迷惑メールの現状と対策

有限会社ジーワークス

佐藤 潔

自己紹介

- taRgrey/Rgrey/Starpit
- 有限会社ジーワークス
- 長野県白馬村
- スキーしたかったから

今やってる仕事

- 8 年ほど前から地方 ISP 様から外注
- サーバの管理やユーザサポート業務

なぜスパム対策を考え始めたか

- ユーザの苦情からスパム対策の必要性を痛感
- メールアドレスへの愛着が既存ユーザをつなぎとめている
- 古くからのユーザほど…
  - メールアドレスを変えたくない
  - スパムが多い

こんなことを話します

- スパムの現状
- 対策手法紹介
- おすすめ対策方針
- メール以外のスパム対策
- まとめ

すみません
話したいことがいっぱいあるんで
かっとばしていきます

スパムの現状

- スパムの数
- スパムのメインストリームは bot
- 日本語スパムはスパム 1.0
- 8:2 の法則
- スパム格差
- 分業化
- スパム対策はすでにセキュリティ対策の一環

スパムの数

- 海外だと 90% 以上という報告が多い
- 日本だと 80% ～ 90% 程度という実感
- 年々増えている
  - 3 年前の日本で 70% ～ 80%
  - 5 年前のアメリカで 70% ～ 80%

スパムのメインストリームは bot

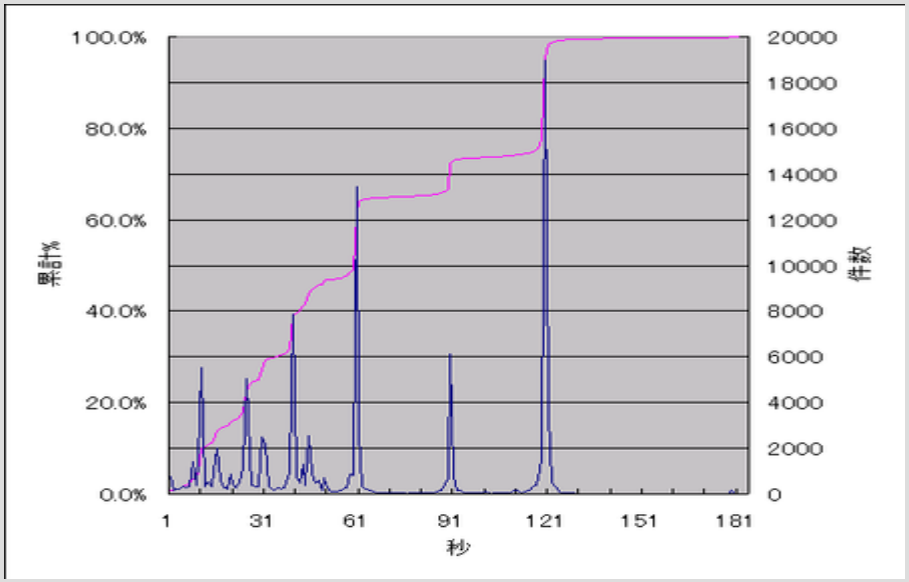
- スパムの大半は海外の動的 IP から
- bot 化した Windows PC で出す
- bot 総数は不明
- とりあえず何 100 万台という単位
- bot が直接送信先メールサーバへ SMTP 接続してくる
- 日本発のスパムは OP25B により激減

日本語スパムはスパム 1.0

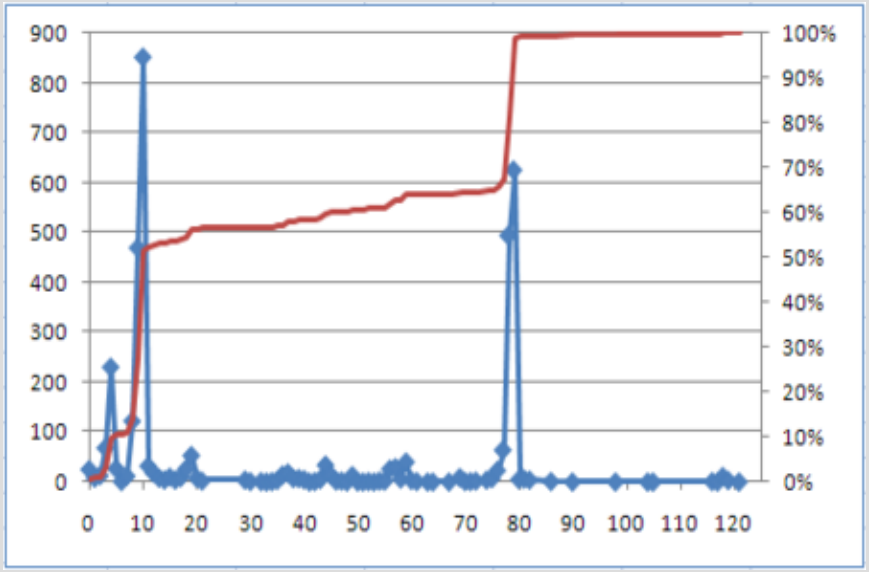
- 日本語スパムは中国韓国等の固定的 IP から
- レンタルサーバか、事務所にサーバを何台も置いて qmail や postfix から出す
- 黒竜江省で摘発された例（タクミ通信摘発）

