Still Curious about Anti-Spam Testing?

Here's a Second Opinion

David Koconis, Ph.D.
Senior Technical Advisor, ICSA Labs
01 October 2010

Outline

- **Introduction**
- **Anti-spam testing synopsis**
- **Components of meaningful testing**
- **Anti-Spam Testing Methodology**
  - Legitimate email corpus
  - Store-and-forward versus live testing
  - Observations
- **Comparison to VBSpam**
- **Conclusion & Future**

Introduction

- **ICSA Labs and me**
- **Enterprise anti-spam products**
- **What was the original diagnosis?**
  - Comparative
  - Unbiased
  - Real email in real-time
  - Statistically relevant (i.e., large corpus)
  - Explain what was done

Definitions

- **Effectiveness**
  - Percent of all spam messages identified as such and not delivered
- **False Positive**
  - Legitimate email misclassified as spam and not promptly delivered
- **False Positive Rate**
  - Percent of all legitimate messages not promptly delivered
- **Corpus (Corpora)**
  - Collection of email messages typically having some property in common

Anti-spam Testing Synopsis

- **Number of spam messages on the Internet far exceeds number of legitimate messages**
- **Want solution that**
  - blocks every spam message (100% effective)
  - promptly delivers every legitimate email (0 false positives)
- **But Nobody's perfect**
- **Legitimate email does get blocked/delayed**
  - End users get mad, Support cost, Missed opportunity
- **Spam gets delivered**
  - Storage and time wasted, possible malicious content
- **Which solution works best?**
- **How can solutions be improved?**

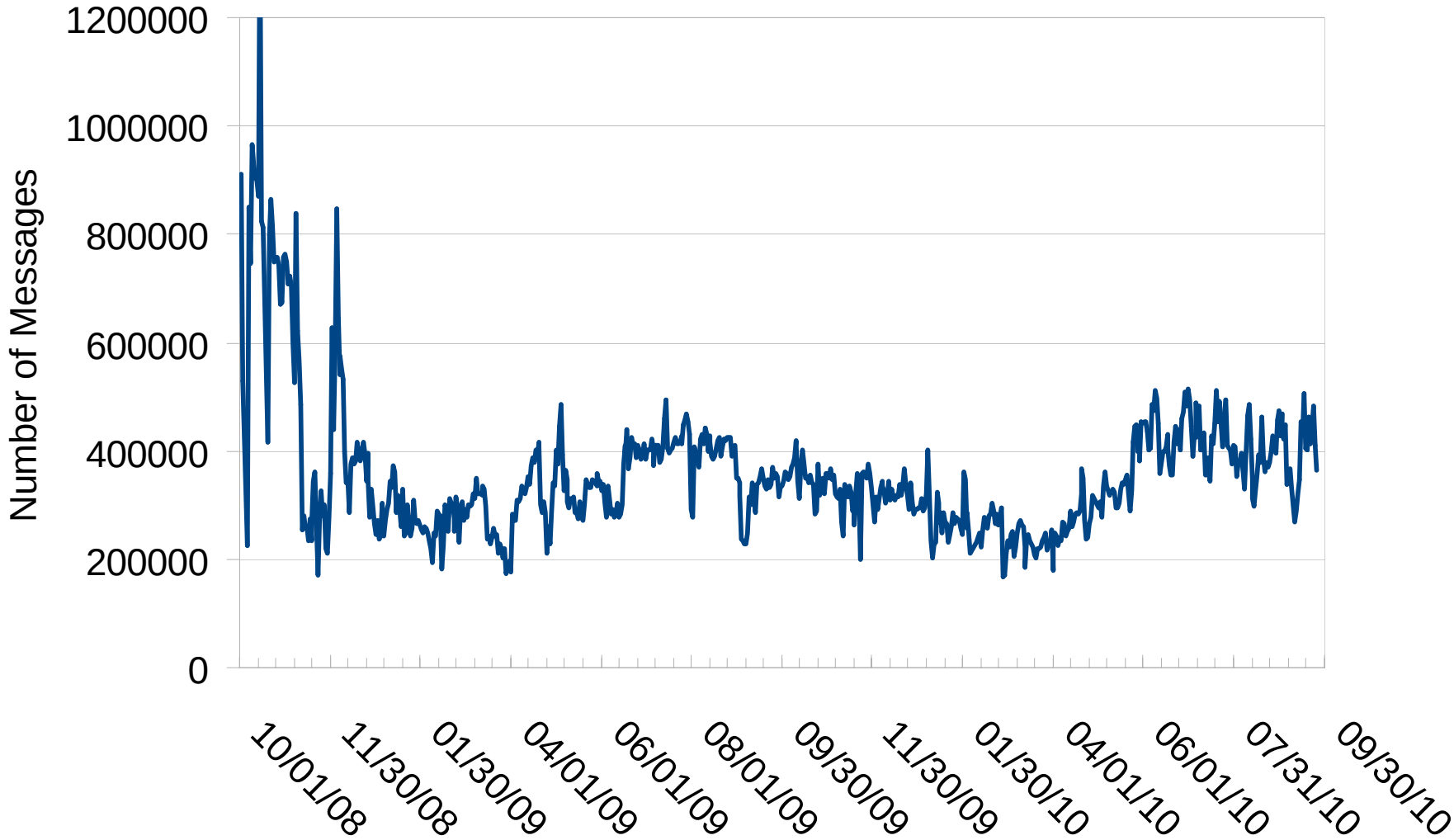
What is needed for meaningful anti-spam testing?

