Exploiting Transport-Level Characteristics of Spam

Robert Beverly[1] Karen Sollins

MIT Computer Science and Artificial Intelligence Laboratory

[1]now at BBN Technologies

{rbeverly,sollins}@csail.mit.edu

August 21, 2008

Conference on Email and Anti-Spam 2008

Outline

The Spam Arms Race

Attackers, scammers and thieves quickly adapt to defenses. Most effective solutions exploit *fundamental* weaknesses of attackers

Current Best Practices:

- Content Filtering ... response: modify word tokens
- Reputation Analysis ... response: dynamic, fresh addresses
- Collaborative Filtering ... response: mail uniqueness
- And the cycle continues: Authentication Schemes, computational puzzles, etc.

The Spam Arms Race

Attackers, scammers and thieves quickly adapt to defenses. Most effective solutions exploit *fundamental* weaknesses of attackers

Current Best Practices:

- Content Filtering ... response: modify word tokens
- Reputation Analysis ... response: dynamic, fresh addresses
- Collaborative Filtering ... response: mail uniqueness
- And the cycle continues: Authentication Schemes, computational puzzles, etc.

CSAIL

The Spam Arms Race

We propose a different approach:

- No panacea; existing solutions all have weaknesses
- Our solution, “SpamFlow,” is distinct from current practice

Question:

Are traffic characteristics a fundamental weakness of spam?

Hypothetical Question

Specifically:

- What is the *transport* (TCP/IP packet stream) character of spam?
- Are there *differences* between spam and ham flows?
- How to exploit differences in a way which spammers cannot easily evade?

Why ask this question?

Hypothetical Question

Specifically:

- What is the *transport* (TCP/IP packet stream) character of spam?
- Are there *differences* between spam and ham flows?
- How to exploit differences in a way which spammers cannot easily evade?

Why ask this question?

CSAIL

Transport-Level Characteristics of Spam

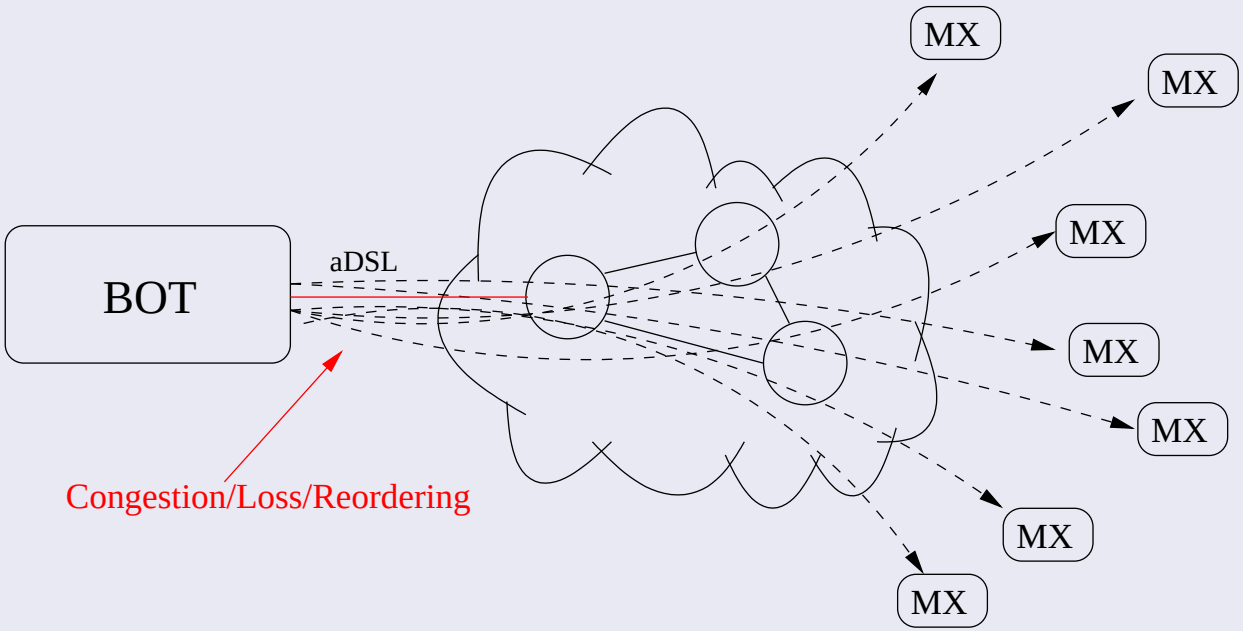
Two Observations

1. Low Penetration:
   - due to existing filters, user ambivalence
   - $\rightarrow$ huge volumes of spam
2. Sending Methods:
   - Open mail relays, email trojans, botnets, dialup
   - $\rightarrow$ Low asymmetric bandwidth, widely distributed

Transport-Level Characteristics of Spam

Combining Observations: Low Penetration + Sending Methods

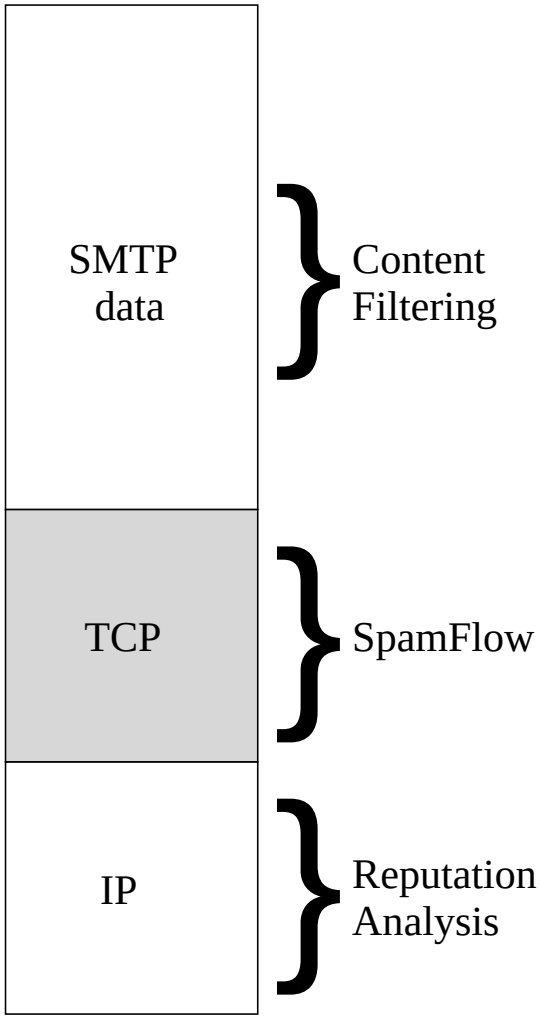
Volume + Methods + Economics → link/host resource contention



Contention:

Contention manifests as TCP/IP loss, retransmission, reordering, etc.

Understanding SpamFlow



- Not looking at IP header
- Not looking at data
- SpamFlow: TCP stream, incl timing
- (look at combining methods later)

Outline

A Brief Diversion on TCP/IP

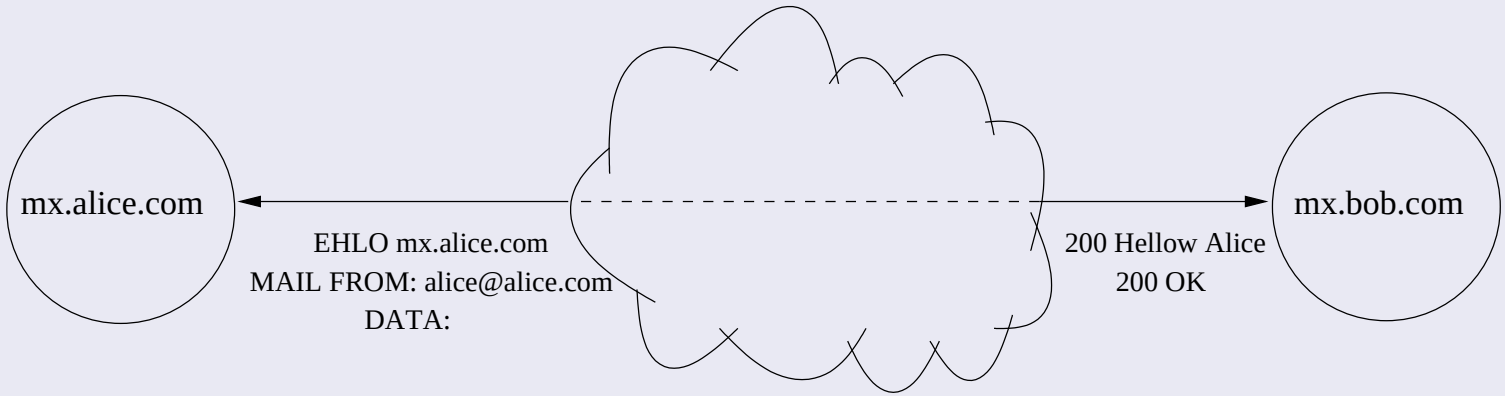
Transmission Control Protocol (TCP):

