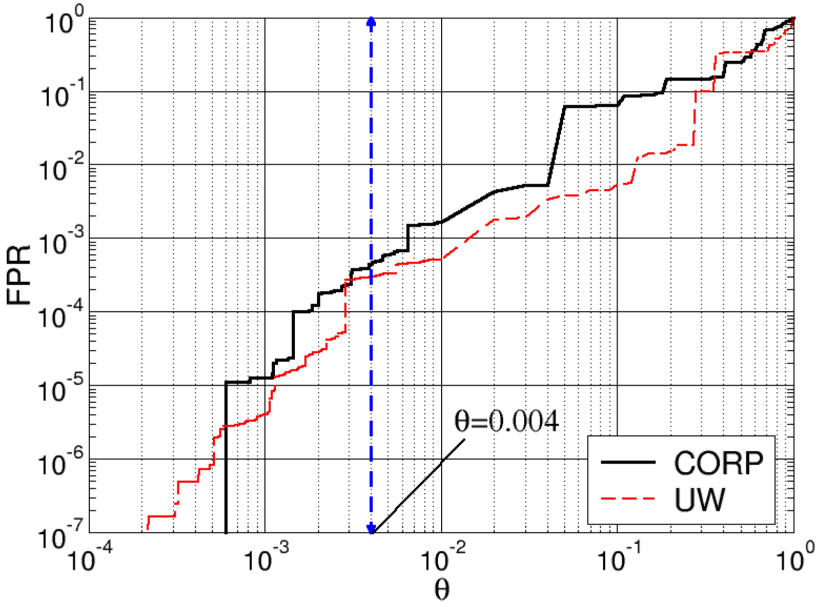
- Determine ham fraction threshold, θ,
  - Number of ham emails/number of total emails on a per signature basis
  - Should be a balance between a good false positive ratio (FPR) and good false negative ratio (FNR)
- Determine the signatures covered by above θ
  - Too much/little coverage?

# Our Signature Threshold

Graphs showing the performance of extracted signatures
under various thresholds



False Negative Ratio –
Fraction of missed spam messages
over the total of spam messages



False Positive Ratio –
Fraction of misclassified ham
messages over the total of ham
messages

# Results

# UW-Top 10 Spam Sending Signatures

| Signature | #Spam | #Ham | #Senders | OS Genre |
|---|---|---|---|---|
| [T16:128:0:44:M536:.] | 14,495,869 | 2708 | 260,955 | UNKNOWN/ Srizbi |
| [16384:128:1:48:M1440,N,N,S:.] | 1,562,732 | 123 | 20,308 | Windows |
| [S45:128:1:48:M1440,N,N,S:.] | 837,353 | 72 | 12,270 | Windows |
| [65535:64:1:52:M1452,N,W2,N,N,S:.] | 679,216 | 54 | 7,537 | UNKNOWN |
| [65535:128:1:48:M1442,N,N,S:.] | 468,074 | 14 | 8,328 | Windows |
| [65535:128:1:48:M1352,N,N,S:.] | 361,652 | 22 | 7,843 | Windows |
| [65535:64:1:52:M1440,N,W2,N,N,S:.] | 298,878 | 37 | 4,331 | Windows |
| [T16:128:0:44:M1360:.] | 262,077 | 21 | 3,147 | UNKNOWN/ Srizbi |
| [T16:128:0:44:M528:.] | 223,246 | 3 | 2,662 | UNKNOWN/ Srizbi |
| [65535:128:1:52:M1460,N,W1,N,N,S:.] | 210,267 | 45 | 3,261 | Windows |

*Fine grain signatures can expose some near only spam-sending signatures*

# Performance of Signatures

The top-100 signatures from April 2008
applied to their respective data sets.

| Set of Signatures | #Spam | %age Spam | #Ham | %age Ham | #Senders | %age Senders |
|---|---|---|---|---|---|---|
| UW | 24,797,823 | 28.2 | 3,485 | .03 | 403,568 | 5.5 |
| CORP | 11,249,690 | 59.8 | 243 | .05 | 1,639,667 | 52.9 |

*100 fingerprints can reduce the amount of spam by 28-59%*

# Signature Stability

Identified top-100 signatures from April 2008 and applied them to subsequent months.

## UW

| Month | Fraction of connections | Fraction of senders |
|---|---|---|
| Apr 2008 | 0.74 | 0.68 |
| Apr 2008 | 0.74 | 0.68 |
| May 2008 | 0.77 | 0.67 |
| June 2008 | 0.78 | 0.69 |

## CORP

| Month | Fraction of connections | Fraction of senders |
|---|---|---|
| Apr 2008 | 0.65 | 0.52 |
| Apr 2008 | 0.68 | 0.51 |
| May 2008 | 0.71 | 0.53 |
| June 2008 | 0.53 | 0.41 |

*Signatures were stable over four month period*

# Signature Accuracy

Performance of two signature sets:

Data Set:

UW

| Set of Signatures | #Spam | #Ham | #Senders |
|---|---|---|---|
| CORP Top 100 | 34,378,320 | 33,756 | 561,278 |
| UW Top 100 | 24,797,823 | 3,485 | 403,568 |
| INTERSECTION | 21,329,958 | 3,211 | 360,627 |
| UNION | 37,846,185 | 34,030 | 604,219 |

CORP

| Set of Signatures | #Spam | #Ham | #Senders |
|---|---|---|---|
| CORP Top 100 | 11,249,690 | 243 | 1,639,667 |
| UW Top 100 | 8,676,986 | 443 | 1,361,959 |
| INTERSECTION | 8,383,147 | 89 | 1,316,314 |
| UNION | 11,543,529 | 597 | 1,685,312 |

*Combining signature sets can increase accuracy and spam sender coverage*

# Signature X aka Srizbi

- [T16:128:0:44:M536:.] ~ [T16:128:0:44:M*:.]
- The top signature is in common among both data sets
    - Investigated separately because of the large amount of spam seen from this signature
    - Supported previous research that identified signature as part of the Srizbi botnet
    - Sends nearly all spam

# Attacks on the System

- Spoofing Signatures
  - Random Signatures
    - would cause signature tracking on SMTP servers to have millions of entries
  - Legitimate Signatures
    - would cause emails to get passed our filtering mechanism

# Related Work

- Ramachandran & Feamster
  - Uses TCP fingerprints to classify spam by OS
  - Studied spam from a sinkhole
- Beverly & Sollins
  - Used characteristics of SMTP flows

# Conclusions

- We have presented an *architecture* and *evaluation* of a router-level spam filter
  - Utilized two data sets
  - Showed that fine grain TCP fingerprints can significantly reduce spam volumes
  - Discovered additional Srizbi signatures
- Future work
  - Exploring the Srizbi signature in detail
  - A longer-term study of TCP fingerprints
  - A prototype version of our system

# Questions