ここまで捨てられる！スパムメール対策術
～技術の現状と展望～

JANOG23

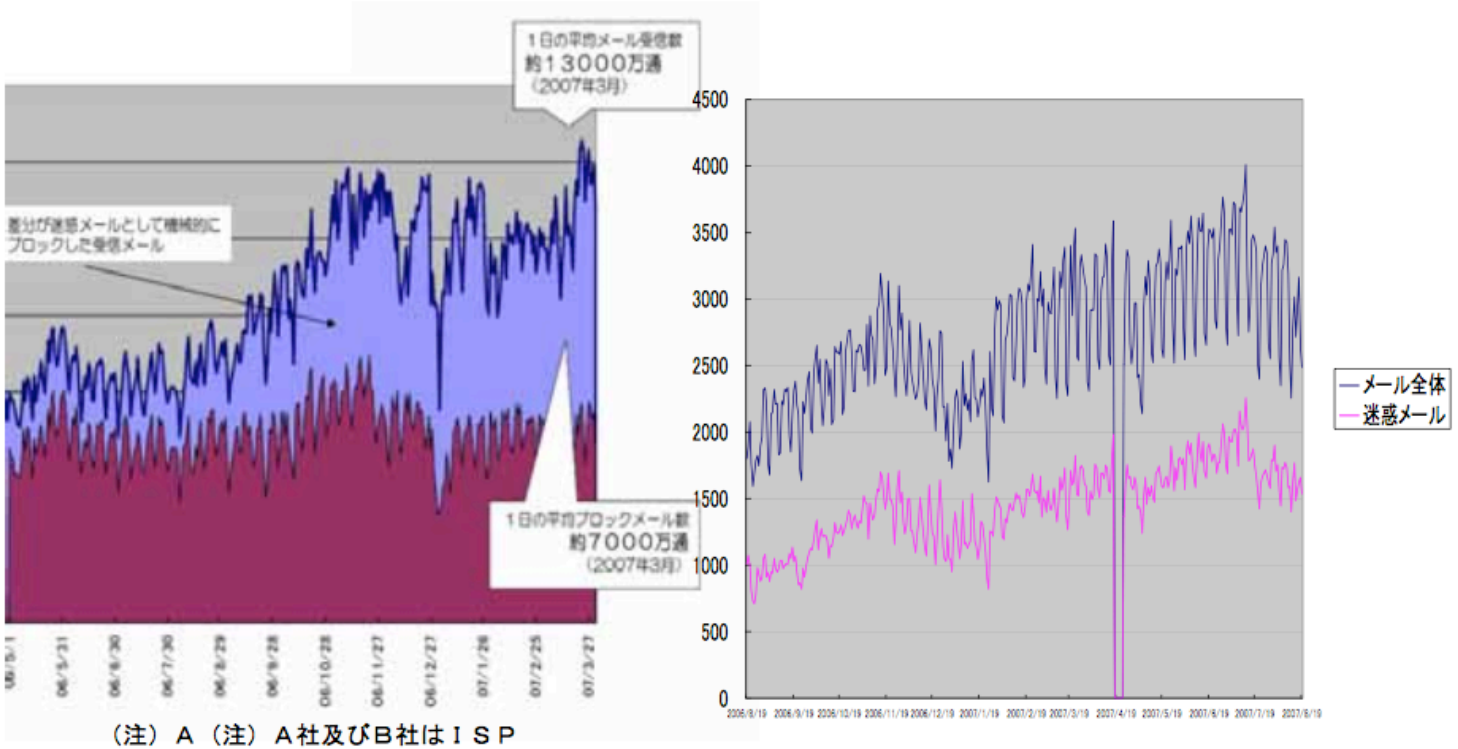
2009/1/22

NTTサービスインテグレーション基盤研究所

森　達哉

止まらないスパム

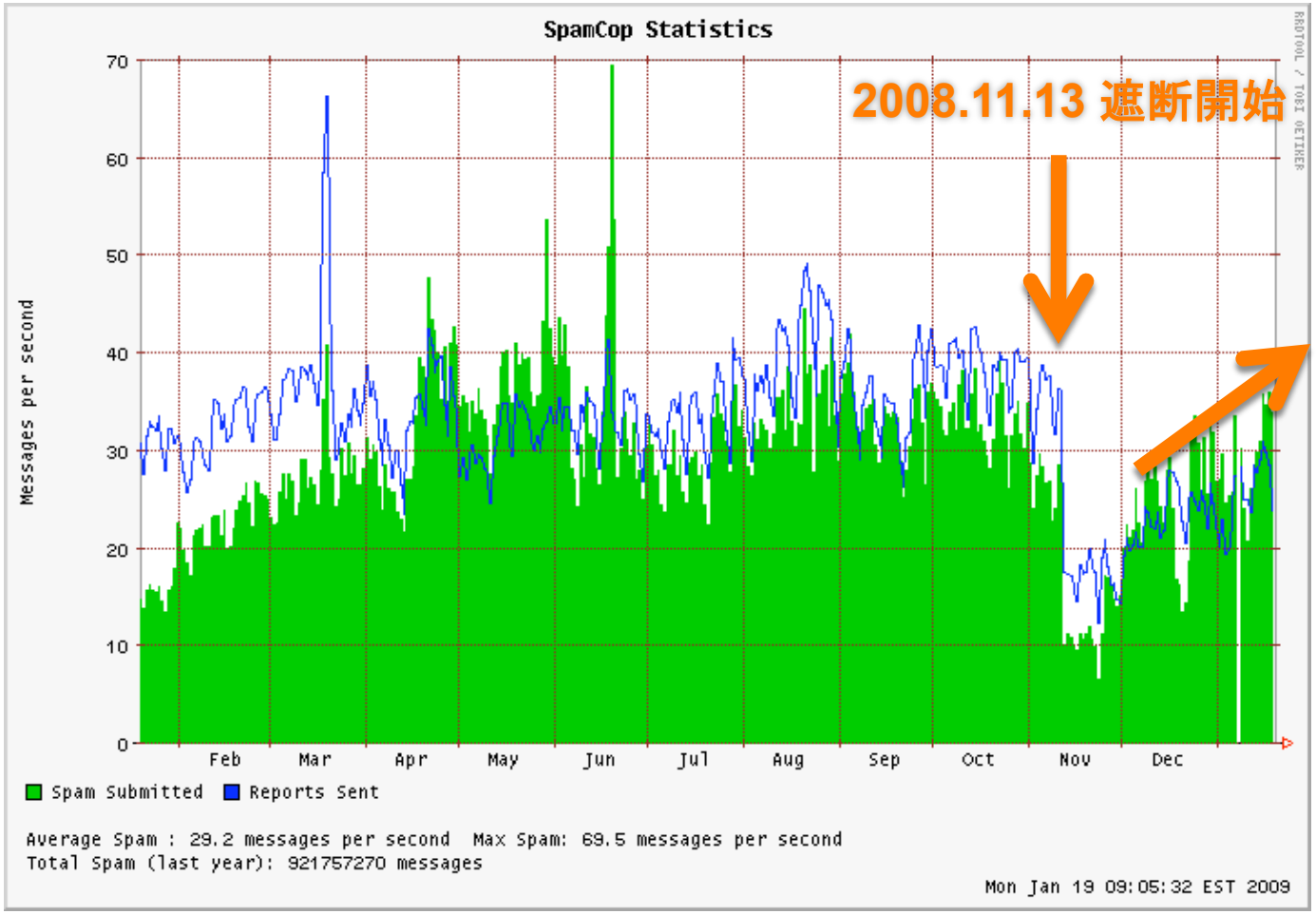
- Since 1978
- 特に2006-2007年より世界的な急増傾向にあり。



総務省「迷惑メールへの対応の在り方に関する研究会」中間とりまとめ報告より引用

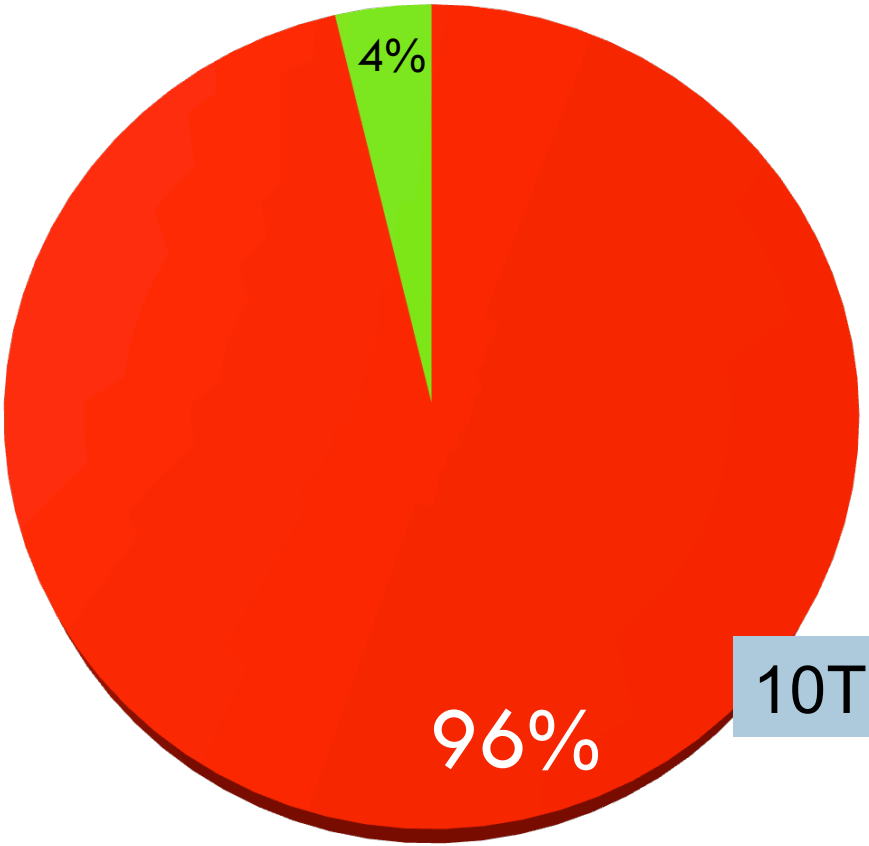
Botnet C&Cサーバホスティング会社(McColo)の上位ISPによる遮断とその後の経過

