ON THE EFFECTIVENESS OF IP REPUTATION FOR SPAM FILTERING

COMSNETS 2010

January 9, 2010

Holly Esquivel (UW-Madison, USA)

Tatsuya Mori (NTT, Japan; presenter)

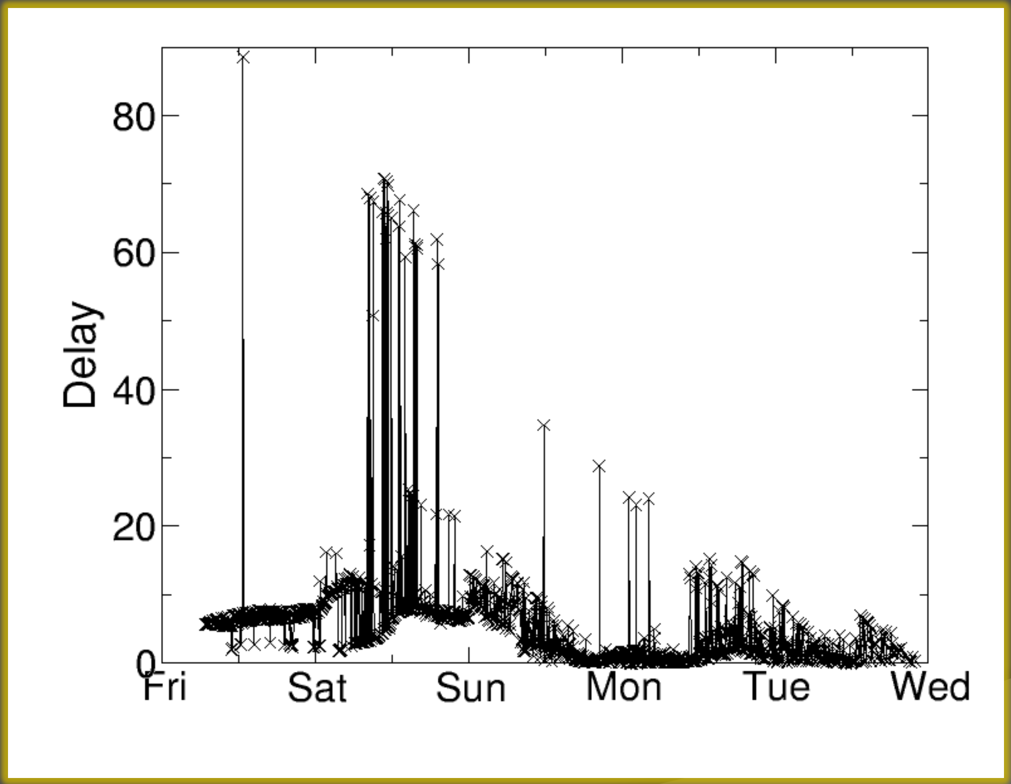
Aditya Akella (UW-Madison, USA)

Torrent of spam today

- More than 95% of email today is spam
  - Major ESPs receive more than 100 million spam messages per day
- Evolution of spamming
  - Present since the beginning (1978), it never stops growing
  - Spamming still has strong incentive as a business
  - Spammers own global-scale distributed spamming infrastructures (botnets)

How Is Receiving Huge Amount of Spam Harmful?

- Spamming is not just a nuisance. It could severely damage our information infrastructure.



Mail delivery delay (hours) at an enterprise mail system.

IP Reputation Services

- One technique to mitigate such spam traffic
- This service provides a score (reputation) for an IP address
- The most light-weight solution that precedes other anti-spam solutions.
- Based on reports from TTP and measurement (e.g., spam traps)
- Major spam appliance companies operate their own IP reputation services
  - Ironport, Symantec, etc.
  - are black boxes to users

Questions:

- What fraction of email can be correctly classified with IP reputation services?
  - Especially white lists since they previously have often been overlooked

- How we can create localized IP reputation services? Are they effective?