- **Lots of appropriate spam**
  - Continually updated corpus
  - Representative of what is seen on the Internet
- **Lots of legitimate email**
  - Personal and subscription lists or newsletters
  - If possible, not proprietary
- **Test methodology that mirrors deployment**
  - Products under test able to query Internet resources
    - »Protection updates
    - »DNS, RBL, SPF, etc
- **Detailed logging and dispute resolution**

Lots of Spam - ICSA Labs Corpus

- **Spam Collector**
  - Internet connected gateway MTA honeypot
  - Pointed to by multiple valid MX records
  - Accepts SMTP connection and generates unique identifier
  - Adds “Received:” header
  - Stores message headers, data and envelope
- **Messages arrive continually**
  - Triggers syslog message and DB insert
    - »Arrival time, Filename, Classification
- **Directory rolled at midnight**
  - Rsync’ed to analysis server
  - Analyze entire corpus

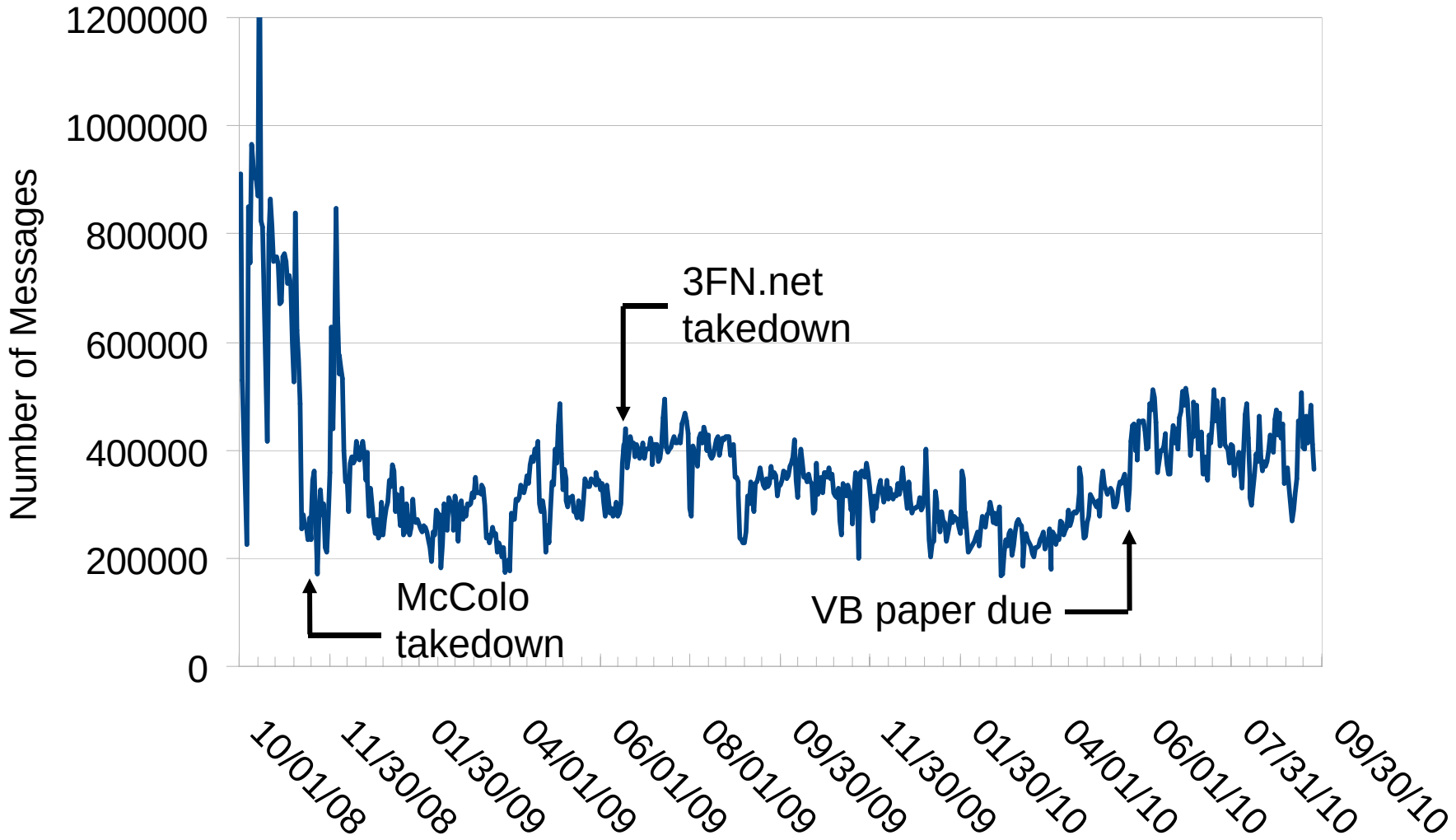
Daily Message Volume at ICSA Labs Spam Trap



Daily Volume vs. events and predictions

- **ISP take downs**
  - November 2008 (McColo)
    - »Media reports spam volume decreased 35-80%
  - June 2009 (3FN.net)
    - »Media reports smaller, if any decrease (spammers learned lesson)
- **Volume predictions for 2010**
  - Peaked in mid 2009 and then returned to 2008 levels
    - »McAfee threat report for Q1 2010
  - 30~40% increase in spam from 2009-2010
    - »Cisco 2009 annual report

Daily Message Volume at ICSA Labs Spam Trap



Message Analysis

- **Extract & save interesting message properties**
  - Sender, recipient(s), size, subject, source, body digest
- **MIME type headers**
  - has attachment? What type?
- **Classification**
  - Most are spam
  - Special accounts for Newsletter subscriptions & Project Honeypot feed
- **Decide if suitable for use in test set**
  - RFC compliant addresses
  - Not duplicate message
  - Not relay attempt