- 2週間ほどは効果があったがその後徐々に戻している



あるメールサーバの1ヶ月のメール受信状況

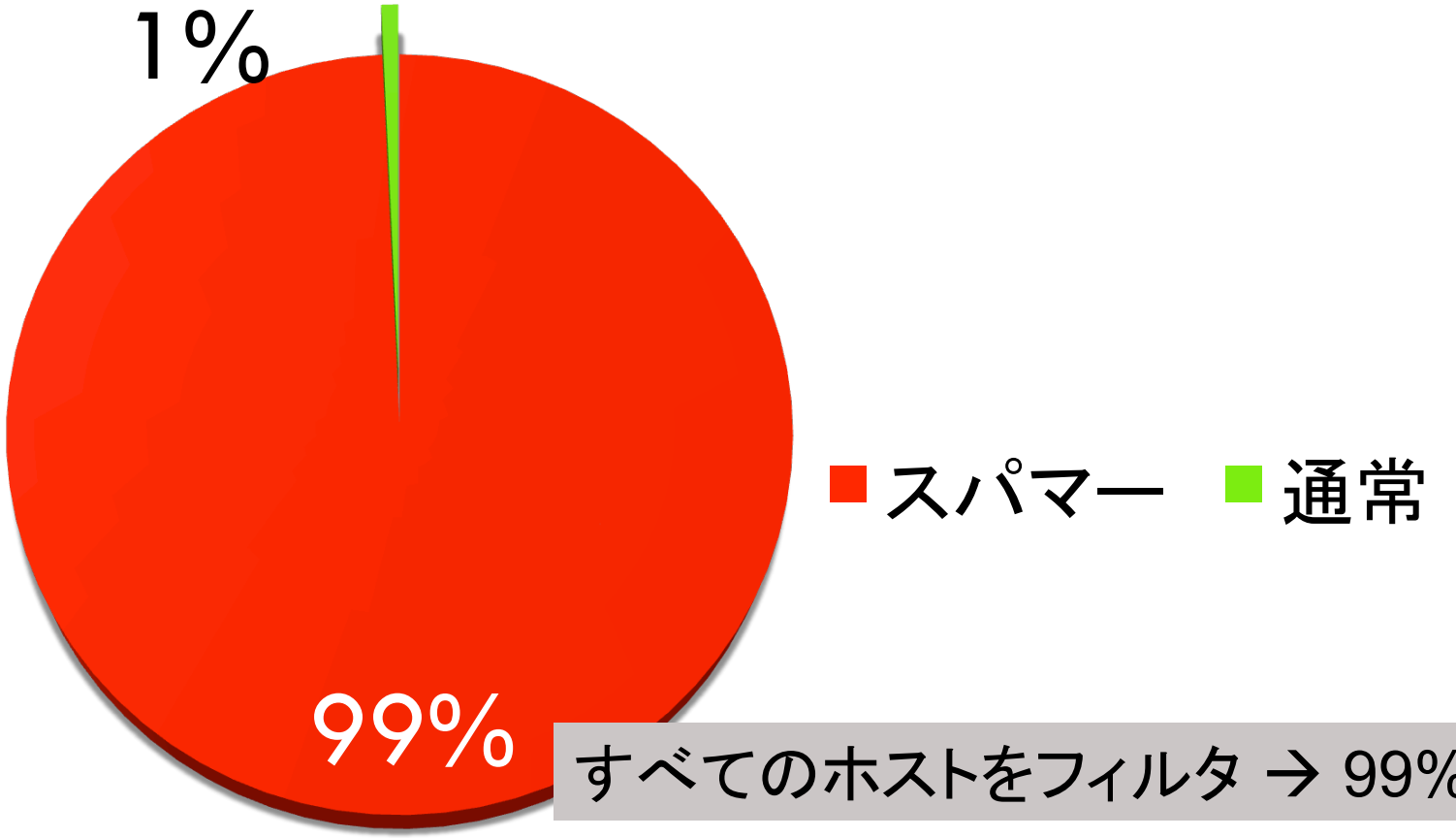
受信メッセージ: 約1800万通の内訳



10TB のストレージ → 9.6TB が無駄!

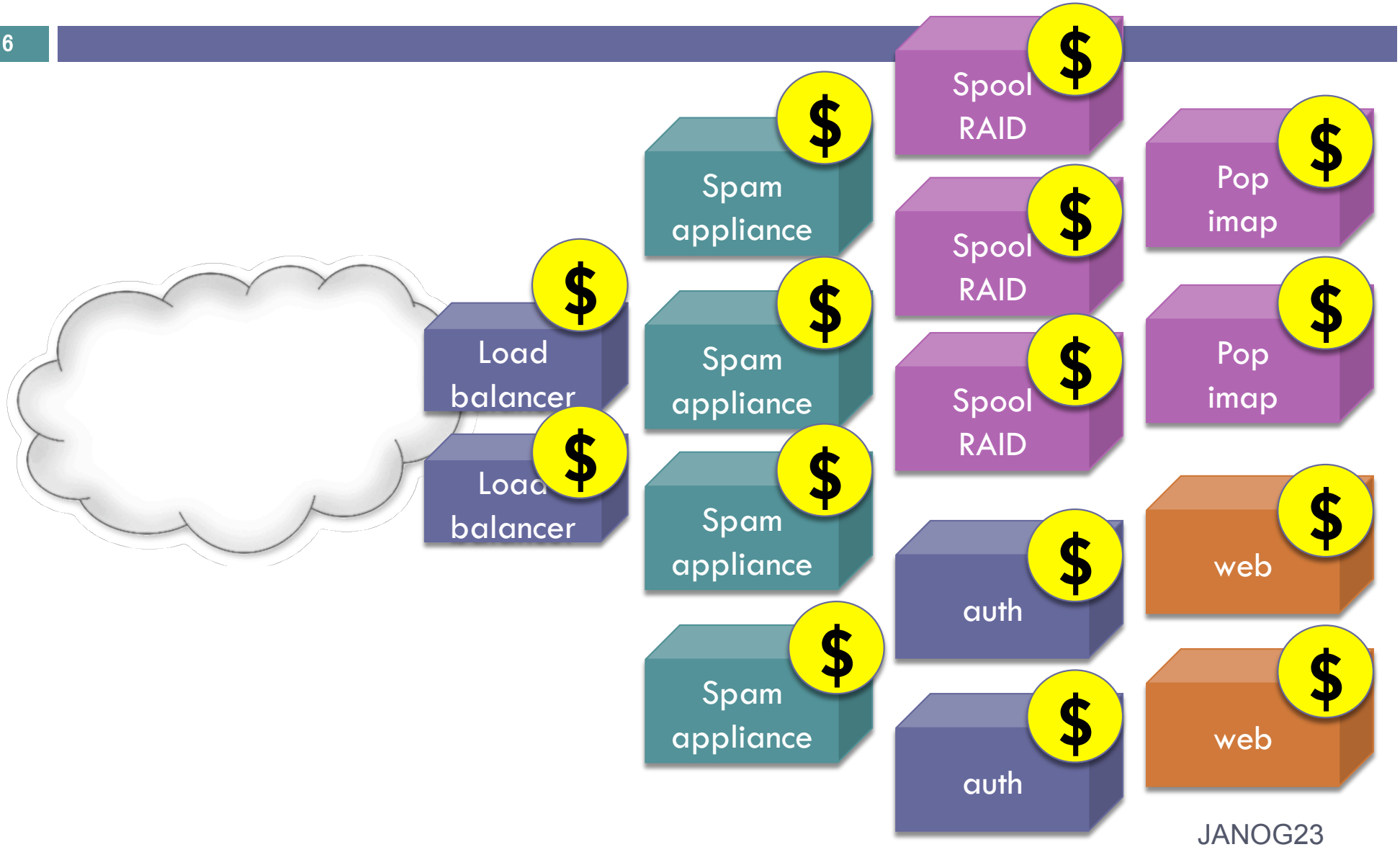
あるメールサーバの1ヶ月のメール受信状況

メール送信ホスト: 約200万ホストの内訳



すべてのホストをフィルタ → 99%正解!!