- Reliable, bi-directional, in-order byte transmission abstraction
  - Acknowledgments
  - State Machine
- Flow and congestion control
  - Reacts to loss, persistent congestion
- Multi-flow fairness and efficient resource utilization (AIMD)
  - Round trip time (RTT) estimation
  - Bandwidth probing

CSAIL

SMTP and TCP

Transmission Control Protocol:



- Simple Mail Transport Protocol (SMTP) uses TCP for transport
- Sequence of SMTP handshaking between Mail Transport Agents (MTAs)
- Mail contents are packetized

How do Spam Connections Behave?

Outline

How do Spam Connections Behave?

...or, a quick look at `netstat`

```
RcvQ   SndQ   Local     Foreign Addr                State
0      0      srv:25    92.47.129.89:49014          SYN_RECV
0      0      srv:25    ppp83-237-106-114.:29081    SYN_RECV
0      0      srv:25    88.200.227.123:25068        SYN_RECV
0      0      srv:25    92.47.129.89:49014          SYN_RECV
0      0      srv:25    ppp83-237-106-114.:29084    SYN_RECV
0      0      srv:25    88.200.227.123:25068        SYN_RECV
0      0      srv:25    88.200.227.123:25069        SYN_RECV
0      0      srv:25    88.200.227.123:25070        SYN_RECV
0      0      srv:25    88.200.227.123:25074        SYN_RECV
0      0      srv:25    84.255.150.15:4232          SYN_RECV
0      25     srv:25    222.123.147.41:50282        LAST_ACK
0      28     srv:25    adsl-pool-222.123.:1720     LAST_ACK
0      31     srv:25    222.123.147.41:50152        LAST_ACK
0      15     srv:25    222.123.147.41:50889        LAST_ACK
0      9      srv:25    88.245.3.19:venus           LAST_ACK
0      25     srv:25    78.184.155.70:1854          FIN_WAIT1
0      23     srv:25    190-48-30-225.spe:50920     FIN_WAIT1
0      23     srv:25    dsl.dynamic812132:48154     FIN_WAIT1
0      23     srv:25    ip-85-160-91-16.e:48093     FIN_WAIT1
0      23     srv:25    88.234.141.158:48389        FIN_WAIT1
0      23     srv:25    p5B0FBB5D.dip.t-d:11965     FIN_WAIT1
...
```

How do Spam Connections Behave?

...or, a quick look at `netstat`

```
RcvQ    SndQ    Local     Foreign Addr                 State
0       0       srv:25    92.47.129.89:49014           SYN_RECV
0       0       srv:25    ppp83-237-106-114.:29081     SYN_RECV
0       0       srv:25    88.200.2
0       0       srv:25    92.47.12
0       0       srv:25    ppp83-23
0       0       srv:25    88.200.2
0       0       srv:25    88.200.2
0       0       srv:25    88.200.2
0       0       srv:25    88.200.2
0       0       srv:25    84.255.1
0       25      srv:25    222.123.
0       28      srv:25    adsl-poo
0       31      srv:25    222.123.
0       15      srv:25    222.123.
0       9       srv:25    88.245.3
0       25      srv:25    78.184.1
0       23      srv:25    190-48-3
0       23      srv:25    dsl.dyna
0       23      srv:25    ip-85-16
0       23      srv:25    88.234.14
0       23      srv:25    p5B0FBB5D.dip.t-d:11965      FIN_WAIT1
...
```

TCP Stuck in States

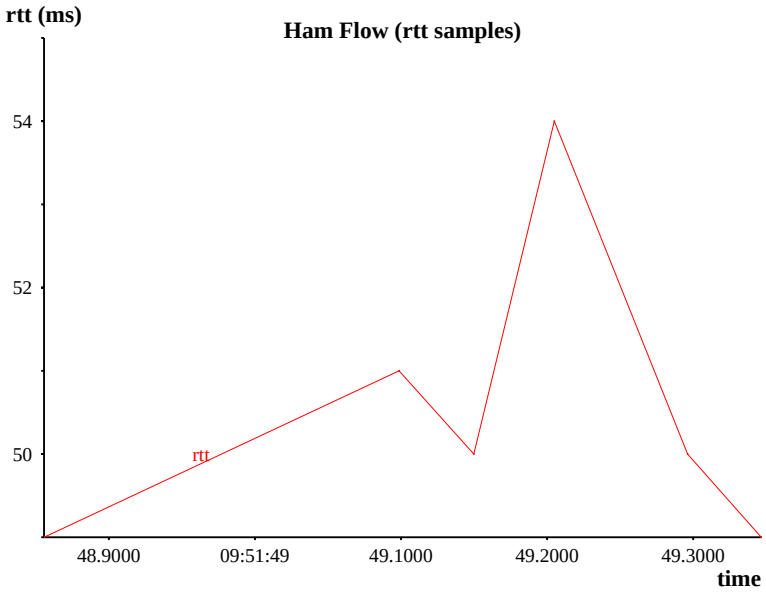
- Stays in these states for minutes
- Half-open connections
- Remote MTAs that “disappear” mid-connection
- Remote MTAs that send `FIN` and disappear

CSAIL

What about RTT?

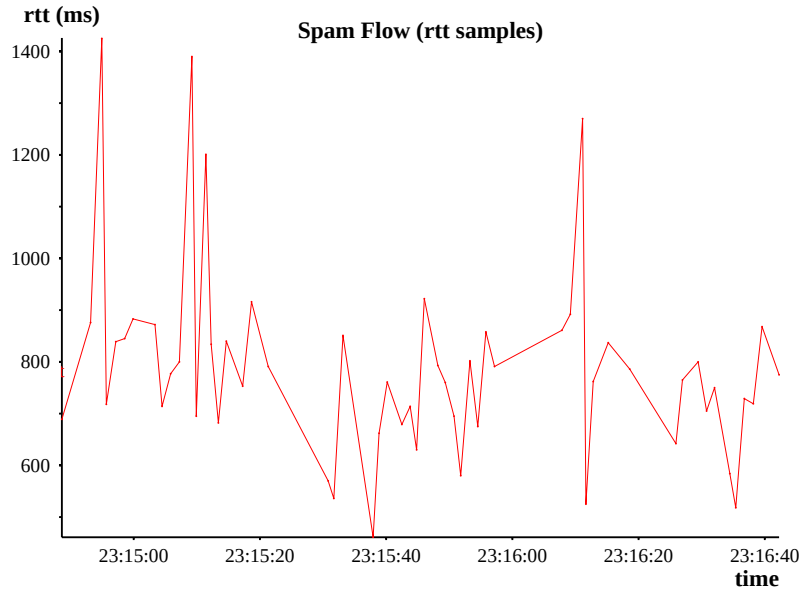
...building more intuition

```
Received:  from vms044pub.verizon.net
From:  "Dr.  Beverly, MD" <b@ex.com>
Subject:  thoughts
Dear Robert,
I hope you have had a great week!
```