The 10 Worst Spam Countries

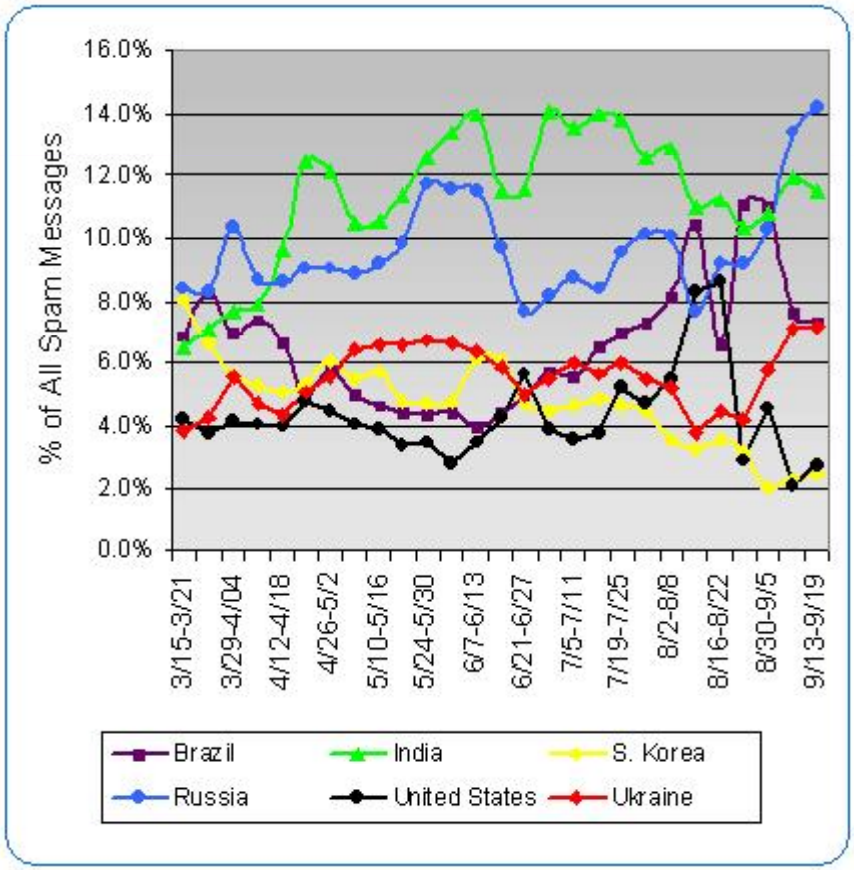
As at 31 May 2010 the world's worst Spam Haven countries for production and export of spam are:

1	**United States**	Number of Current Live Spam Issues: 2303
2	**China**	Number of Current Live Spam Issues: 692
3	**Russian Federation**	Number of Current Live Spam Issues: 479
4	**United Kingdom**	Number of Current Live Spam Issues: 292

"… compiled from the SBL database using the number of currently listed SBL records for each network (ISP/NSP) sorted by country."

Data from Spamhaus 31-May-2010, http://www.spamhaus.org/statistics/countries.lasso

Spam message source



- **Source means IP that connected to ICSA Labs**
- **Where does the U.S. rank?**
  - First by far
    - »Spamhaus, Symantec
  - First, but only by a hair
    - »Sophos
  - Second
    - »Cisco 2009
  - Not even top 5
    - »Panda Security
    - »ICSA Labs

From ICSA Labs Spam Data Center
https://www.icsalabs.com/technology-program/anti-spam/spam-data-center

Lots of Legitimate Email

- **Legitimate email separated into 2 categories**
- **Newsletters**
  - Subscribe to press releases, announcements and newsletters
    - »Google Alerts, Bankrate.com, U.S. State Department, etc.
  - Messages arrive at spam collector with unique RCPT
- **Person-to-person email**
  - Business related
    - »Meeting minutes, sales forecast, customer queries
  - Non-business related
    - »After hours or weekend plans, family photos, etc.
  - One or more recipients
  - Occasional attachments