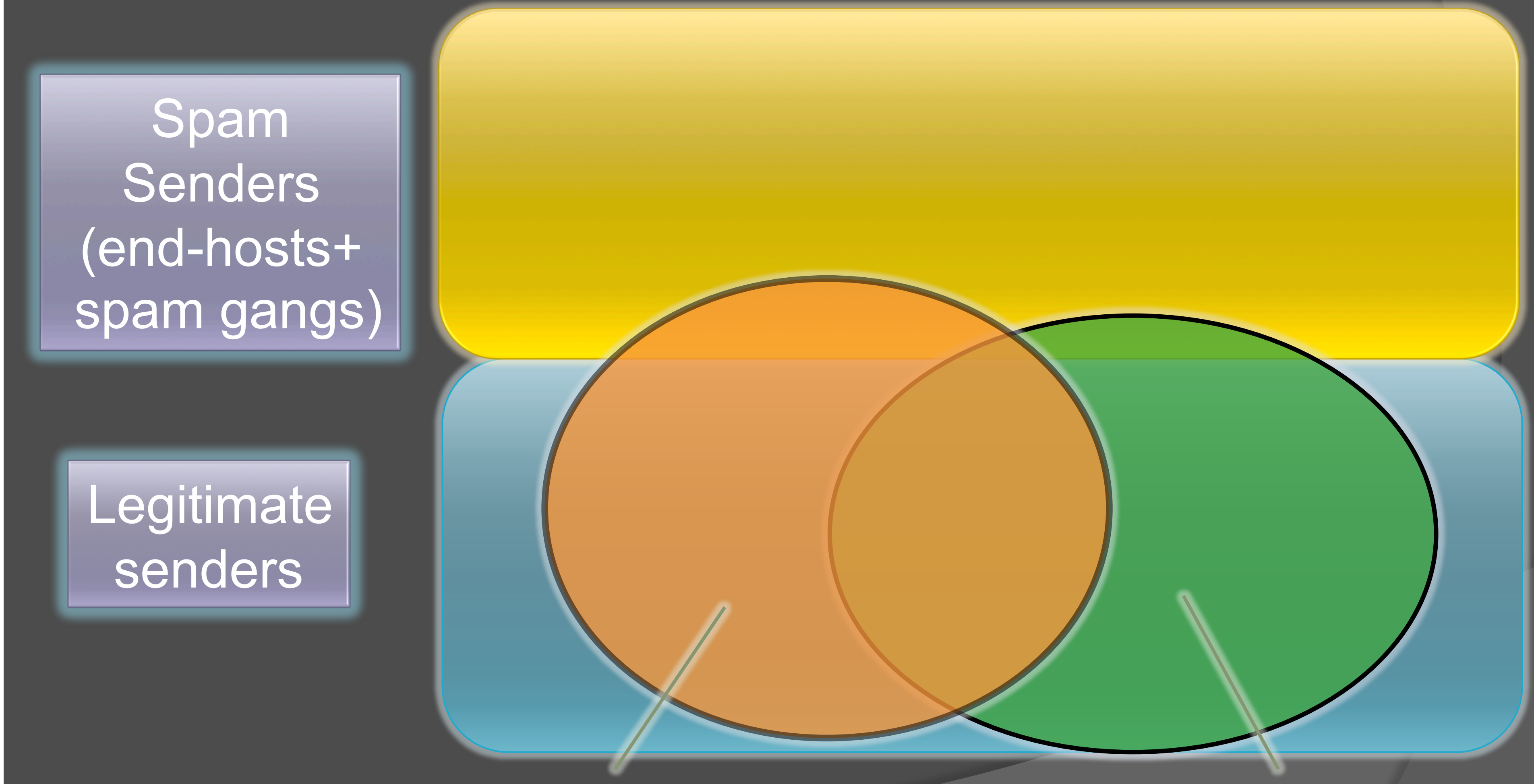
Our Contributions:

- Classify email senders into three primary categories and study the effectiveness of IP reputation services for each category
- Present methodologies to build custom local IP reputation lists
- Study other sources of email senders (open proxy, hijacked prefix)
- Study the characteristics of spamming for each category of senders

Three Categories of Email Senders

- Legitimate servers
  - MTA for legitimate ISP, ESP, Companies, Universities, …
- End-hosts
  - Compromised end-hosts (botnets)
- Spam gangs
  - Bullet-proof hosting servers
  - E.g., Russian Business Network

Performance Evaluation of IP Reputation Lists



Spam Senders (end-hosts+ spam gangs)

Legitimate senders

Existing global white lists

Custom local white lists

Review of DNS SPF

- SPF: Sender Policy Framework
- A simple authentication mechanism that associates domain and IP addresses
  - E.g., ieee.org → v=spf1 ip4:72.236.151.122/32 …
- Some spammers also use SPF to pass the simple authentication checks
  - We can use this to cluster their domains and addresses

Building Custom IP Reputation Lists -- Legitimate Servers --

- WL1: Legit-Popular (web)
  - Compile a list of legitimate domains manually and extract associated IP addresses
- WL2: SPF-good (history-based)
  - Collect domains with good scores and extract associated IP addresses
  - Sufficient history required

Building Custom IP Reputation Lists -- End-hosts --

- BL1: Hostname (Naming heuristics)
  - Compile heuristics for hostnames, e.g., ppp222.foo.com, dyn34-13-7-12.bar.com
  - Check the RDNS of all the IP addresses
- BL2: Srizbi (Malware heuristics)
  - Check the TCP header of a sender
    - If the pattern matches to special case, it is likely a bot.