Ham

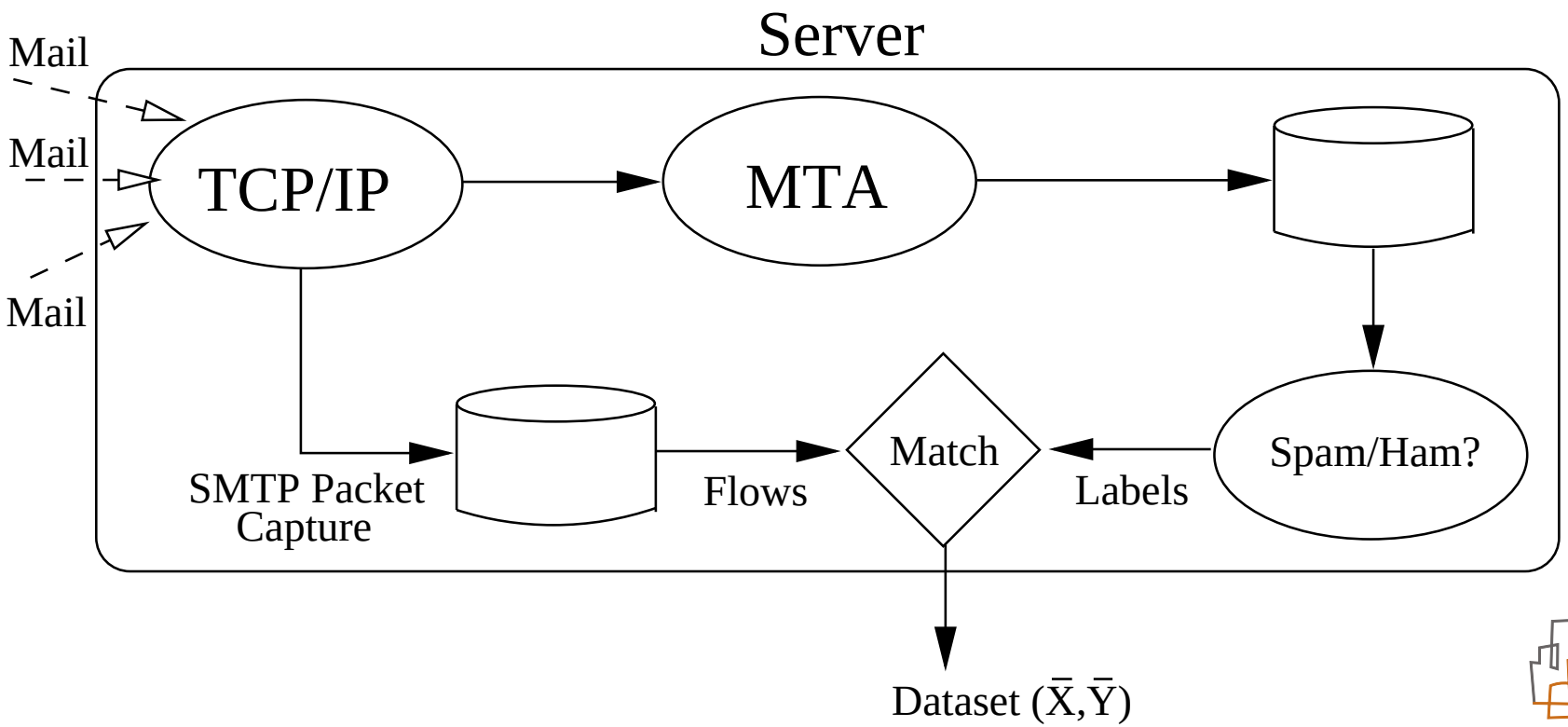
```
Received:  from unknown (59.9.86.75)
From:  Erich Shoemaker <ried@ex.com>
Subject:  Repl1ca for you
A T4g Heuer w4tch is a luxury statement
on its own.
In Prest1ge Repl1cas, any T4g Heuer...
```



Spam

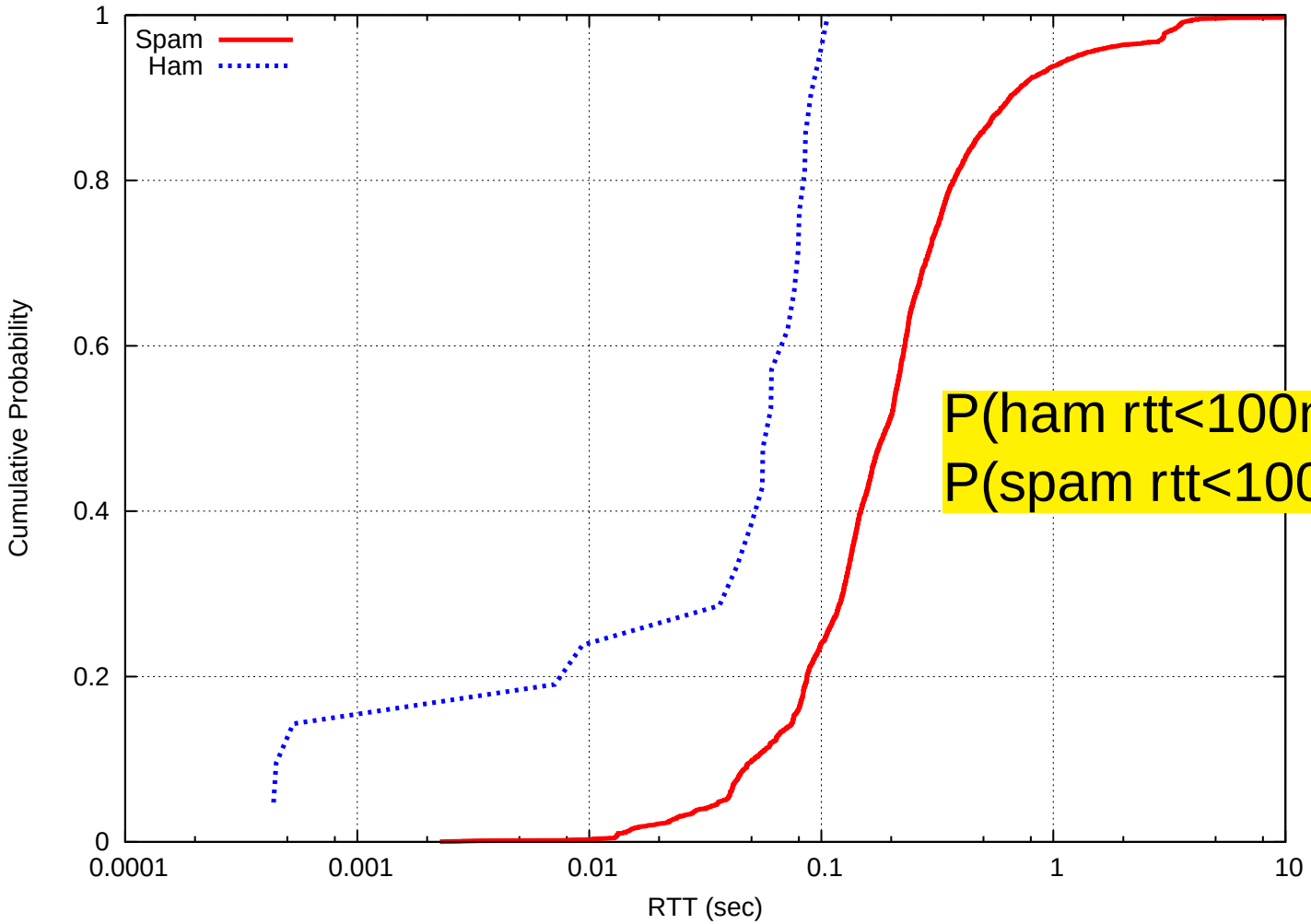
Data Collection

- Instrument a Mail Transport Agent (MTA) server
- Collect SMTP packet trace
- Match labeled emails to packet flows



Outline

Round Trip Time



P(ham rtt<100ms) ∼ 1;
P(spam rtt<100ms) ∼ 0.2!