8:2の法則

- 上位スパム送信業者で大多数のスパムが出されている
- スパムでは95:5くらい？
- 右下図の80秒のピークだけで1スパム業者が1/3のスパムを出していることがわかる



2007/7



2009/5

スパム格差

- 海外ではまんべんなく届く
- 日本ではスパムが届く人と届かない人の差が激しい
- 理由はメールアドレスの収集方法
  - Webからクロール
  - 総当たりや辞書アタック
- Webに載ってるアドレスや単語のアドレスには山ほど届く

分業化

- 世界では分業化が進んでいる
- スパム送信業として専業化
- botネットを持つ
- 日本ではあまり分業されていない感じ

スパム対策はすでにセキュリティ対策の一環

- スパムでWebサイトへ誘導
- 誘導したページでFlashの脆弱性などを突いて感染
- bot化
- そのbotを使ってスパムを撒く

どんな対策手法があるのか

スパム対策する箇所は大きく３つ

- クライアントでの対策
- メールサーバでの対策
- 送信の制限での対策
  - OP25B
  - SPF/DKIMなどの送信ドメイン認証情報の付加

サーバ管理者の方が多いでしょうからメールサーバでの対策手法を紹介

送信者アドレスのブラックリスト

- 一番単純で最初に思いつく対策
- 送信者アドレスは偽装出来るため無力化
- SMTPはインターネットの古き良き時代の性善説で設計されてるからなんでも偽装可能
- 送信者アドレスを偽装できなければ対策しやすいのに…

→ 送信ドメイン認証

送信ドメイン認証

- 送信者アドレスのドメインが偽装されていないかを確認するもの
- SPF
- DKIM

詳しくはumqさんの解説を！

見えにくい・変えにくい箇所のブラックリスト

- HELOのブラックリスト
  - botのスパムには使えない場合が多いが日本語スパムには有効
- 送信者ドメインのNSのブラックリスト
  - 送信者アドレスのメタ化
- 軽くて誤検出が少ないよう設定できる
- 汎用的ではない

禁止ワード

- スパムに良く含まれる単語「VIAGRA」「出会い」とかを禁止ワードにする
- 誤検出の問題
- 「V1AGRA」とかで無力化
- 禁止ワードを自動学習出来れば…

→ ベイジアンフィルタ

ベイジアンフィルタ

- キーワードとその「スパムらしさ」をメールから自動学習
- 強力で非常に広まった
- 学習にSVM使うものとかもある
- ローカライズの問題
- 重い
- スパマー側の対抗策
  - 画像スパム
  - ワードサラダ

画像認識(OCR)

- アンチ画像スパム
- ローカライズの問題
- 糞重い
- CAPTCHAとか分割するとか画像じゃなくRTFにするとかのアンチアンチ

URLBL

- 本文に入っているURLのブラックリスト
- 結構効果が高い
- URLも画像にする対抗策

送信者IPアドレスのブラックリスト

- 送信者の接続元IPがブラックリストに入っていたら拒否
- 動的IPから出す手法で無力化
- ブラックリストの更新が追いつかない。リストの共有出来ないか…

→ DNSBL

DNSBL

- IPアドレスのブラックリストをDNSのシステムを利用して共有
- Botが増えすぎて無力化しつつある
- 他のユーザの巻添え
- IPアドレス帯で登録
- サービス運用者と国が違うとき

動的IPを制限

- S25R
  - 動的IPっぽい逆引き名
- DUL (Dialup User List)
  - 動的IPのDNSBL
- 当然誤検出あるよ

地域でメタ化

- GeoIP
- 中韓台フィルタ
- もちろん誤検出あるよ

NNIPF

- 最近傍識別
- スパム送信者のIPアドレス集合との「距離」で判定
- 動的IPや地域でのメタ化を自動学習しているとも言えそう

greylisting

- 一時拒否して再送してきたら受け取る
- メールの遅延
- 再送しない「正しい」サーバや、違うIPから再送される場合誤検出
- nolisting
  - セカンダリmxで受け取る
- selective greylisting
  - 怪しいところだけgreylisting掛ける

tarpitting

- SMTPセッションの返答を何10秒か待たせる
- せっかちなクライアントは接続を切ってくる
- メールマガジンなどで誤検出起りやすい
- プロセス数の増大
- Starpit
  - selective tarpitting
- taRgrey
  - tarpittingを再送で救済

p0f

- TCP/IPのパケットから接続元のOS種を判定
- 動的IPっぽいWindowsPCからの接続なら怪しいよね

コラボレーションフィルタ

- 参加者にメール内容のシグニチャをブラックリスト登録してもらう
- Pyzor/Razor、CLOUDMARKなど
- 商用で良く使われている
- DCC
  - 多数の同じメールが届いているならスパムだろう
- 人間が判定してるんだから誤検出はない？
  - 他言語の場合など
- 同じセグメントに属する参加者の数が少ないとダメ