中～大規模なメール配送システムにかかるコスト



今メールサービスオペレータが困っていること

- 大量のスパムメール処理に要する配送システムのオーバーヘッドの増大
  - 特にコンテンツフィルタリングの処理が重い
    - ウィルススキャン・情報漏洩監査
    - スパムフィルタリング
  - 負荷増大により、サービス断してしまうケースが散見される(ISP, 企業, 大学等)
  - 大量に消費されるディスクスペース
    - ほとんどは無駄なビット…
- 一方で通信の秘密による規制もあり

何をすれば良いか

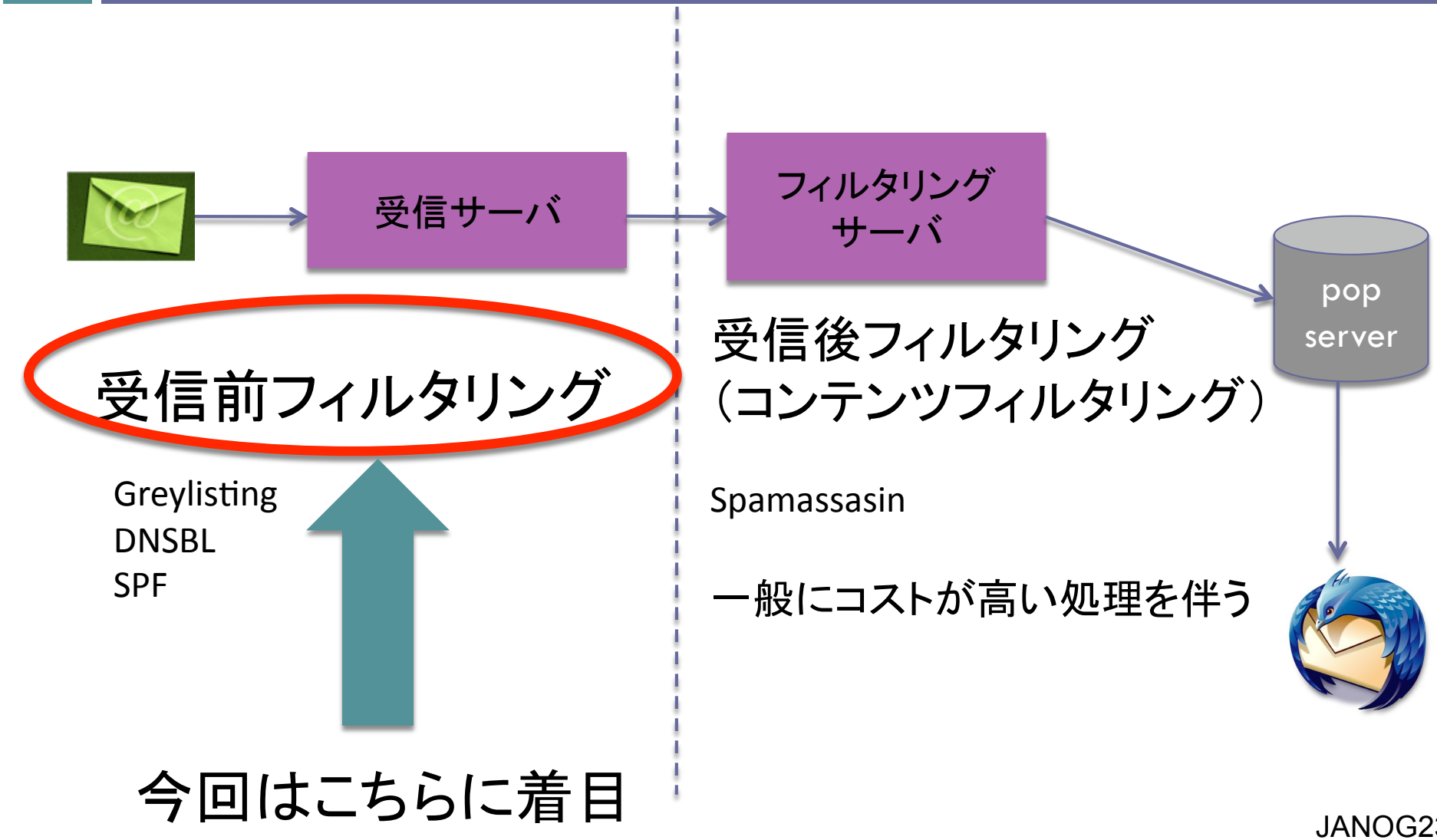
- まず現状の把握
  - 統計として実態を明らかにする
  - 既存の対策技術の有効性を評価

今回の
発表内容

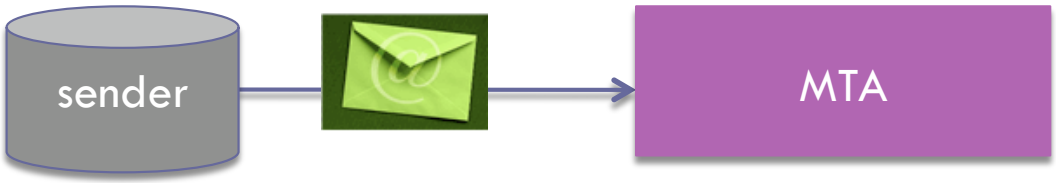
- 今後の対策技術に求められるものは何か？
  - 議論しましょう！

- 将来のために継続的な計測・分析が重要
  - 対処療法はあくまでも一時的なもの
  - トレンドは変わる

代表的な迷惑メール対策技術



Greylisting



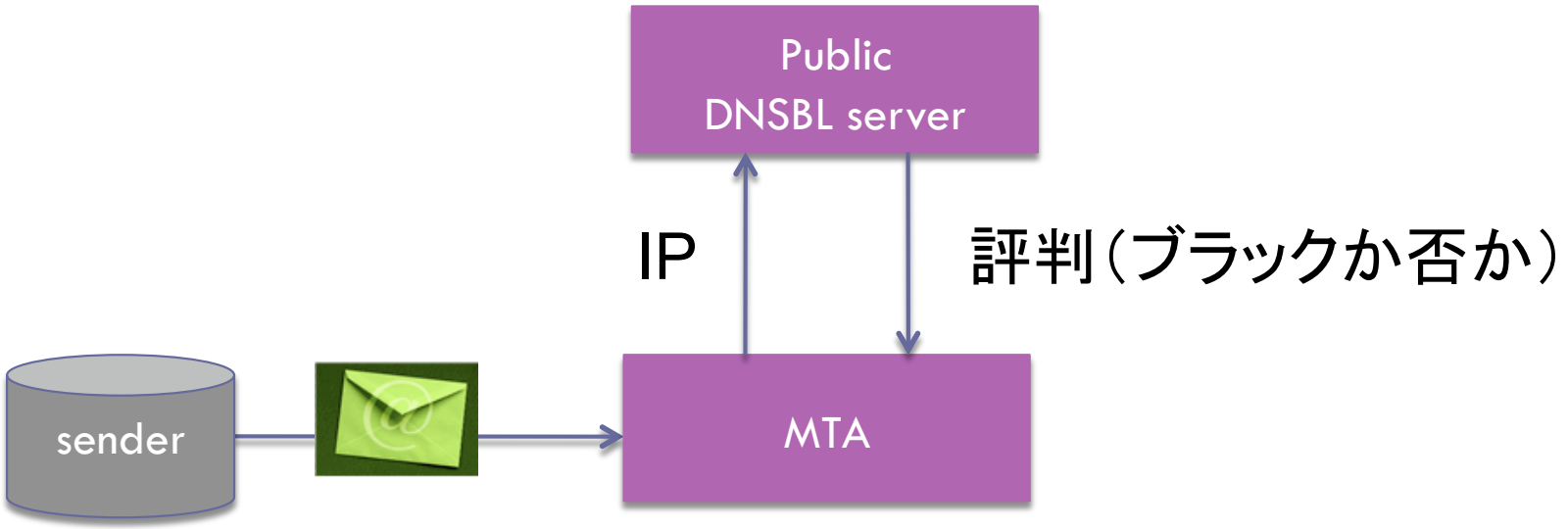
{ip, from, to}

初見の組み合わせ→ 拒否
二度目以降→受信

botnet のようなスパム送信ホストは再送しないというのが前提
IPアドレスやメッセージの中身に依存しないホストの挙動を利用した手法

DNSBL (DNS Black list)