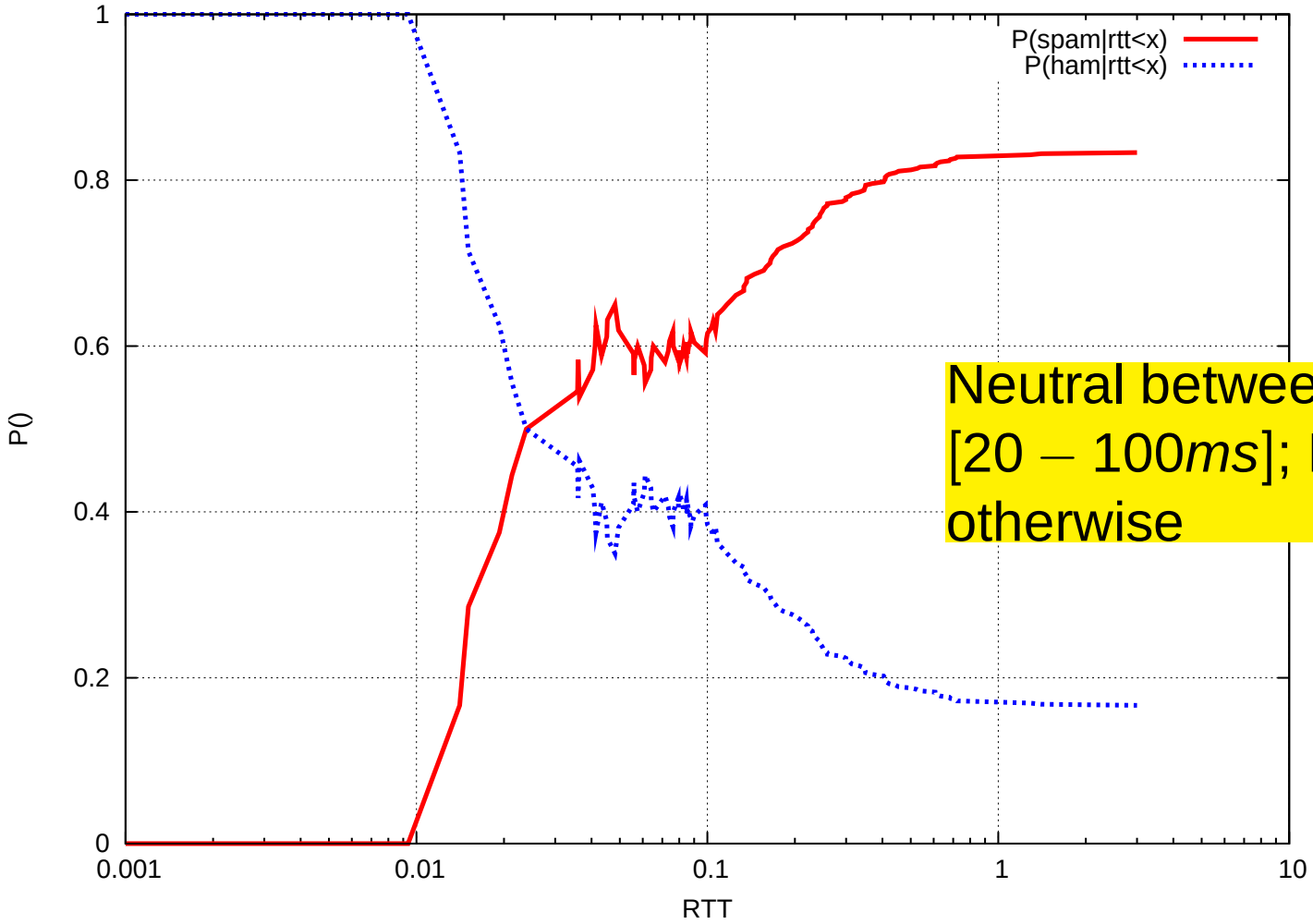
Round Trip Time
cont'd

Bayes' Rule

- Use causal information to form diagnosis

$$P(spam|rtt > x) = \frac{P(rtt > x|spam)P(spam)}{P(rtt > x)} \quad (1)$$

Round Trip Time
cont'd



Neutral between $[20 - 100ms]$; Highly biased otherwise

Selecting Features

Wait! You're disenfranchising distant servers!

- Yes; may be a good thing
- $\simeq 5\% > 1s$
- More importantly...

Other Transport "Features:"

- Packets, Retransmits, OutOfOrder, RSTs, FINs
- Zero Window, Minimum Cong. Window, Max Idle, Jitter, etc.
- *Adaptable* per-user, per-network

Key Insight

Statistical flow properties can provide differentiation

CSAIL

Selecting Features

Wait! You're disenfranchising distant servers!

- Yes; may be a good thing
- $\simeq 5\% > 1s$
- More importantly...

Other Transport "Features:"

- Packets, Retransmits, OutOfOrder, RSTs, FINs
- Zero Window, Minimum Cong. Window, Max Idle, Jitter, etc.
- *Adaptable* per-user, per-network

Key Insight

Statistical flow properties can provide differentiation

CSAIL

Outline

Non-Features

Non-Features

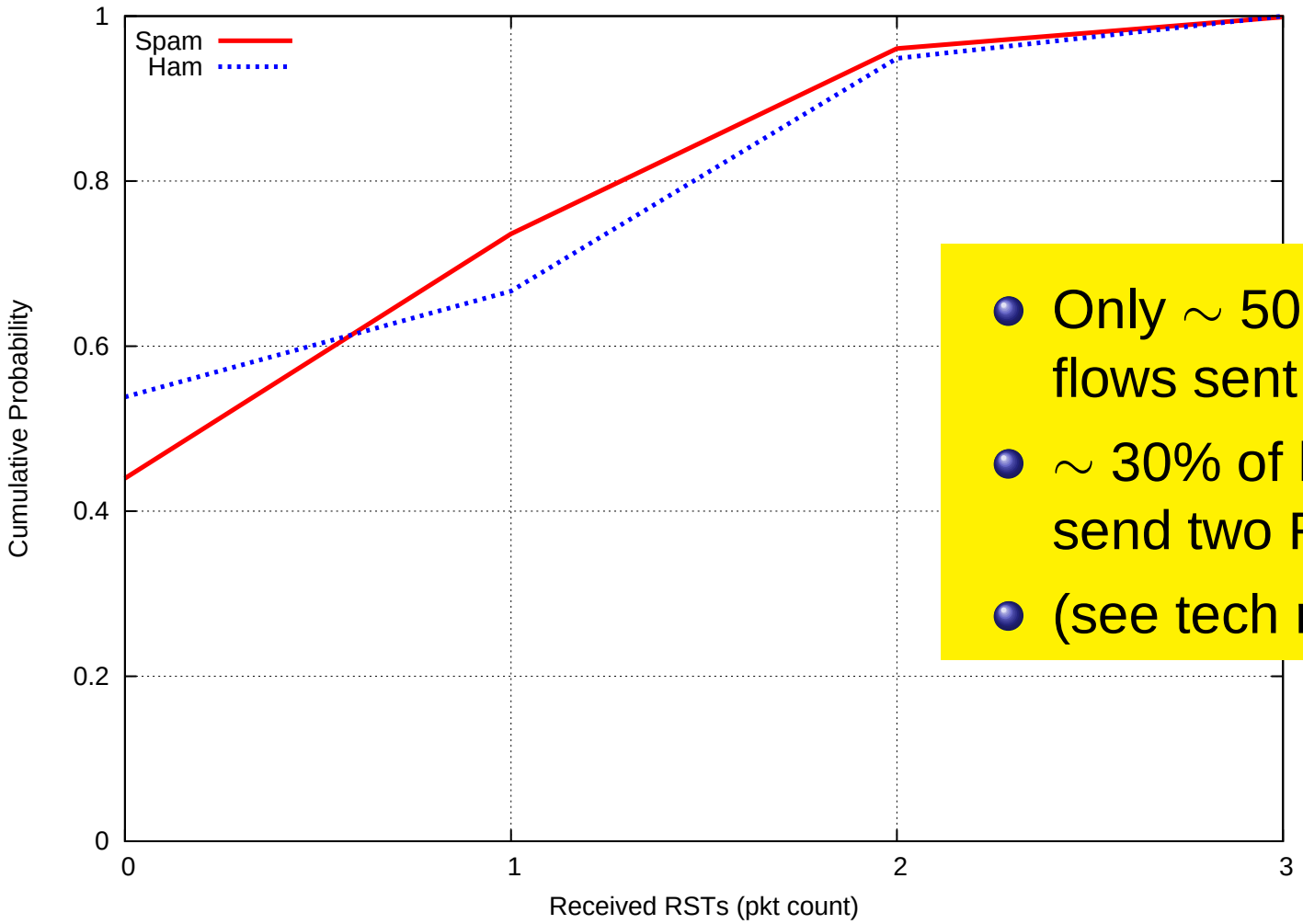
- Many intuitively “good” features turn out not to be
- Strength of statistical approach

One Example in Detail:

- RSTs as abortive close on socket
- A good indication of misbehaving flows?