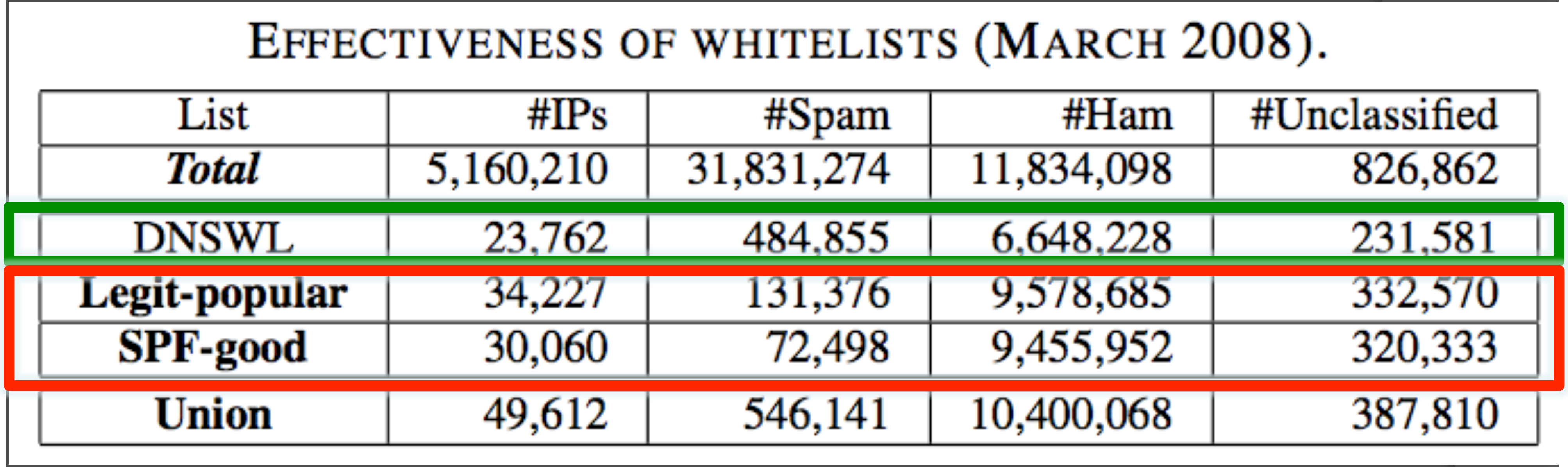
Building Custom IP Reputation Lists -- Spam Gangs --

- BL3: Bad Blocks (history-based)
  - Extract blocks (clusters) of IP addresses with bad history
  - Clustering with BGP prefix and some heuristics (/29-based aggregation)
- BL4: SPF-bad (history-based)
  - Same as SPF-good except for bad domains and their associated IP addresses

Data Sets

- SMTP logs
  - Timestamp, sender IP, sender domain, score
  - Collected at University of Wisconsin-Madison
- Tcpdump
  - Used for compiling custom blacklists (BL2)

Performance of IP Reputation (1)