Legitimate email generation framework

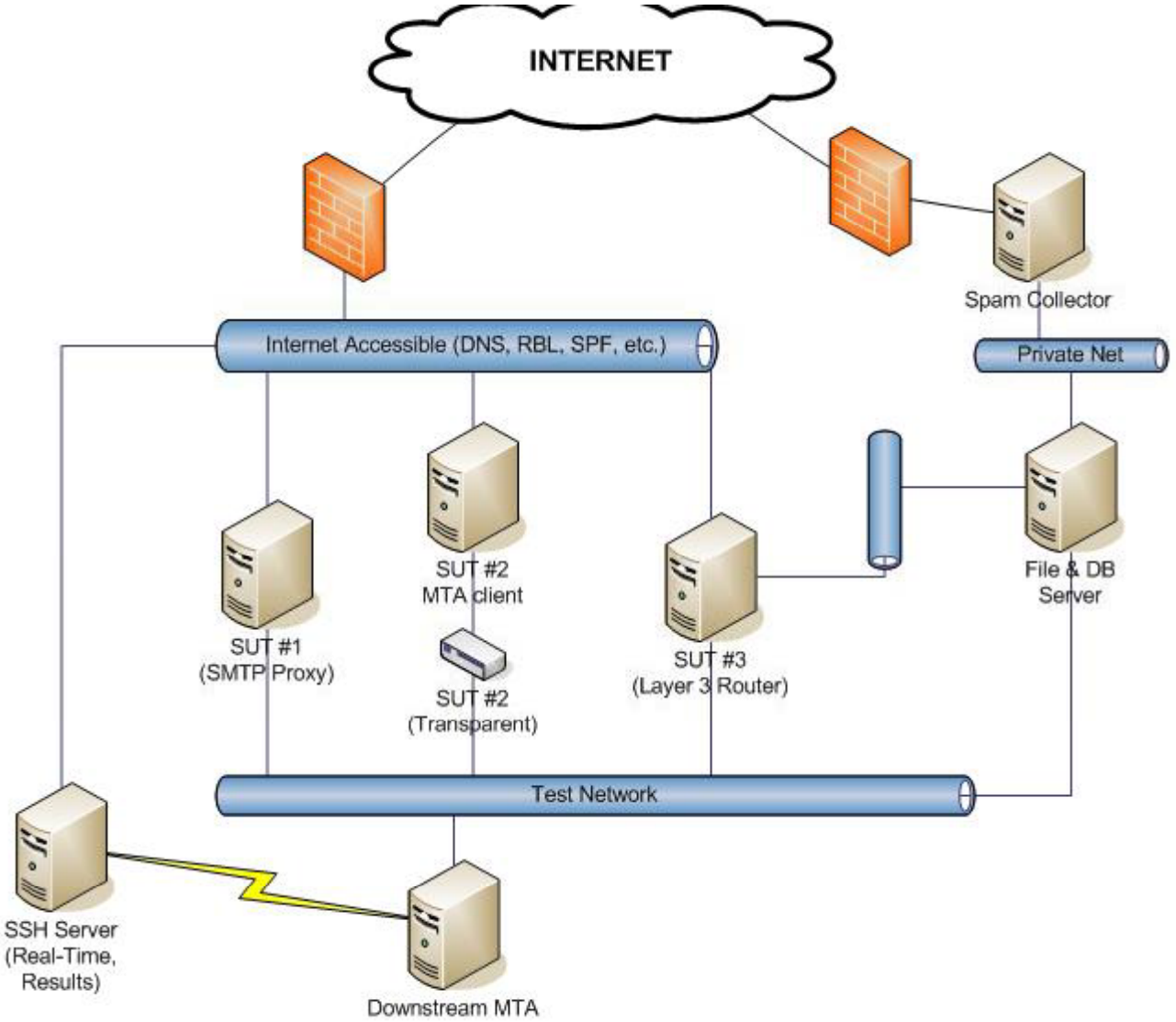
- **Message bodies from real email**
  - list postings, non-proprietary msgs, personal accounts
- **Assorted MIME types**
  - 40% text/plain, 40% text/html, 20% multipart/alternative
- **Repository of attachments**
  - 15% get attachment
- **Sender and Recipient addresses in DB table**
  - Users: Name, address, title
  - Companies: MX host, domain, email address convention, SPF
- **Number of recipients probability-driven**
  - 80% single recipient, 20% up to 4

Legitimate email generation framework (cont.)