Non-Features

Example: Received RSTs



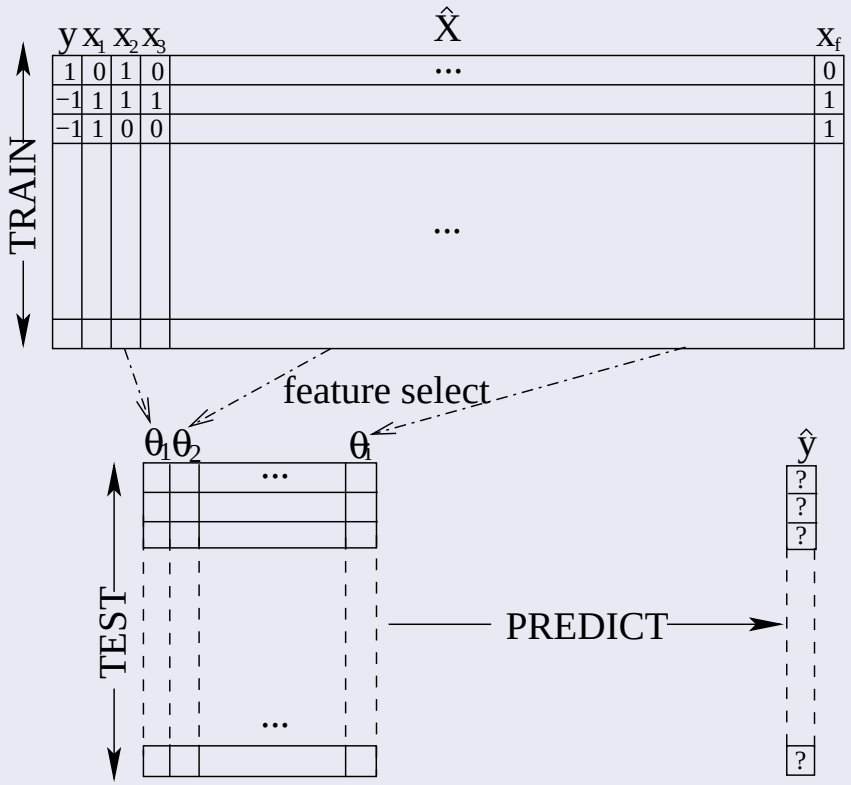
- Only $\sim$ 50% of ham flows sent no RSTs!
- $\sim$ 30% of ham flows send two RSTs!
- (see tech report for why)

Outline

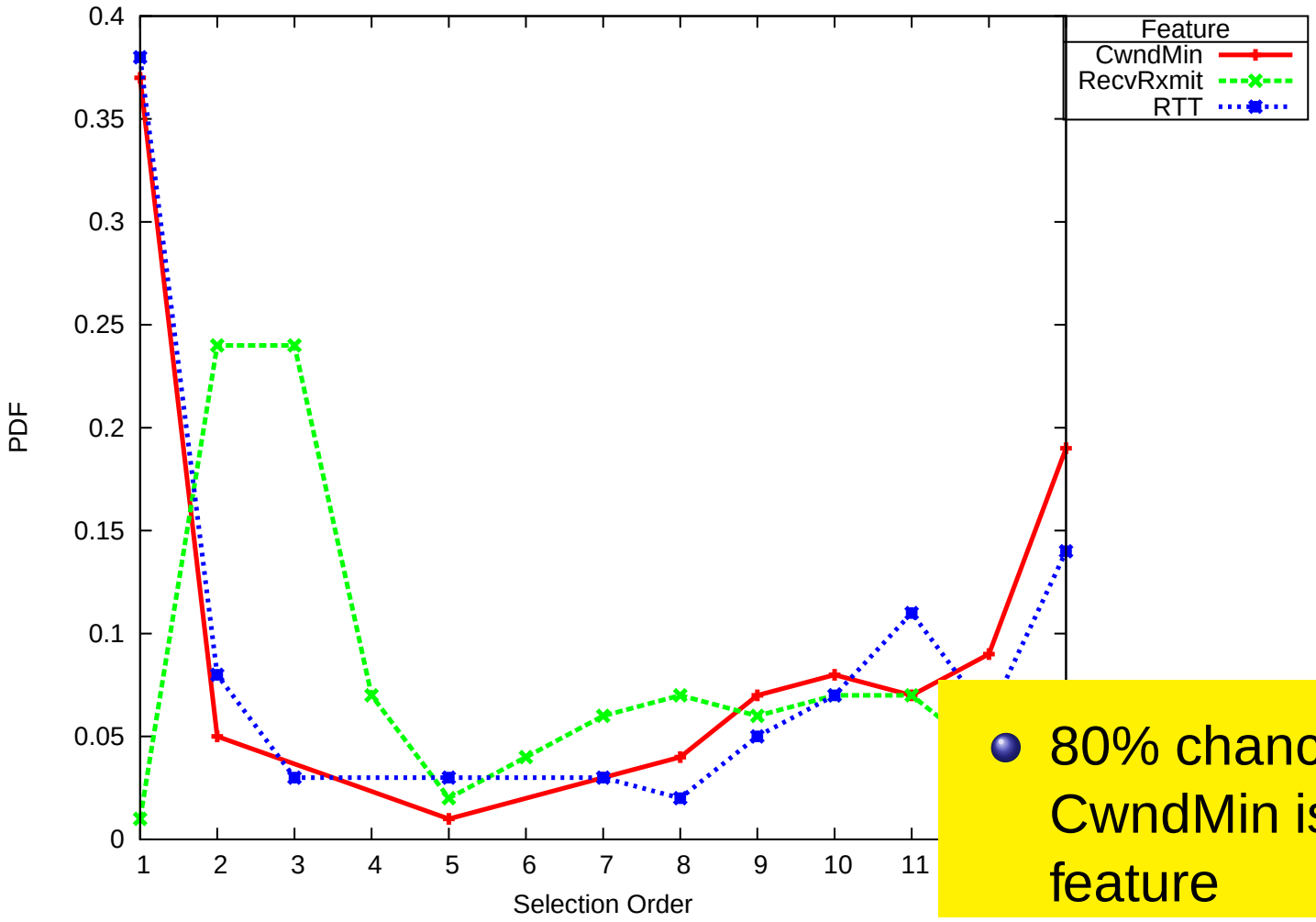
Picking Features

So, which features provide discrimination?

- Feature selection
- Simple method is forward fitting
- Greedily choose one available feature to minimize training error