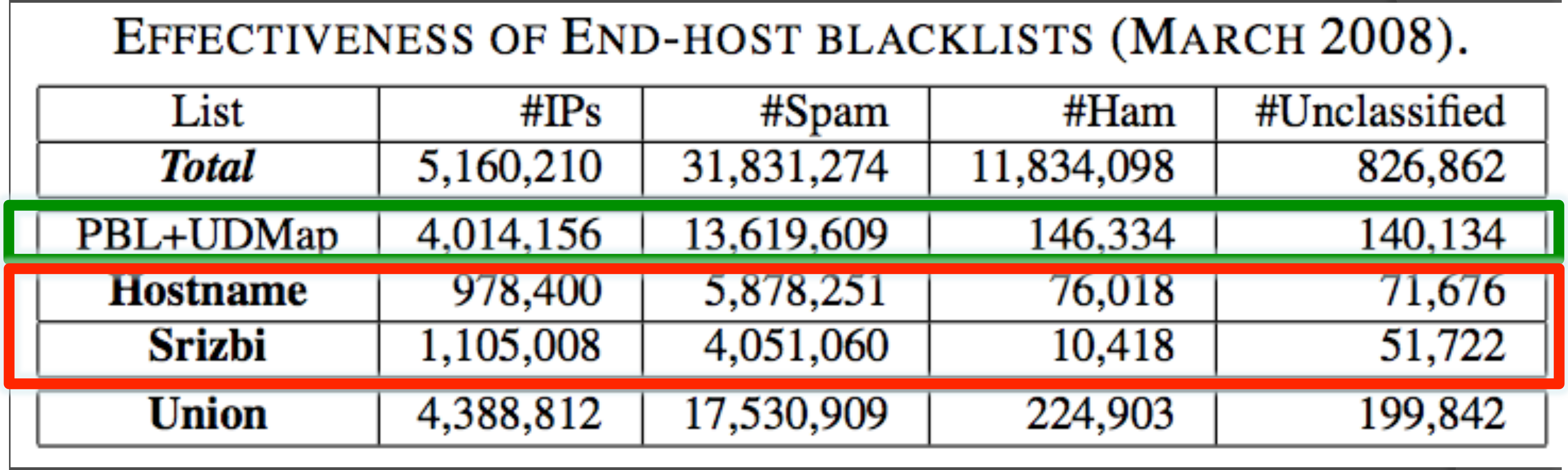


EFFECTIVENESS OF WHITELISTS (MARCH 2008).

List	#IPs	#Spam	#Ham	#Unclassified
Total	5,160,210	31,831,274	11,834,098	826,862
DNSWL	23,762	484,855	6,648,228	231,581
Legit-popular	34,227	131,376	9,578,685	332,570
SPF-good	30,060	72,498	9,455,952	320,333
Union	49,612	546,141	10,400,068	387,810

- Custom reputation lists cover more ham and less spam
- In total, reputation lists cover roughly 90% of ham

Performance of IP Reputation (2)

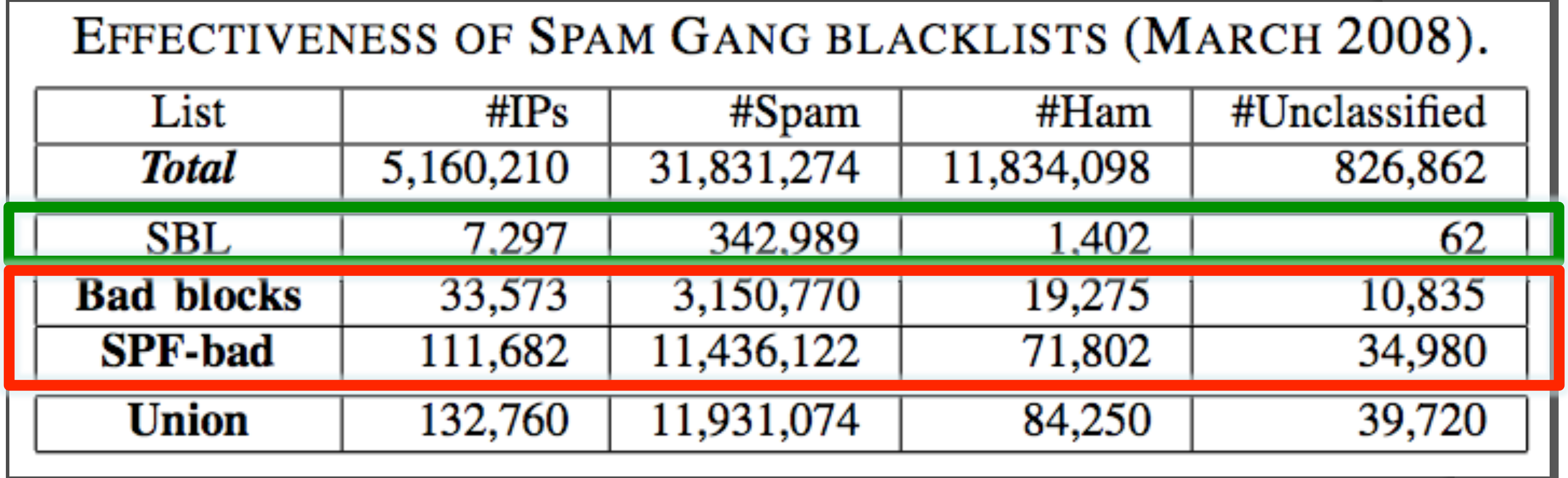


EFFECTIVENESS OF END-HOST BLACKLISTS (MARCH 2008).