- **Isn't this what spammer's are trying to do?**
  - Yes, but
- **It's our MTA receiving messages**
  - Received header passes SPF check
  - Other SMTP headers also valid
- **Not used for newsletter ham**
- **No malicious content attachment**
- **Product developers can appeal**
  - Results are available in real-time

Spam Testing Methodology

- Test bed overview


INTERNET
Spam Collector
Internet Accessible (DNS, RBL, SPF, etc.)
Private Net
SUT #2
MTA client
File & DB
Server
SUT #1
(SMTP Proxy)
SUT #3
(Layer 3 Router)
SUT #2
(Transparent)
Test Network
SSH Server
(Real-Time,
Results)
Downstream MTA

Spam Testing Methodology

- Test bed overview
- Message test set determination

Anatomy of a test set

- **Message order driven by probabilities**
  - Main classification (90% spam / 10% ham)
  - Secondary classification of ham (95% personal / 5% newsletter)
- **First decide how many messages in the set**
- **Start with first message pick classification**
- **Then identify message file**
- **Repeat**

Spam Testing Methodology

- **Test bed overview**
- **Message test set determination**
- **Evolution of the testing process**
  - Began with store-and-forward
  - Transitioned to Live

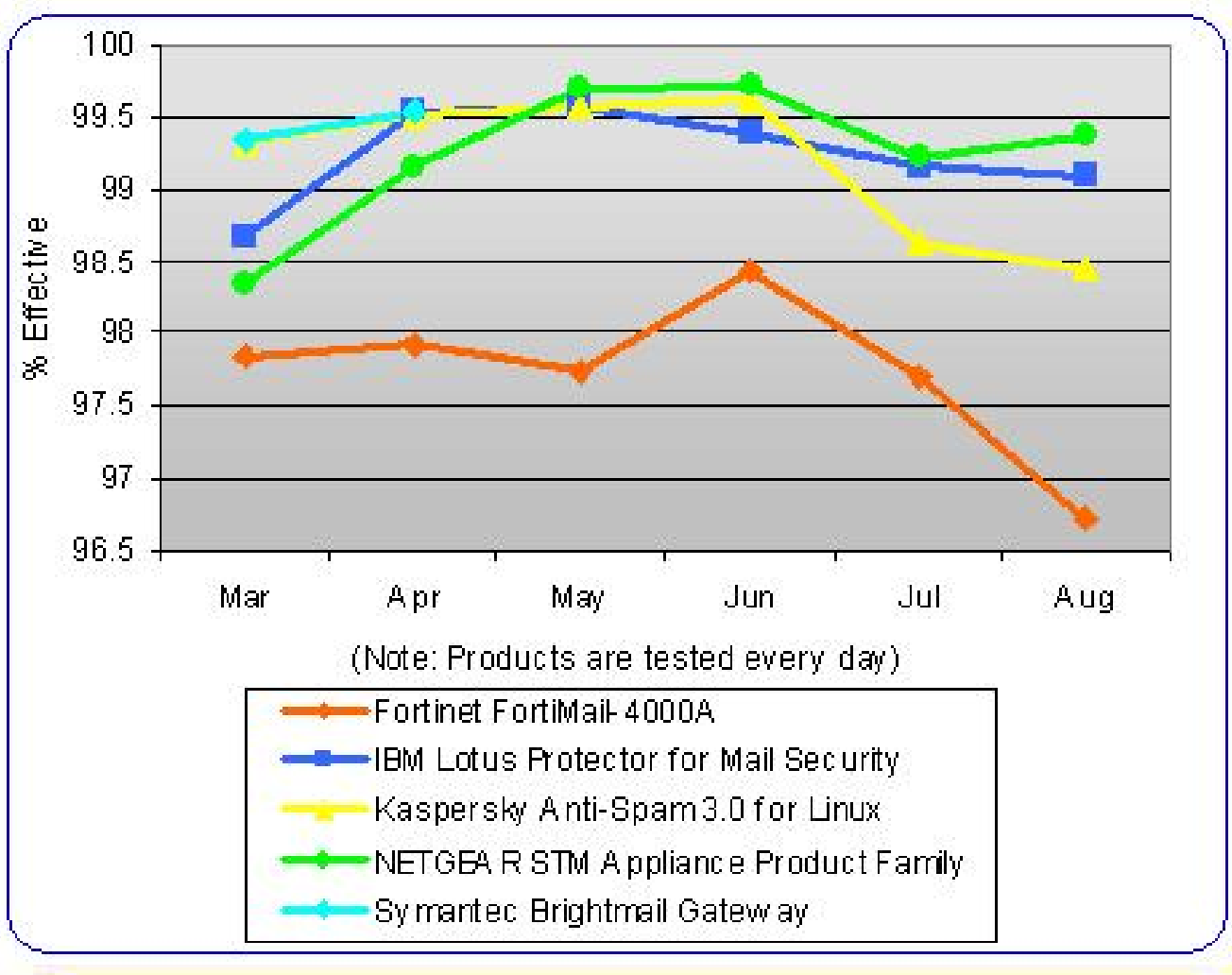
Store-and-Forward Testing (batch)