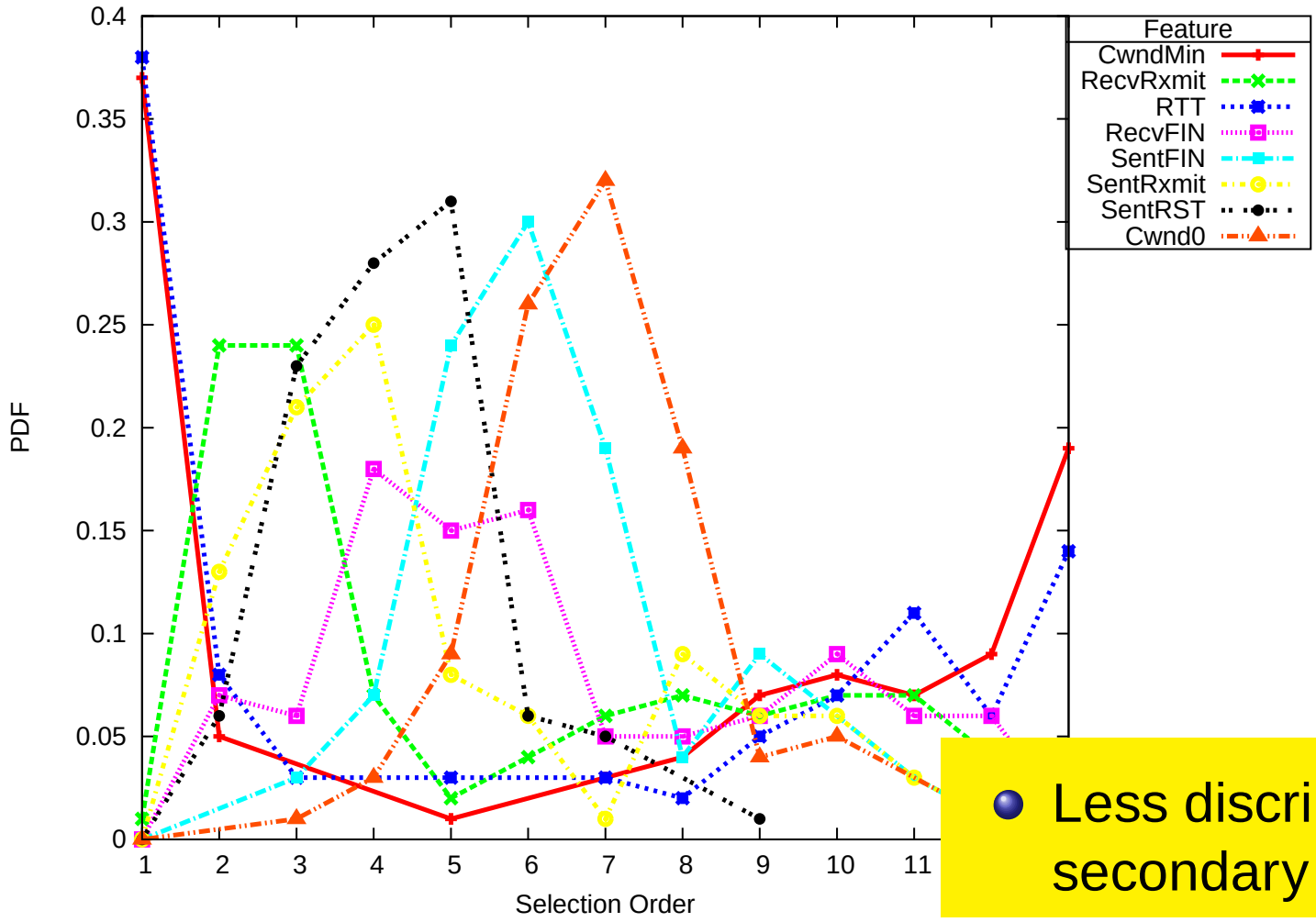


Picking Features
cont'd



- 80% chance that RTT or CwndMin is best single feature

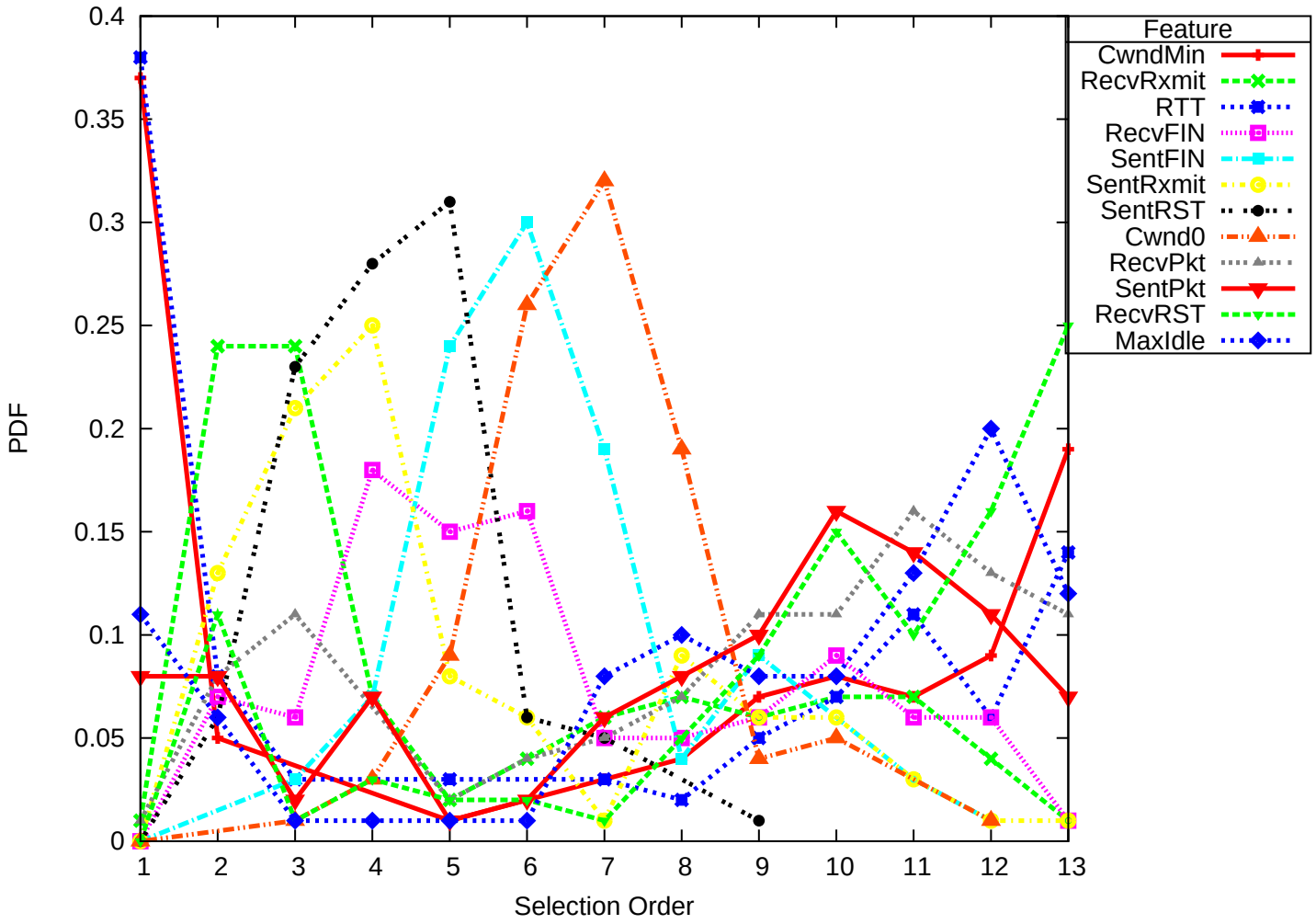
Features
cont'd



- Less discriminatory secondary features

CSAIL

Features
cont'd



Outline

SpamFlow

Based on observations, build a model

- Supervised learning, binary classification
- E.g. Bayes Nets, Support Vector Machines, etc.

SpamFlow

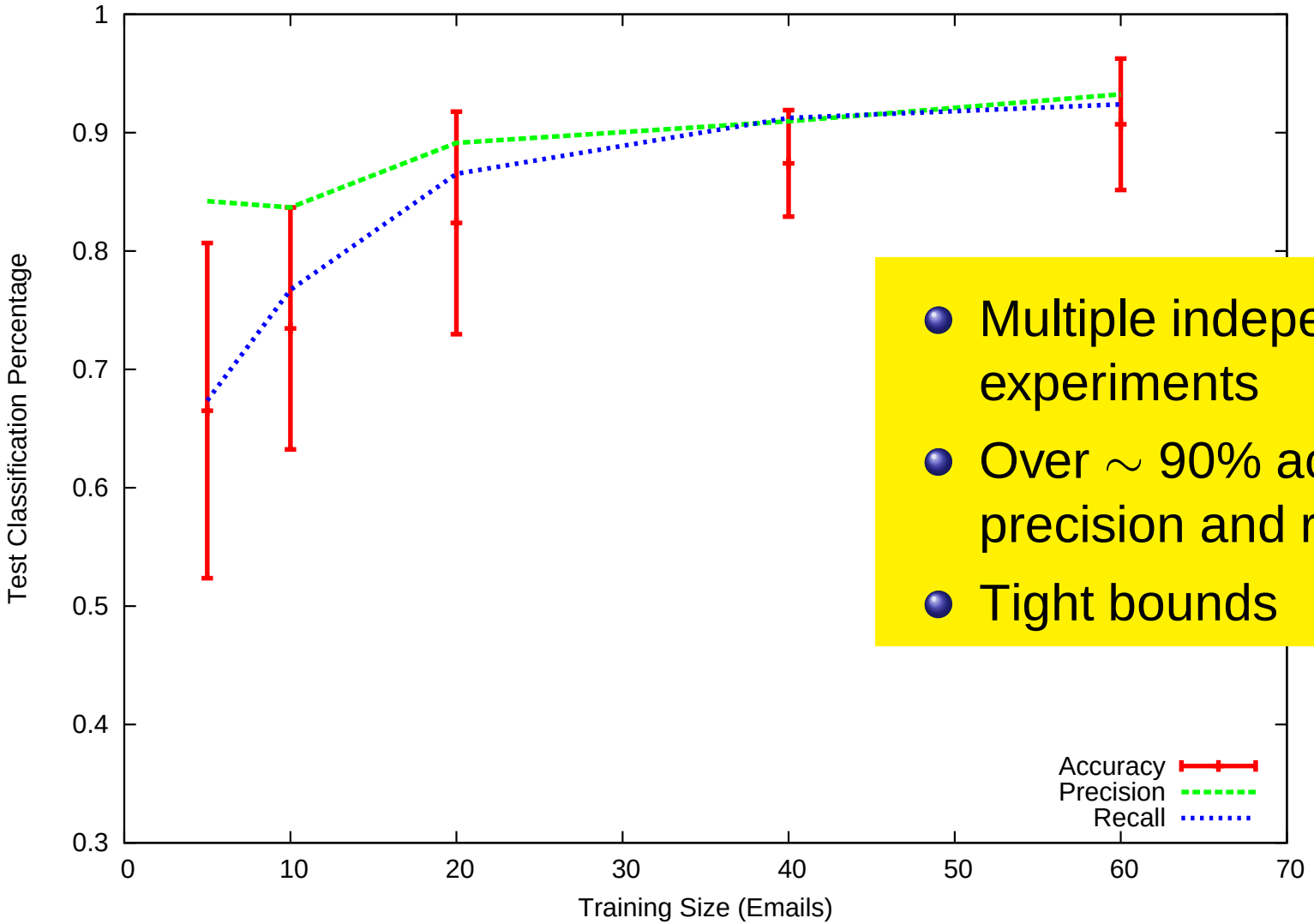
- A working implementation of the ideas using SVMs