List	#IPs	#Spam	#Ham	#Unclassified
Total	5,160,210	31,831,274	11,834,098	826,862
PBL+UDMap	4,014,156	13,619,609	146,334	140,134
Hostname	978,400	5,878,251	76,018	71,676
Srizbi	1,105,008	4,051,060	10,418	51,722
Union	4,388,812	17,530,909	224,903	199,842

- Custom lists complement the coverage by 22%
- In total, the reputation lists cover more than 54% of spam

Performance of IP Reputation (3)

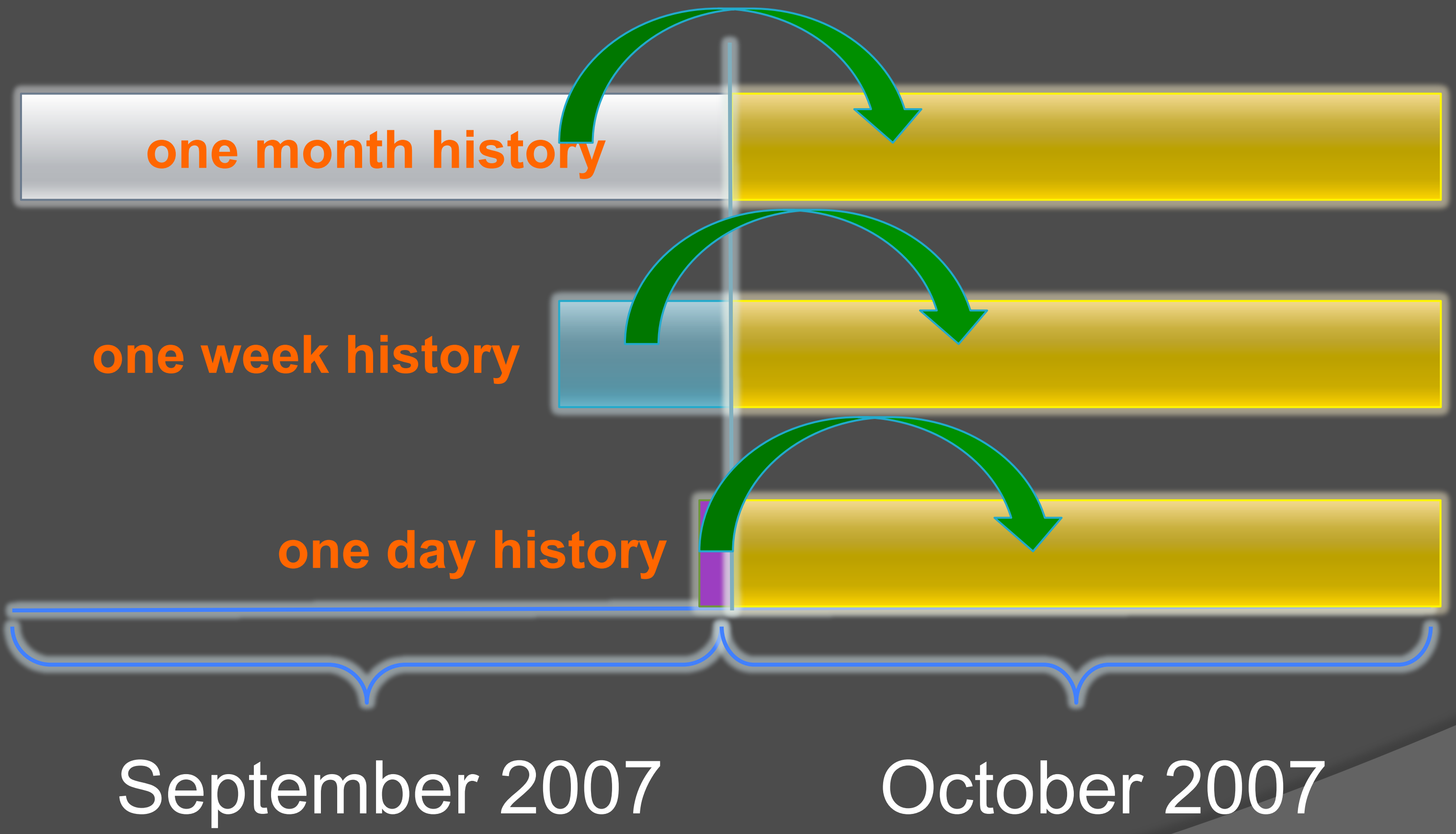


EFFECTIVENESS OF SPAM GANG BLACKLISTS (MARCH 2008).