- Wait for whole spam corpus from previous day to be analyzed
- Generate corpus of legitimate messages
- Assemble message test set
- Test daily beginning at 0300
- Every product sees same messages in same order
- But faster products finish earlier

Transitioned to Live Testing

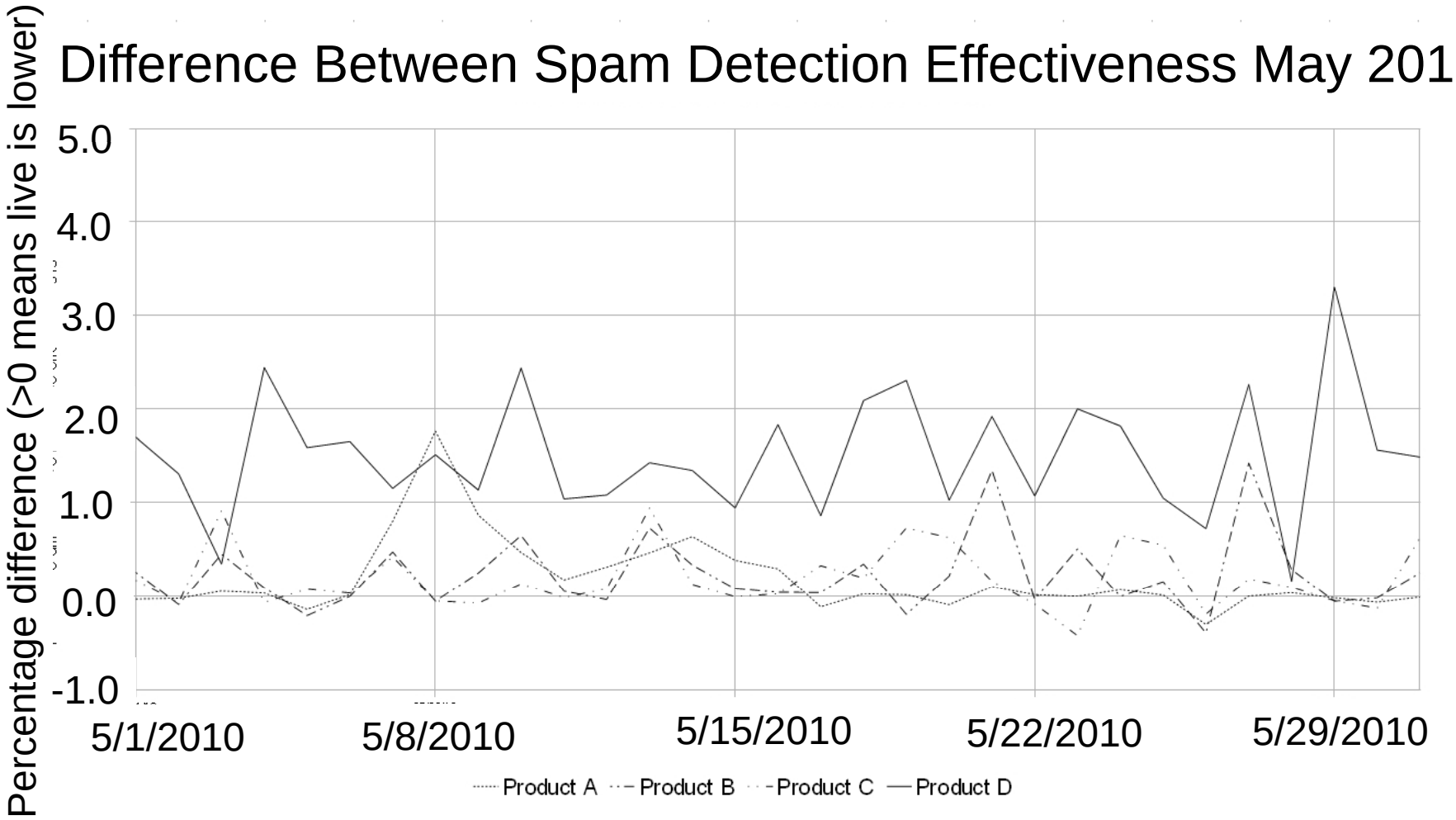
- **Predetermine message set classification order**
- **Proceed through list and**
  - Retrieve message from spam collector in real-time