ユーザ毎のスパム率を利用

- スパムを多く受け取るユーザのメールはスパムの可能性が高い
- ユーザ毎に重みづけされるDCCとも考えられる

統合型

- SpamAssassinなど
- 各種手法での判定結果をポイント付けして判断
- 商用のものはコラボレーションフィルタとベイジアンフィルタあたりの統合型が多そう

こういう手法が有効みたい

- 自動学習
- メタ情報を扱う
- ユーザからの情報を利用

- ローカライズが必要
  - 英語圏の対策をそのまま持ってきても性能が落ちる

対策の方針

- 誤検出が無いことを重視
- なるべくログを残す
- 各手法を組み合わせて弱点をカバー
- 我々のアドバンテージは「ワンオブゼム」
- 自分が利用しているスパム対策

誤検出が無いことを重視

- 重大な誤検出が起こるとユーザは使ってくれない
- 逆に仕事が増える

なるべくログを残す

- 誤検出は必ず起こると考える
- スパム業者の傾向をつかむためにも

各手法を組み合わせて弱点をカバー

- 一つの仕組みでシンプルなシステムにこだわると誤検出や負荷の問題がおこりやすい
- スパムらしさは加算ではなく乗算

SMTPセッションレベルでの組み合わせ

- postfixだとアクセスポリシーを多段で利用
- sendmailでは？milter managerを使おう！

詳しくはkouさんの解説を！

我々のアドバンテージは「ワンオブゼム」

- 超大手、Gmailやhotmailでは使えない手法も多い
- 特定のメールサーバやサービスが狙われた場合無力なもの
- ちょっと届かないサーバがあっても対策してこない

自分が利用しているスパム対策

- 1次フィルタにSMTPセッションでの「軽い」フィルタ
- 2次フィルタにコンテンツフィルタの「重い」フィルタ
- 1次フィルタにtaRgrey
- 日本語スパム用個別フィルタ
- 2次フィルタにSpamAssassin

1次フィルタにtaRgrey

- S25Rに引っかかったものだけ
- tarpittingを90秒くらい掛ける
- 遅延を待てなくてもちゃんと再送されれば許可(greylistingを利用)

日本語スパム用個別フィルタ

- 素のpostfixやqmailで送られてくるものを防ぐ
- 特に日本語スパム
- HELO/sender fromのブラックリスト
  - HELOが固定だから有効
- NSのブラックリスト
  - ドメインを大量に取って送ってくるもの
- taRgreyと個別フィルタで90%強くらい

2次フィルタにSpamAssassin

- ぶっちゃけ松田さんのTLECレシピ使えば良い
- 日本語化パッチでベイジアンフィルタの誤検出率を下げる
- 2次のコンテンツフィルタでも90%強落とせれば十分
- だから無理に域値を下げなくてもよい

メール以外のスパムへの応用

- PukiWiki用のスパム対策 spam_filter も書いてます
- メールのスパム対策は他のサービスのスパム対策にも応用が効く
  - コメントスパム
  - トラックバックスパム
  - スパムブログ

  などなど

サービス設計時点から
スパム対策を考えておくべき

- 広まったサービスにはすべてスパムがやってくる
- スパムによりそのサービスが機能しなくなる
  - トラックバック
  - コメントスパムの山でブログ廃棄
- あとからプロトコルを変えるのは難しい

共有ブラックリスト

- DNSBL/URLBL
  - そのまま使えるが、セグメントが違うので、コメントスパム用のブラックリストでないと効果が低い
- スパムちゃんぷるー
- DCC
  - 同一の内容の書き込みやトラックバック
  - 大規模なブログサービスならその中だけで十分DCCが働くはず

メタ情報でブラックリスト

- URLのNSのブラックリスト
- フィッシング用のドメインを使い捨ててくる場合に
- GeoIP
  - クライアントやURLの国情報を見る

スパム業者があまり気にしてないところをチェック

- Accept-Language情報
  - CNだったり無かったりしたら怪しい
- ボタンの押下位置情報
  - 画像ボタンだとクリックした位置も送られてくる

selective CAPTCHA

- 「怪しいもの」に対してだけCAPTCHAを掛ける
- 複合条件、例えば「内容に日本語がなくてURLが3件以上ある場合、CAPTCHAを掛ける」とか

ポートスキャンのtarpitting

- ポートスキャンの反応をTCPのWindowサイズを0にしてtarpitting

まとめ

- スパム対策はすでにセキュリティ対策の一環
- 誤検出が無いことを重視
- なるべくログを残す
- 各手法を組み合わせて弱点をカバー
- メールのスパム対策は他のサービスのスパム対策にも応用が効く
- サービス開始時からスパム対策を考えておくべき