List	#IPs	#Spam	#Ham	#Unclassified
Total	5,160,210	31,831,274	11,834,098	826,862
SBL	7,297	342,989	1,402	62
Bad blocks	33,573	3,150,770	19,275	10,835
SPF-bad	111,682	11,436,122	71,802	34,980
Union	132,760	11,931,074	84,250	39,720

- Custom lists cover much more spam with low fraction of false positives
- In total, the reputation lists cover more than 38% of spam

Effectiveness of history-based reputation over time



one month history

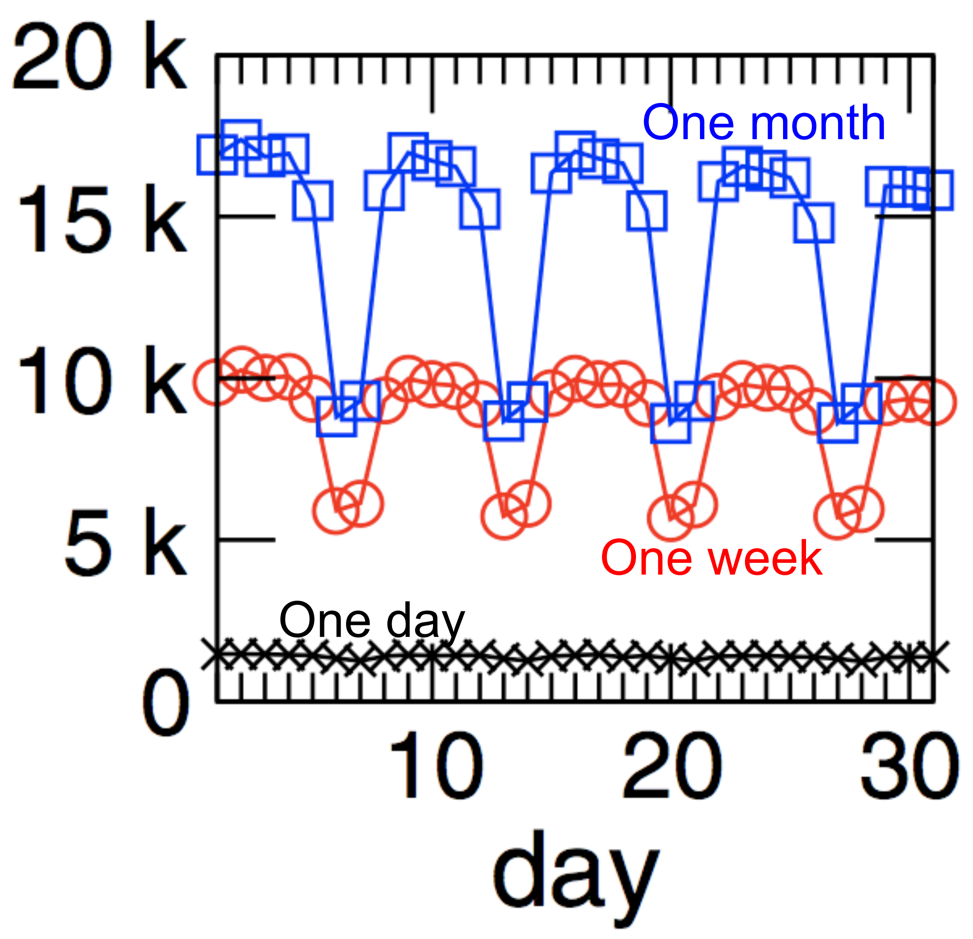
one week history

one day history

September 2007

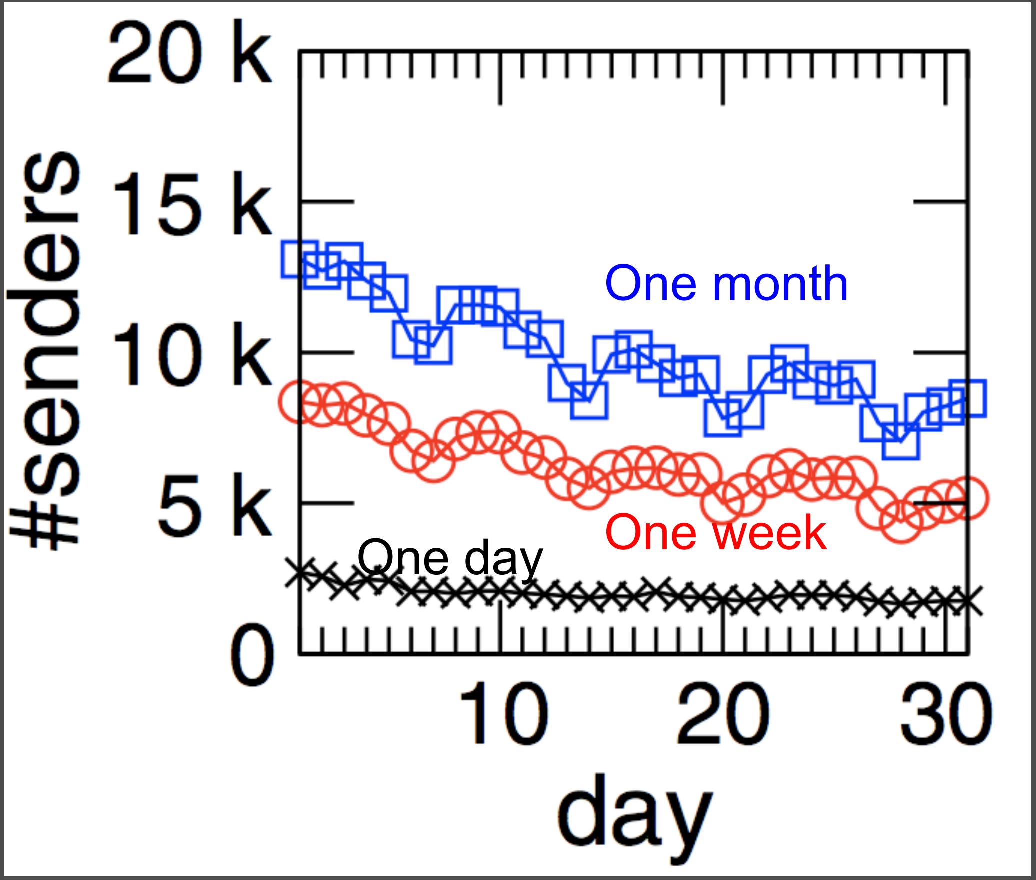
October 2007

Coverage of SPF-good over time



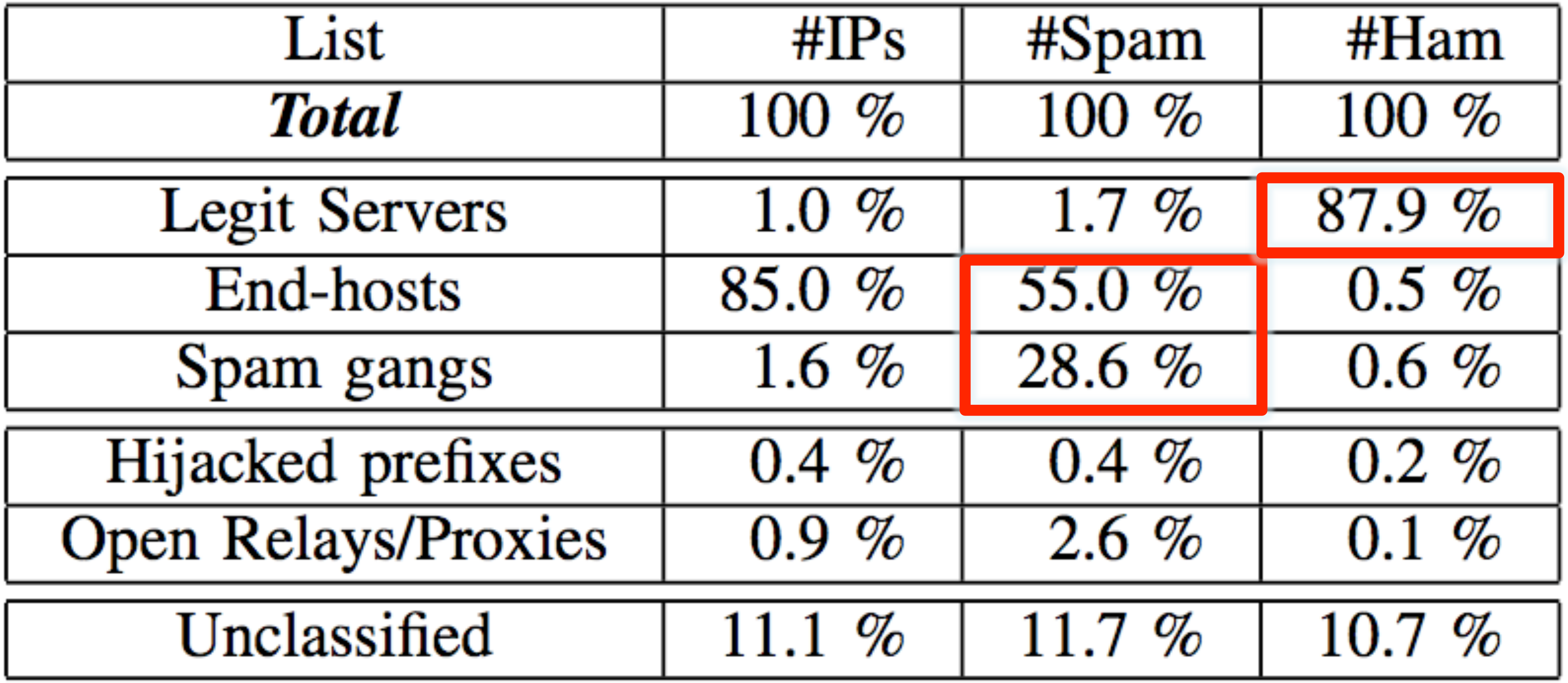
- Constant over time
  - Good ones are stable
- Cyclic patterns
  - Human activity
- Longer learning covers more senders
  - One week is comparable to one month