Evaluation

- $FP =$ ham marked as spam
- $FN =$ spam marked as ham
- $accuracy = \frac{TP+TN}{P+N}$
- $precision = \frac{TP}{TP+FP}$

CSAIL

Prediction Performance



- Multiple independent experiments
- Over $\sim$ 90% accuracy, precision and recall
- Tight bounds

SpamAssassin False Negatives

False Negatives

- Against our data set, SpamAssassin gives 127 false negatives
- SpamFlow detects 78% of those
- $\longrightarrow$ useful to combine methods!

For example...

SpamAssassin False Negatives

```
Received: (qmail 12851 invoked from network); 24 Jan 2008 05:14:58 -0000
Received: from 201-213-46-215.net.prima.net.ar (201.213.46.215:8963)
  by ralph.rbeverly.net with SMTP; 24 Jan 2008 05:14:58 -0000
Received: from unknown (HELO deviant) (192.168.0.5) by mail6.colossal.com
 with SMTP; Thu, 24 Jan 2008 00:14:58 -0500
Date: Thu, 24 Jan 2008 00:14:58 -0500
To: rbeverly@grdata.com, rcmsjm@grdata.com, reb3@grdata.com, roots.nojunk@grdata.com, russell_s
From: "Jordan  Abrams" <inclusionVito@familyhistree.com>
Subject: Canadian Pharmcy Online! - 70-80% OFF!
Content-Length: 76
Lines: 6



Re" Your Pharmacy order # 85493899

Pls Go  ’ www.protectfair ’  dot com
```

SpamAssassin False Negatives

```
Received: (qmail 12851 invoked from network)
Received: from 201-213-46-215.net.prima.net.
  by ralph.rbeverly.net with SMTP; 24 Jan 200
Received: from unknown (HELO deviant) (192.16
 with SMTP; Thu, 24 Jan 2008 00:14:58 -0500
Date: Thu, 24 Jan 2008 00:14:58 -0500
To: rbeverly@grdata.com, rcmsjm@grdata.com, r
From: "Jordan  Abrams" <inclusionVito@familyh
Subject: Canadian Pharmcy Online! - 70-80% OF
Content-Length: 76
Lines: 6



Re" Your Pharmacy order # 85493899

Pls Go  ' www.protectfair '  dot com
```

SpamAssassin:

```
X-Spam-Status:  No,
score=3.5 required=5.0
tests=BAYES_50,
FS_OBFU_PRMCY,
SORTED_RECIPS,
UNPARSEABLE_RELAY
autolearn=no version=3.2.3
```

SpamAssassin False Negatives

```
Received: (qmail 12851 invoked from network); 24 Jan 2008 05:14:58 -0000
Received: from 201-213-46-215.net.prima.net.a
  by ralph.rbeverly.net with SMTP; 24 Jan 200
Received: from unknown (HELO deviant) (192.16
 with SMTP; Thu, 24 Jan 2008 00:14:58 -0500
Date: Thu, 24 Jan 2008 00:14:58 -0500
To: rbeverly@grdata.com, rcmsjm@grdata.com, r                    ll_s
From: "Jordan  Abrams" <inclusionVito@familyh
Subject: Canadian Pharmcy Online! - 70-80% OF
Content-Length: 76
Lines: 6



Re" Your Pharmacy order # 85493899

Pls Go  ’ www.protectfair ’  dot com
```

SpamFlow:

```
SntPkt:  45 RcvPkt:  29
SntRxmit:  0 RcvRxmit:  1
SntRST: 0 RcvRST: 0
SntFIN: 1 RcvFIN: 1
Cwnd0:  0 MinCwnd:  65280
MaxIdle:  1.366636
RTT: 0.162413
```

CSAIL

Open Questions

Spam is an Arms Race:

- How would spammers react?
- Adapt by slowing down, sending less mail
- Could spammers tweak TCP stacks and circumvent?

Future Work:

- Gather additional data sets
- Package, distribute
- Explore method's potential in other domains

Summary

- Attacking spam at a different layer
- Correct predictions with over 90% accuracy, precision and recall *without* content or reputation analysis
- SpamFlow finds 78% of SpamAssassin false-negatives
- No implementation hurdle, easily combined with existing techniques

Thanks!

Questions?


CSAIL

SpamFlow FAQ

1. *Can SpamFlow be more conservative in using RTT:* Yes, even a highly conservative filter can still leverage RTT to eliminate extremely large RTT spam flows.
2. *Doesn't SpamFlow privilege well-connected senders?* Personal, home or small business servers do not have the same volume requirement as spammers and thus are unlikely to induce the same TCP congestion effects we observe. SpamFlow only discriminates against sources that are *both* poorly connected *and* injecting large volumes of mail.
3. *What about email lists?* In contrast to spam, which must be sent continually, email list traffic can be scheduled in order to not cause local congestion.

Support Vector Machines

Dual-Form, Constrained Optimization:

$$\sum_{t=1}^{n} \alpha_t - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \mathbf{K}(\phi(x_i), \phi(x_j)) \text{ s.t. } C \geq \alpha_t \geq 0, \sum_{t=1}^{n} \alpha_t y_t = 0 \quad (2)$$

- Separate training set into two classes in most general way
- **Main insight**: find hyper-plane separator that maximizes the minimum margin between convex hulls of classes
- **Second insight**: if data is not linearly separable, take to higher dimension
- **Result**: generalizes well, fast, accommodate unknown data structure

CSAIL

What's going on here?

Example: Received RSTs

Google sends SMTP `QUIT`, then active close, then `RST`s passive close

```
11:55:57.807504 googl > srv:  P 187089:187095(6) ack 143 win 5720
11:55:57.807510 googl > srv:  F 187095:187095(0) ack 143 win 5720
11:55:57.807628 srv > googl:  .  ack 187096 win 32614
11:55:57.807863 srv > googl:  P 143:167(24) ack 187096 win 32614
11:55:57.808181 srv > googl:  F 167:167(0) ack 187096 win 32614
11:55:57.834759 googl > srv:  R 46149836:46149836(0) win 0
```

Yahoo! sends SMTP `QUIT`, srv performs active close. Yahoo! then sends three `RST`s when srv goes to `TIME_WAIT`

```
11:20:35.023406 srv > yahoo:  P 113:137(24) ack 1426 win 32120
11:20:35.023782 srv > yahoo:  F 137:137(0) ack 1426 win 32120
11:20:35.023983 yahoo > srv:  F 1426:1426(0) ack 113 win 33304
11:20:35.024073 srv > yahoo:  .  ack 1427 win 32120
11:20:35.076591 yahoo > srv:  R 776208340:776208340(0) win 0
11:20:35.076969 yahoo > srv:  R 776208340:776208340(0) win 0
11:20:35.077381 yahoo > srv:  R 776208341:776208341(0) win 0
```