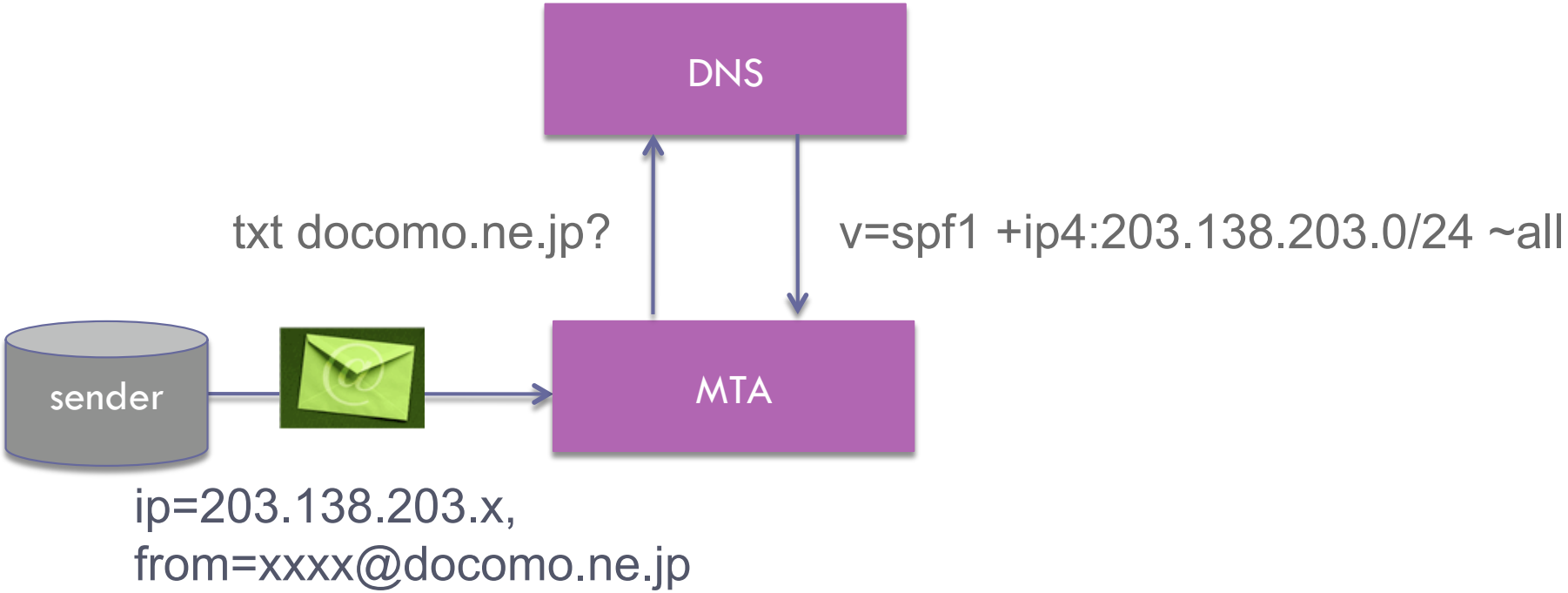
IPアドレスの評判(black list)をDNS インタフェースで提供




```
% nslookup 90.57.60.129.sbl.spamhaus.org
** server can't find 90.57.60.129.sbl.spamhaus.org: NXDOMAIN

% nslookup 93.12.186.222.sbl.spamhaus.org
Non-authoritative answer:
Name:   93.12.186.222.sbl.spamhaus.org
Address: 127.0.0.2
```

SPF (Sender Policy Framework)

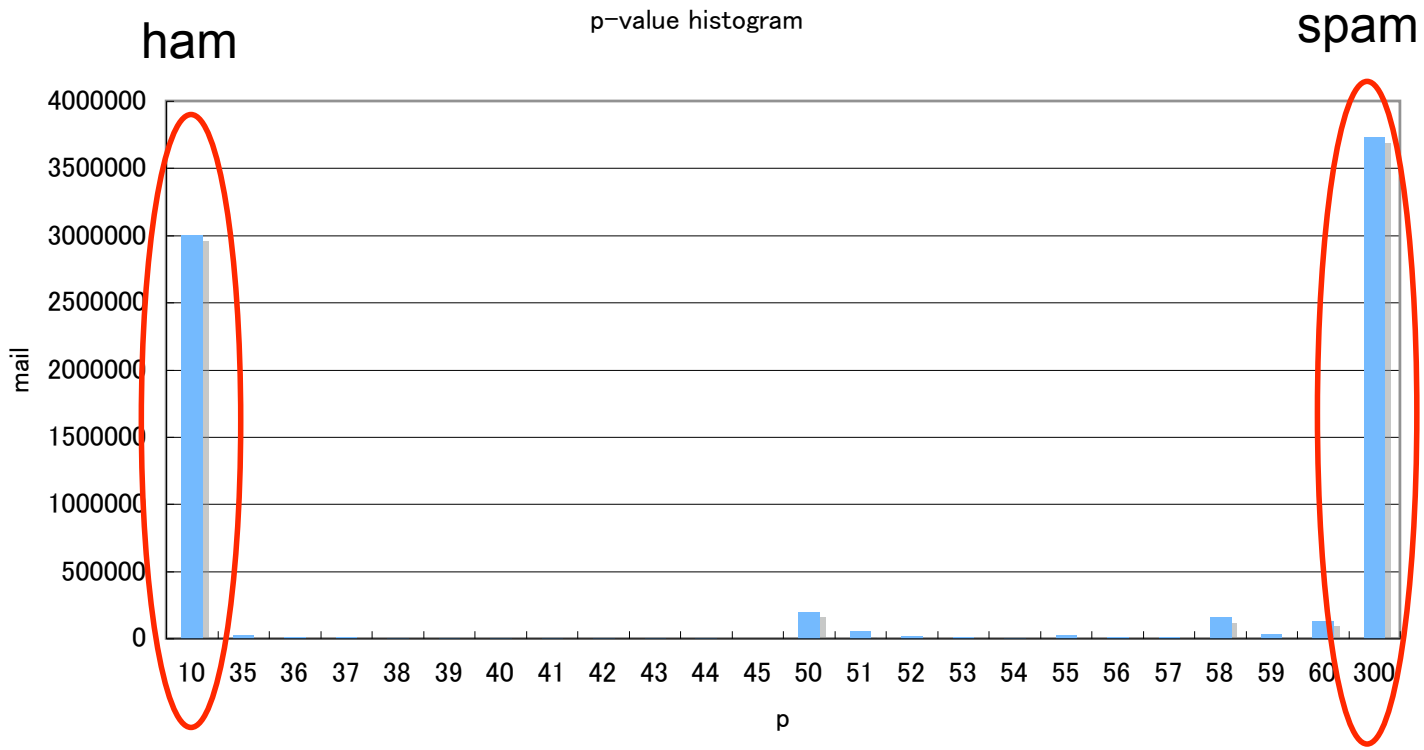


送信してきたホストのIP が該当ドメインのSPFに記載されている
アドレスにマッチするか否かを調べる。
※マッチしない → 詐称というわけでは必ずしもない（移動先での
メール送信など）。

分析したデータ

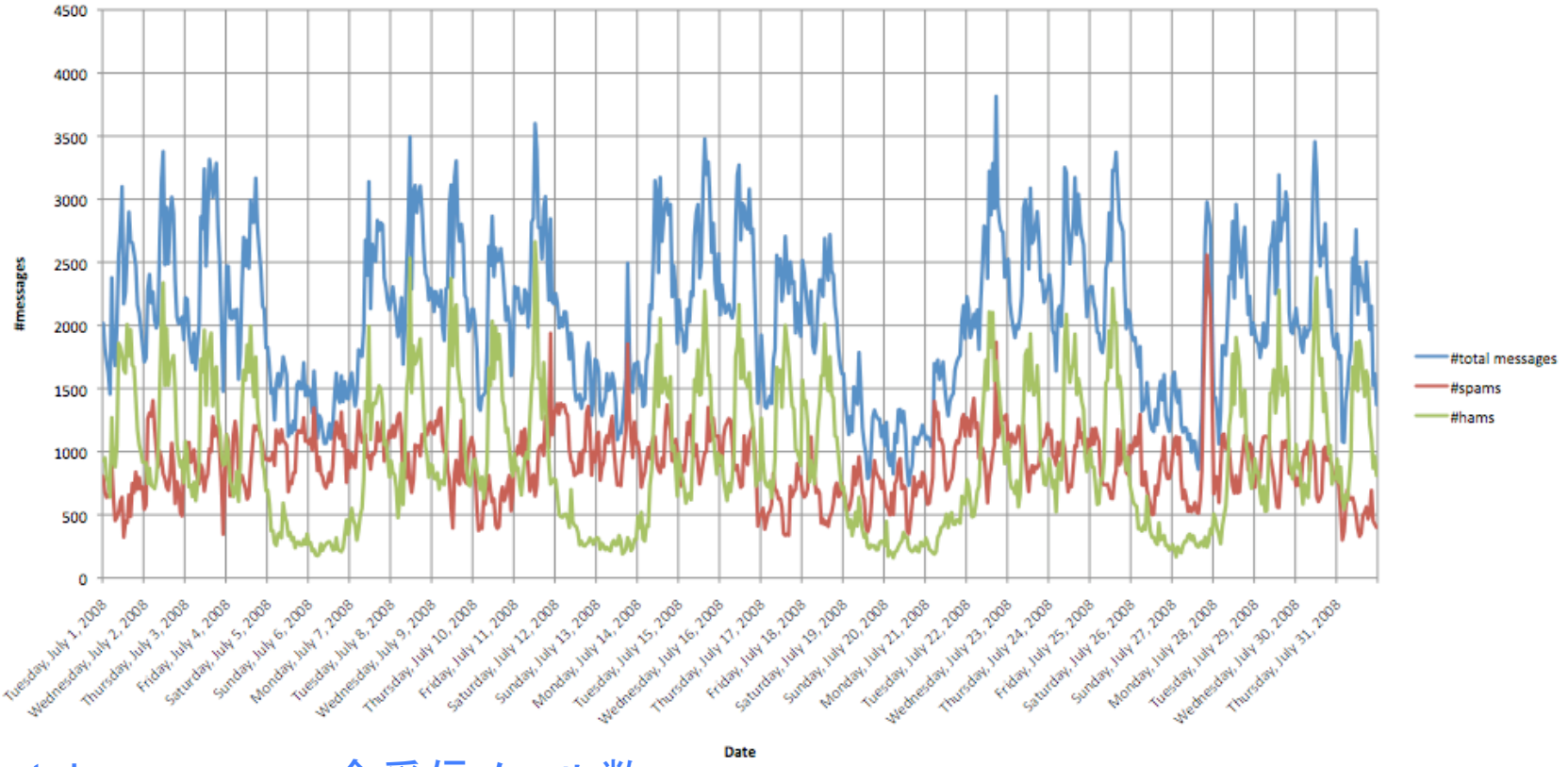
- 対象
  - 企業網メール
  - 計測期間: 2008年4月～2008年7月
- 運用
  - greylisting を運用中
  - 商用スパムアプライアンスにより、すべての受信メッセージにスコアがつく