Coverage of SPF-Bad over time



- Degraded over time
  - Bad ones are not stable
- Weaker cyclic patterns
  - Machine activity
- Longer learning covers more senders
  - One week is comparable to one month

Contribution of each category



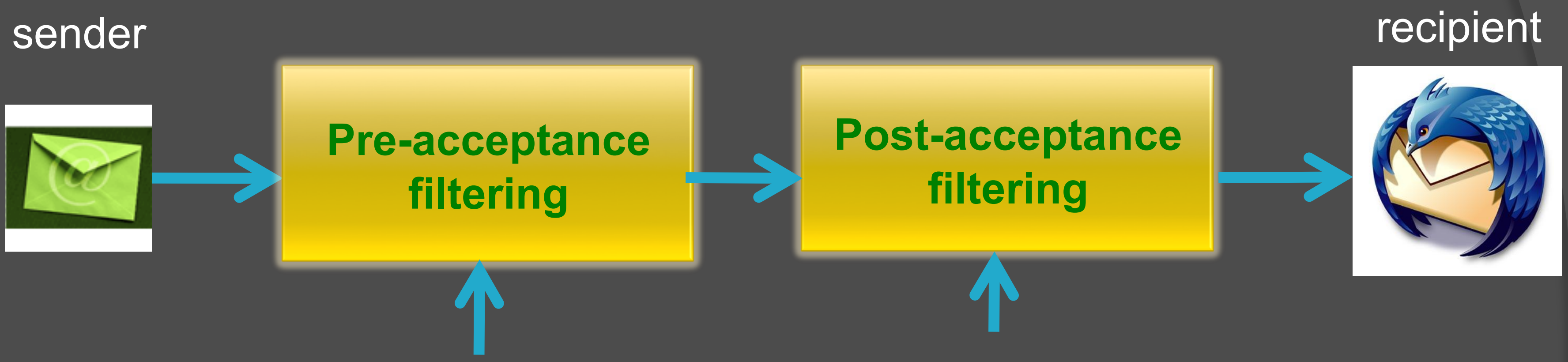
List	#IPs	#Spam	#Ham
Total	100 %	100 %	100 %
Legit Servers	1.0 %	1.7 %	87.9 %
End-hosts	85.0 %	55.0 %	0.5 %
Spam gangs	1.6 %	28.6 %	0.6 %
Hijacked prefixes	0.4 %	0.4 %	0.2 %
Open Relays/Proxies	0.9 %	2.6 %	0.1 %
Unclassified	11.1 %	11.7 %	10.7 %

Summary and Future Work

- Empirically showed up to 90% of spam and ham can be classified with IP reputation services if compiled correctly.
- Local reputation lists can complement global IP reputation services.
- Good IPs are stable over time. Reputation lists for spam gangs need frequent updates.

- Aggregating IP reputation lists using machine learning techniques a viable direction for improving lists further

Existing anti-spam solutions



- IP reputation (DNSBL)
- Greylisting
- Greet pause

- Text mining
- Binary pattern matching
- OCR

Light-weight
Limited information

⟷

Heavy-weight
Detailed information