  or
  - Generate legitimate personal message
- **Analyze it on-the-fly (only essential checks)**
- **Initiate connection to every product at the same time for every message**
- **Execute live test event twice daily (0300, 1700)**



From ICSA Labs Spam Data Center
https://www.icsalabs.com/technology-program/anti-spam/spam-data-center

Difference Between Spam Detection Effectiveness May 2010



Lessons learned

- **Measured Spam Effectiveness Differs**
  - Always better with stored corpus
  - But, relative ranking of products was same
- **Suggests that delay allows propagation of signature/knowledge to device being tested**
- **Misclassified messages included in batch test set**
  - 2nd Exposure effectiveness
  - No correlation between age of message and length of delay
- **However, products sometimes forget**
  - A spam message blocked in live test is later delivered

Comparison to VBSpam

- **Similarities**
  - Relay messages to products from single IP
  - Include original src IP, etc. in Received header
  - Require tested product to make a decision (not quarantine)
  - Use “live” spam feed
  - Disallow Whitelisting of senders

Comparison to VBSpam

- **Differences**

	ICSA Labs	VBSpam
Message delivery rate	~2300/hr	~600/hr
Spam feed	On-site MTA	PHP, Abusix
Message classification	Pre-classified (before)	By consensus (after)
Frequency	Daily (11.5 hours/day)	Quarterly (24/7 for 3 wks)
Pre-DATA filtering?	IP in Received header	XCLIENT extension
Final Score	Report Effectiveness & FP	Combined measure

And one more …

There's more than effectiveness and false positives

- **You're kidding. Right?**
- **Shouldn't there be**
  - Authenticated access to administer the product over the network
  - A way to configure the network settings
  - A way to change or configure the policy being enforced
  - Automatic spam protection updates
  - Logging of
    - password changes to an administrative account
    - attempts by a remote user to authenticate (success/failure)
    - message delivery decisions
  - Sufficient and accurate documentation
- **List of criteria requirements developed with consortium input**
- **Methodology includes test cases to verify each requirement in the criteria**

Conclusion & Future Work

- **Creating a fair, accurate unbiased test requires considerable expertise and development**
- **Testing with stored spam corpus may overestimate the effectiveness products**
- **Investigate sensitivity to time of test**
  - Effectiveness better during business hours or at night?
  - On weekdays or weekends?
- **Incorporate more spam feeds**
  - Project Honey Pot
  - Verizon Cloud Services