メッセージについたスコアの分布



スコアの真偽の信頼性は高いことを検証済（マニュアル調査）
以下ではスコアが正しいものとして分析

受信メッセージの変動パターン



#total messages: 全受信メール数

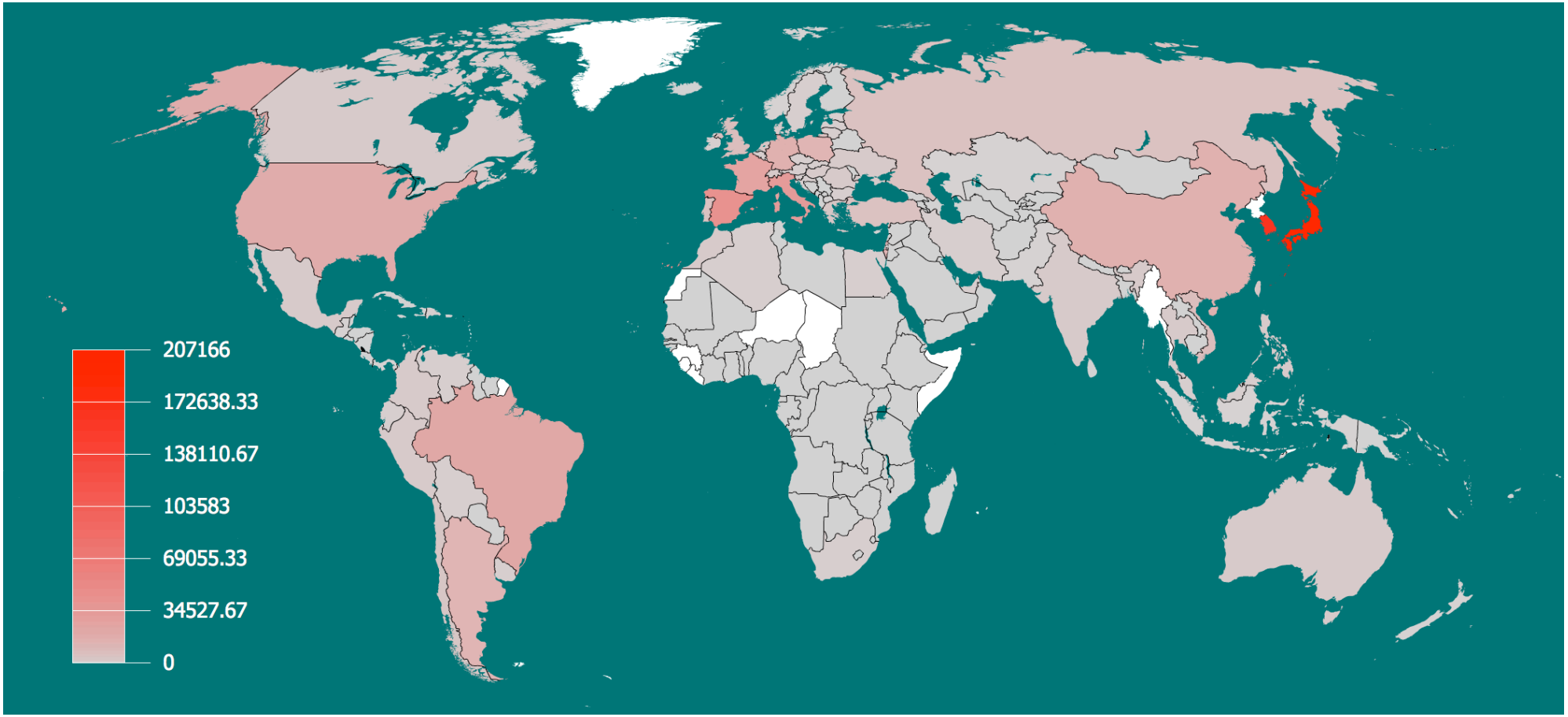
#spams: スパムメール数

#hams: ハムメール数

ハムの配信数は人間活動と相関のあると思われる日変動を示す。

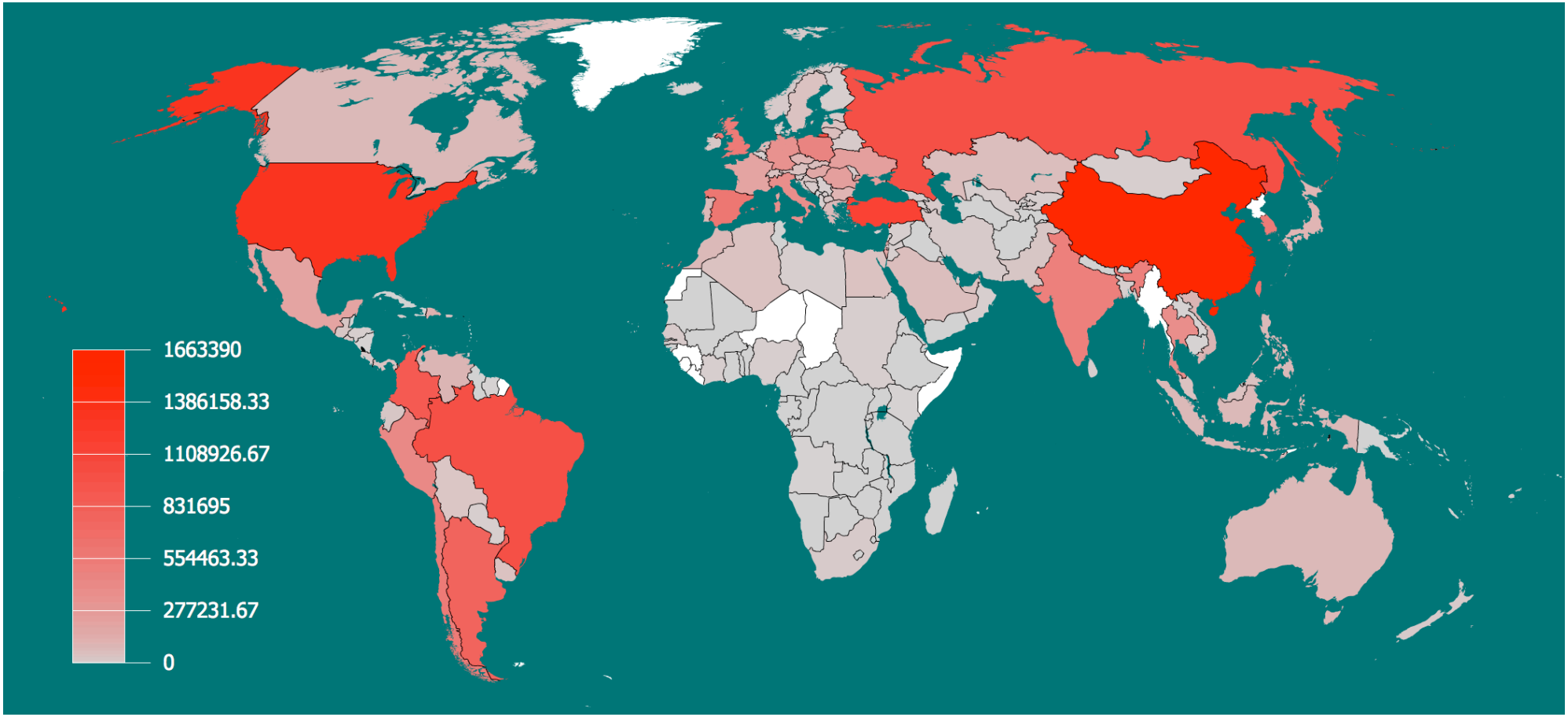
スパムの配信数もやや日変動傾向あり。タイムゾーンが7-8時間ほどずれている。

(受信)スパムの送信元



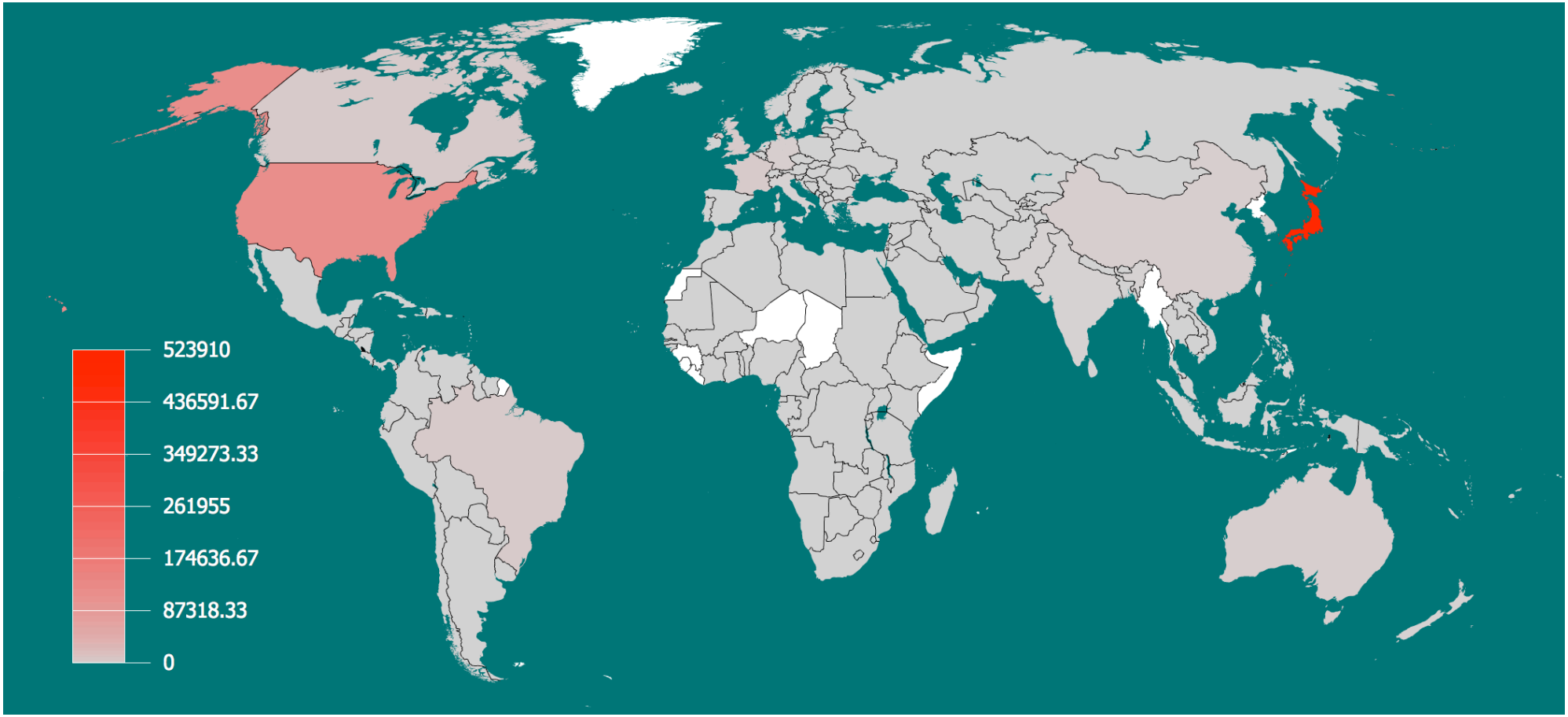
実際に配送されたスパムは国内・韓国発が多かった
※これらのスパムは greylisting で落ちていない