Abortive close in Postfix source; normal behavior

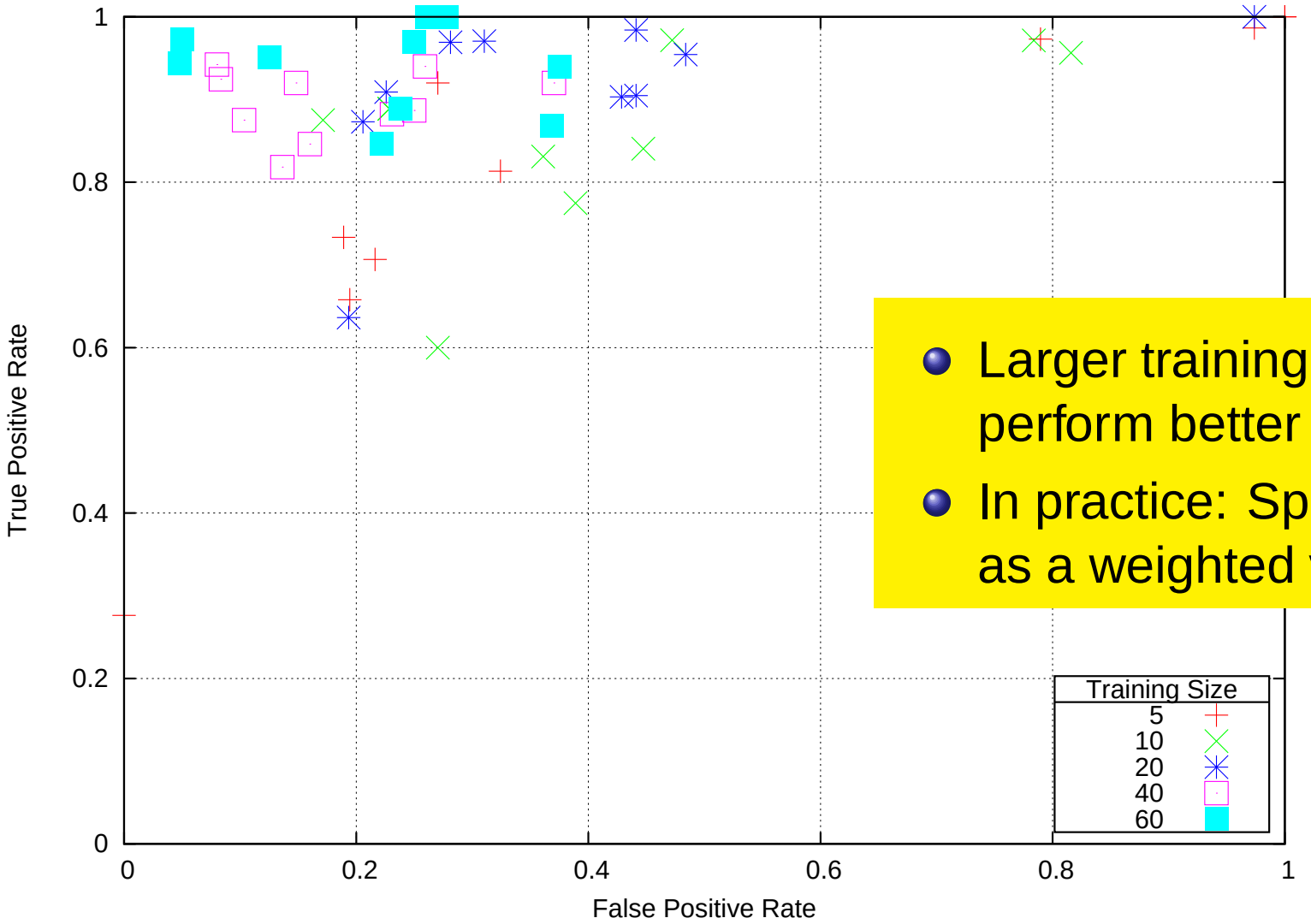
What's going on here?
Example: Received RSTs

Is abortive close a common "normal" SMTP technique?

Postfix Source

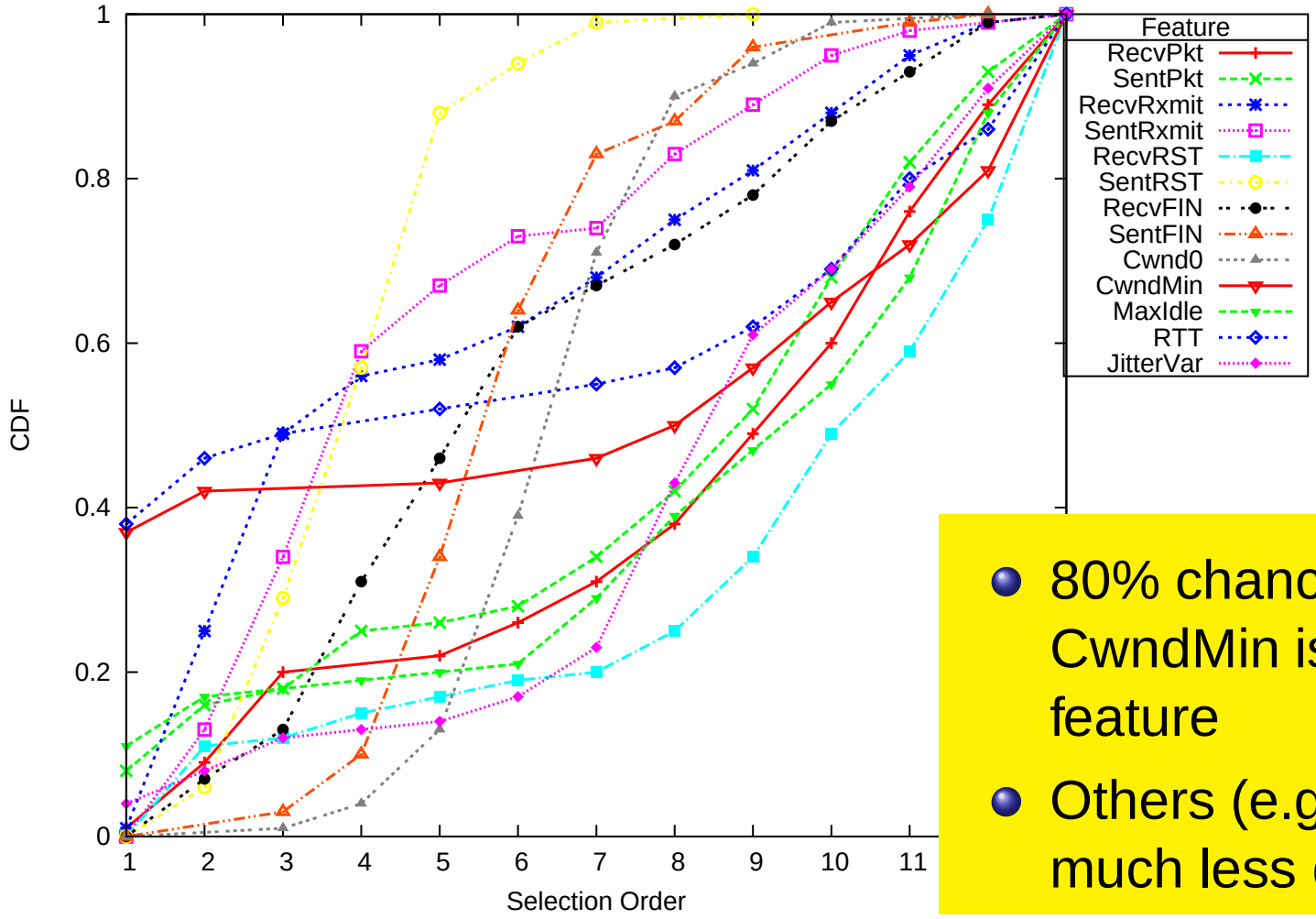
```
static void start_connect(SESSION *session) {
   int fd;
   struct linger linger;
   linger.l_onoff = 1;
   linger.l_linger = 0;
   if (setsockopt(fd, SOL_SOCKET, SO_LINGER, (char *) &linger,
     sizeof(linger)) < 0)
     ...
```

ROC Curve



- Larger training sizes perform better
- In practice: SpamFlow as a weighted voter

Features
cont'd



- 80% chance that RTT or CwndMin is best first feature
- Others (e.g. RecvRST) much less discriminatory

CSAIL

Data Collection

Dataset:

- One week, January 2008
- $\sim 18k$ emails, only ~ 200 legitimate ham
- Normalize spam and ham count for each experiment, randomly select spams
- Dataset is small; future work corrects this

This talk: method, intuition, validation