(非受信)スパムの送信元



Greylisting でフィルタされたメッセージ(非受信スパム)は BRICs 諸国を中心とした一部の国に集中

ハムの送信元

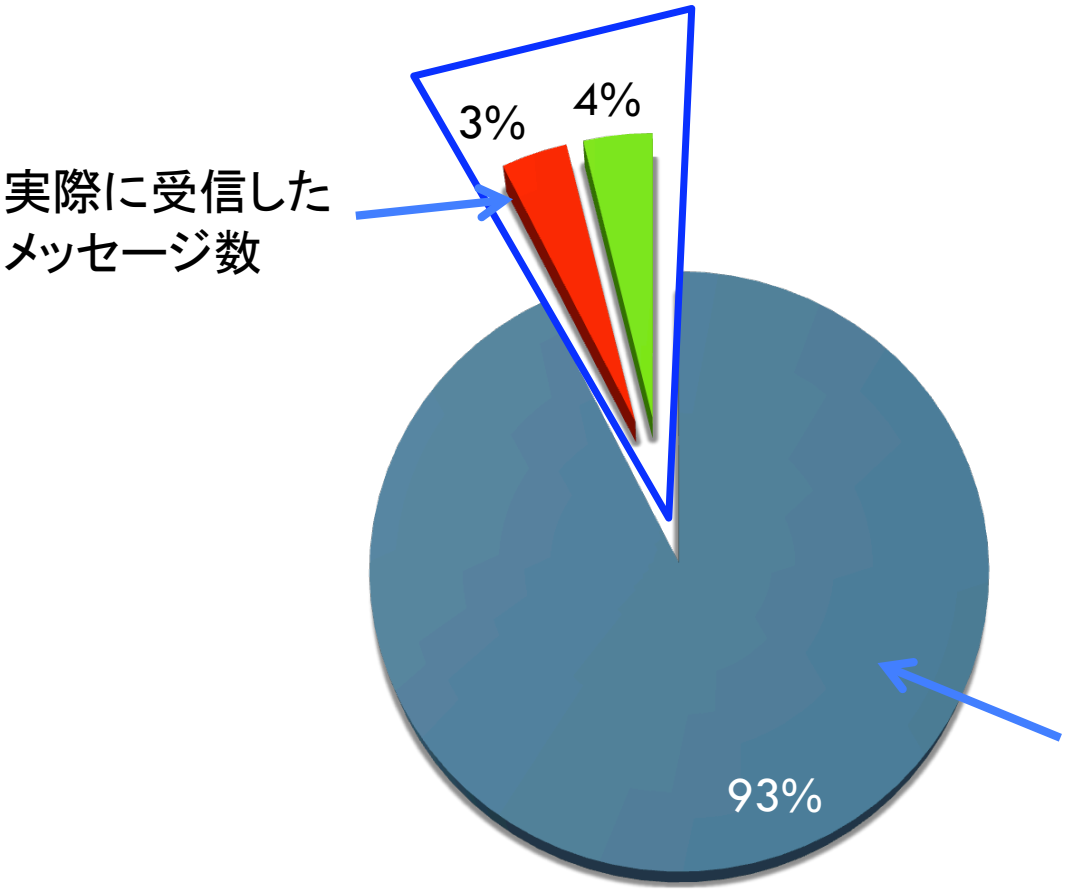


通常のメッセージの送信元は日米に集中

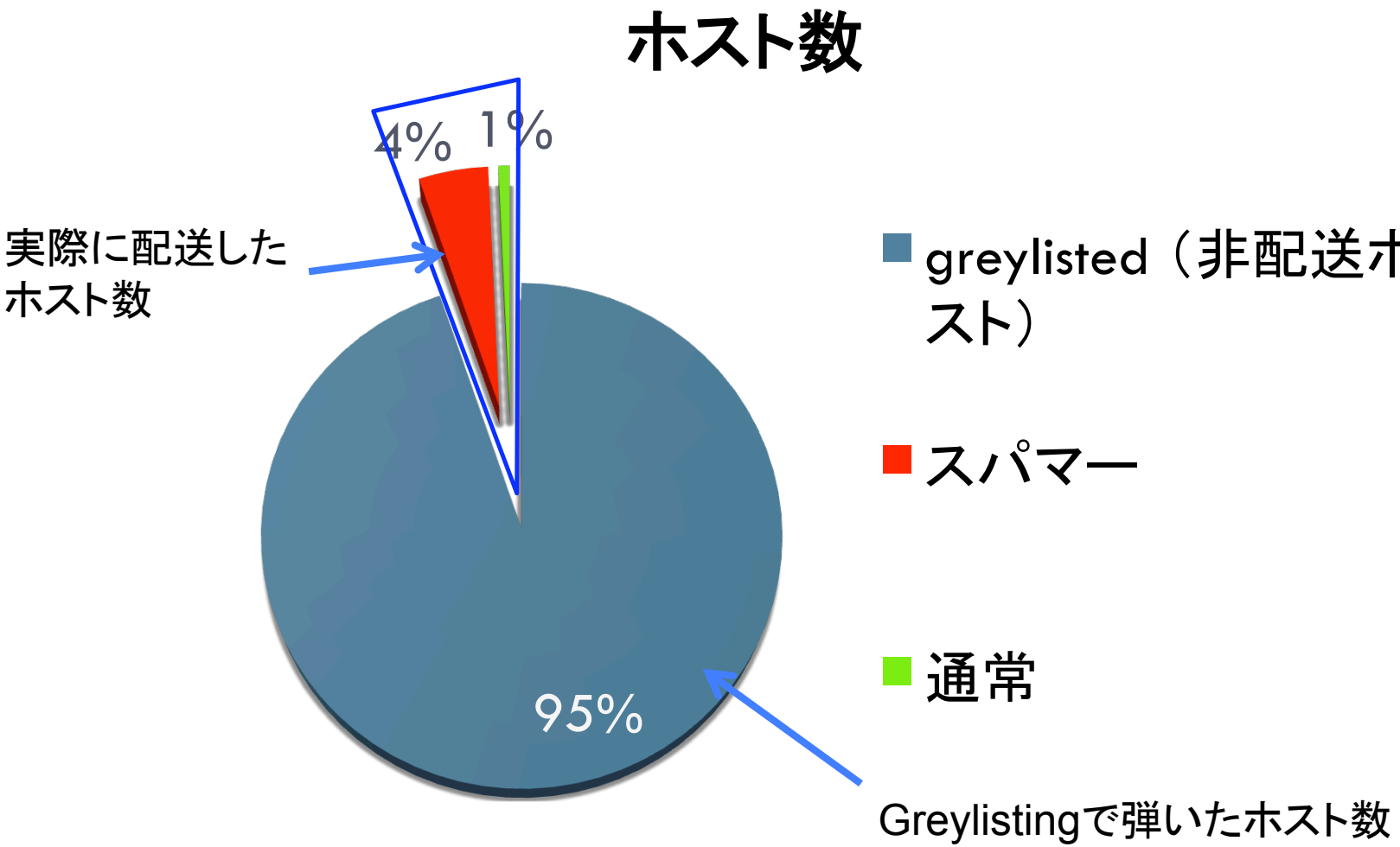
各対策技術の有効性

Greylistingの効果

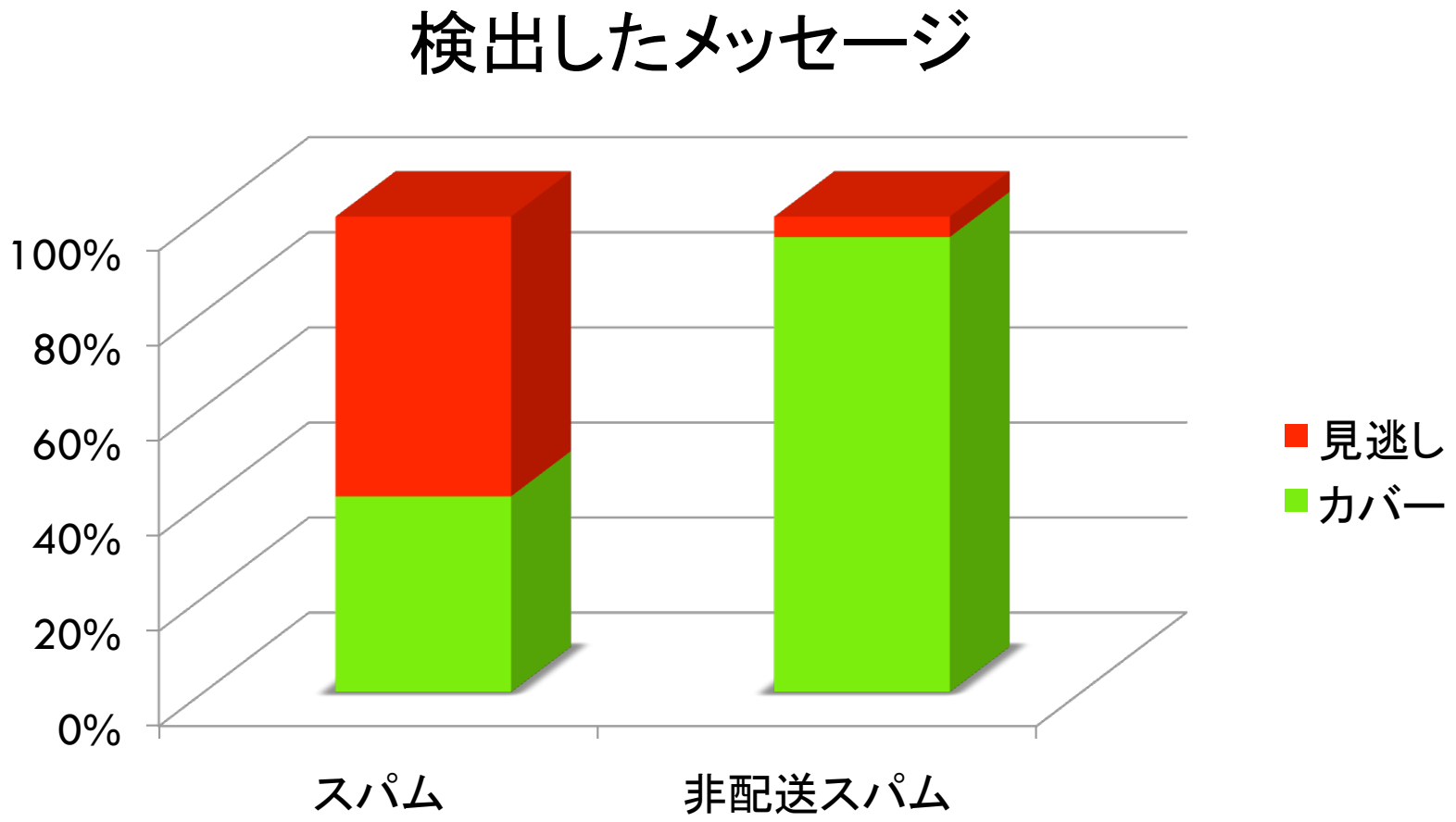
メッセージ・コネクション数



Greylisting の効果

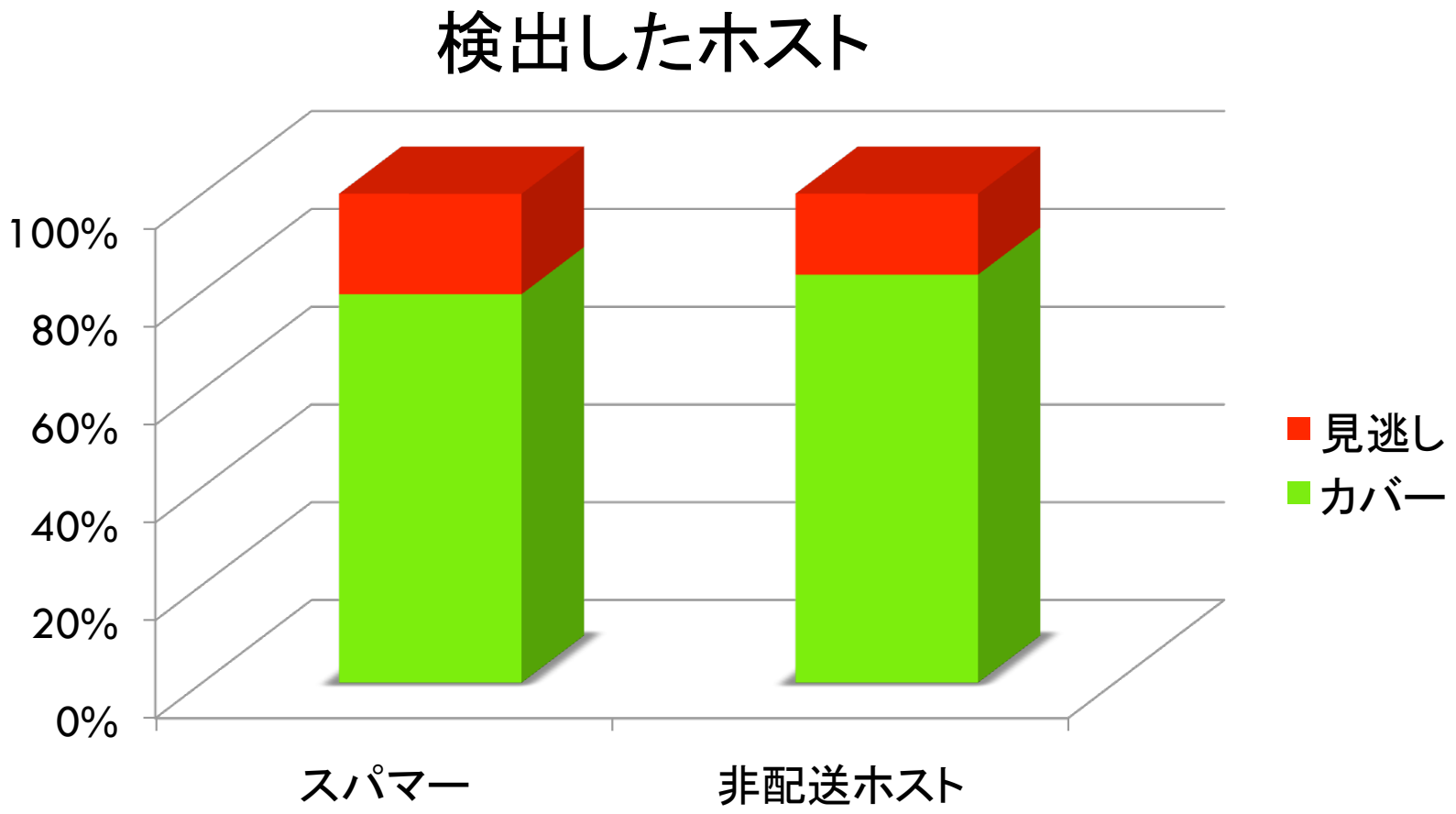


DNSBLの効果（カバー率）



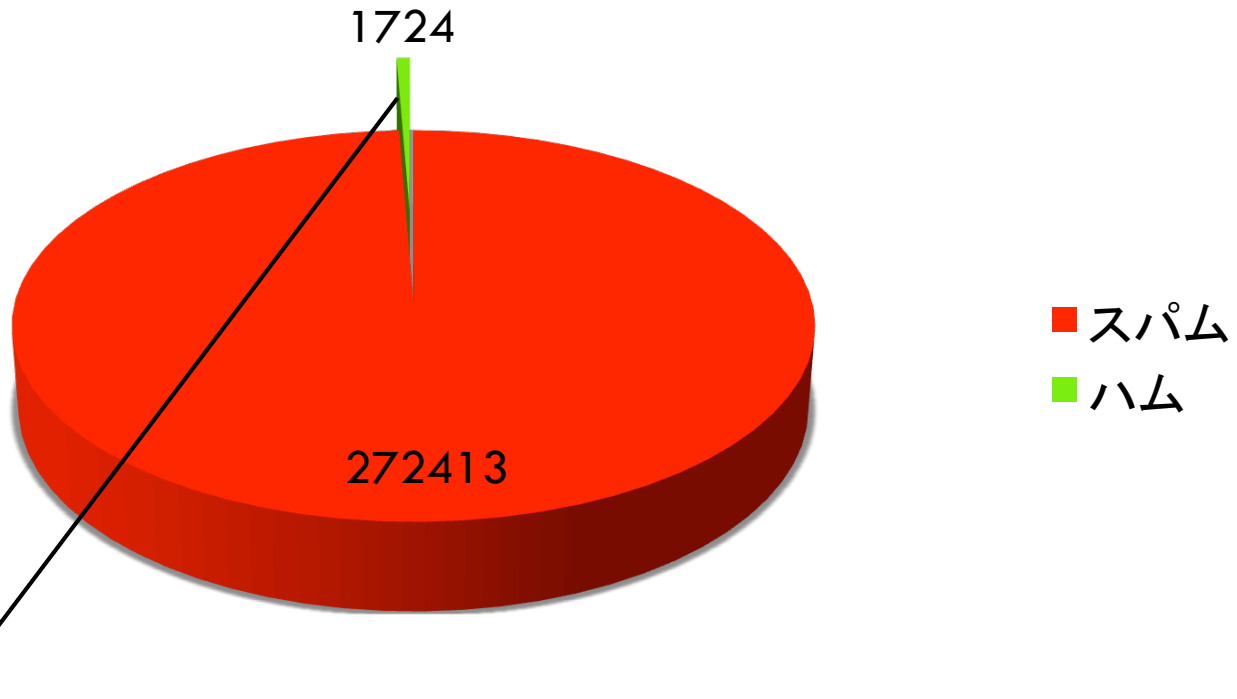
国内発のスパムの大半が greylisting と同様に DNSBL で検出されなかったため、受信スパムのカバー率はそれほど高くない

DNSBLの効果（カバー率）



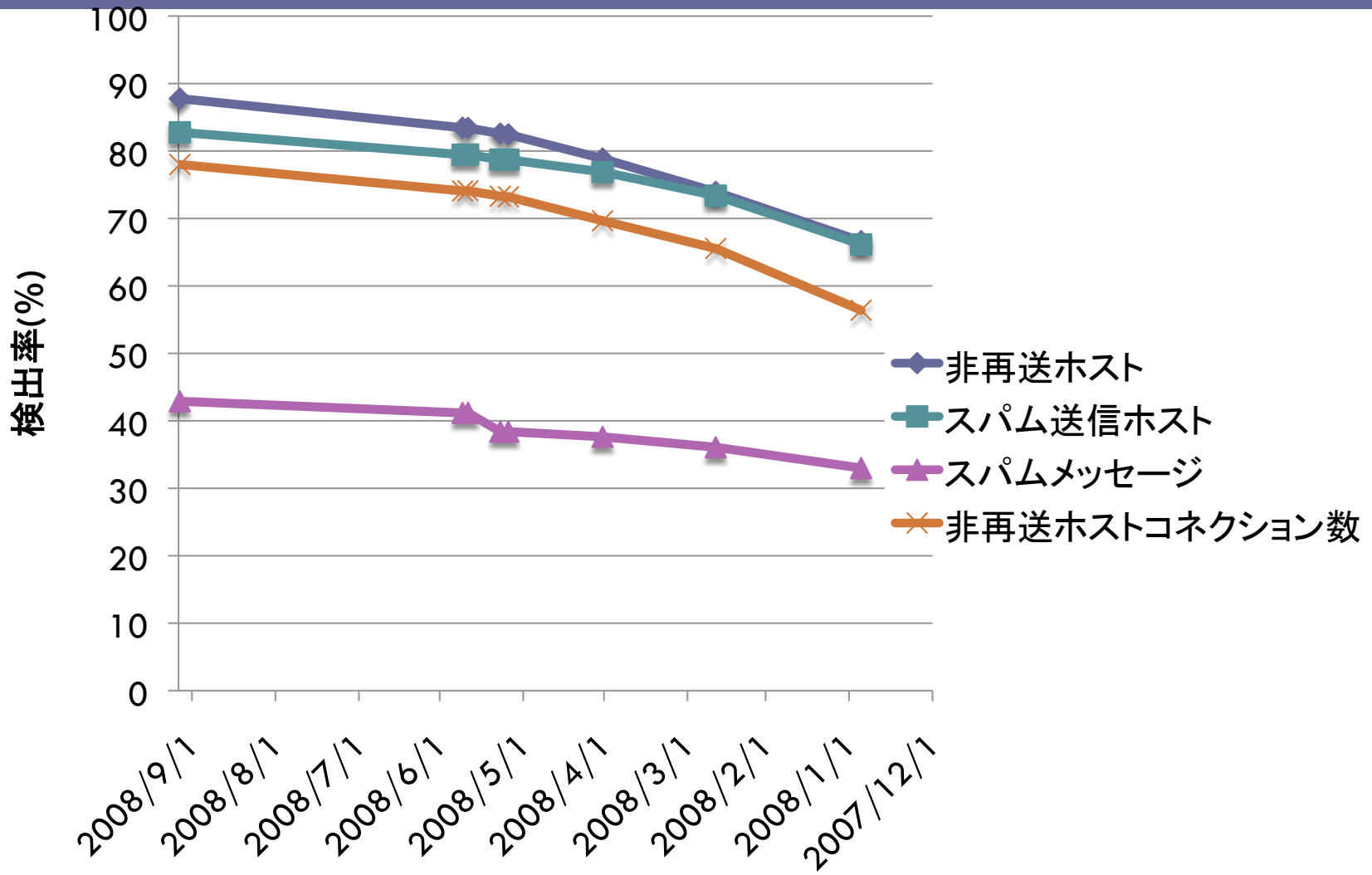
DNSBL(PBL) の検出精度

検出したメッセージの内訳



DNSBLで検出したメッセージにハム(とスコアがついたメッセージ)が含まれている
エラー率 = 0.63%

DNSBL(PBL)の有効性の時間変化



SPFの統計 (ドメイン)

- 調査データ(約4ヶ月のログ)のうち有効な文字で構成される送信元ドメインの数 2,651,037
  - Aレコード,MXレコードが存在しないドメインも含む

- 上記のうち、有効なSPF レコードが存在したドメインの数 128,360 (うち、JPドメインは12,175)

SPFの統計 (IPアドレス・メッセージ)

	IPアドレス数	スパム数	非配送スパム数	ハム数
全ホスト	2,295,918	661,703	17,042,222	689,931
有効なSPFレコードで指定されたホスト	516,873	169,733	4,117,640	521,711

スパムの1/4, ハムの3/4は「有効な」SPFレコードで指定されたIPアドレス発

有効ではないSPFレコードの例

- xxxxx (当日のみ)
  - これらはいずれも “128.0.0.0/1” をかえす
- インターネットの半分は俺んとこ！
  - 世界中の botnet に送らせるため．あるいはmis-configuration?
  - すべて国内の出会い系サイト
  - Greylisting でフィルタされているケースが多い

- 有効ではないとみなした基準
  - Prefix length が 7 bit 以下のアドレス空間を指定(マニュアルでチェック)
  - bogon prefix
  - 無効な dotted decimal 表示 333.333.333.333 など

SPFの利用実態

- SPFを正しく利用しているホスト → SPF-good
  - ハム送信率が高い(9割以上)

- SPFを濫用しているホスト → SPF-bad
  - スパム送信率が高い(9割以上) もしくはメール配送無し
  - ドメイン認証による簡易なフィルタリングを回避するのが目的。敢えてSPFを使う。
  - 踏み台にされている可能性もある。

SPF-good, bad の分析結果

	IPアドレス数	スパム数	非配送スパム数	ハム数
合計	2,295,918	661,703	17,042,222	689,931
有効なSPFレコード有	516,873	169,733	4,117,640	521,711
SPF-good	10,764	1,820	0	425,339
SPF-bad	504,945	117,930	4,117,640	1,184

Good, Bad の両極に概ね集中

各技術の傾向・分析結果

- Greylisting, DNSBLの効果
  - いずれも全体としてのカバー率はいずれも高い
  - 特にメール非配送ホスト(bot等)に対して効果的
  - これらの技術により、ある程度は止まる（フィルタできる）
- SPFの利用実態
  - SPFは送信者詐称検出に使うことが期待できるが、その裏をついているスパマーが存在するので要注意
  - SPFに記載された情報を過信はできない

対策技術の課題

- 有効性はネットワークによって異なる可能性がある
  - 国内発の非ボットによるスパムが多いようなケース
- 誤判定の可能性あり
  - 適用してしまうと申告があるまでわからない
- DNSBLは定期的なアップデートが必要
- カバー率をとるか、精度をとるか（トレードオフ）

議論

- みなさんのところで具体的にどのような対策をしていますか？
- 日々の傾向を分析していますか？
  - 何を見ているか
  - 誤検知してるか
- 今後の対策としてどのような技術が有効でしょう？
  - ユーザの申告・自動化(Feedback Loop), ARF (Abuse Feedback Reporting Format) 等
  - ポリシーの変更が可能なBL (カバー率重視 or 精度重視)
    - 電子情報通信学会 IN/NS 3月研究会にて発表予定
  - その他…

議論

- 他のスパムとの関連性は？
  - SPIT, SPIM, blog コメントスパム, splog, SNS, twitter, …
- 社会科学的（？）なアプローチ
  - スパムがなくならない根源的な理由は何か
- mailop.jp とか作りましょうか?

謝辞

- 早稲田大学CS学科・後藤滋樹教授、下田晃弘氏、魏元氏、石原寛之氏、本嶋悠也氏に感謝します