spam 送信ホストの見分けかた

前野年紀 † 鈴木常彦 ††

近年の spam はいわゆるゾンビ PC から送られてくるものが大部分をしめている．このゾンビ PC を見分けることにより spam を受信拒否する方法を検討した．接続元 IP アドレスの DNS 逆引き情報を利用するとともに SMTP セッションでの振舞いを観察することで見分けるものである．しばらくの間は spam 受信拒否の方法として有効である．

Experiments about distinguishing spam sending hosts

TOSHINORI MAENO† and TSUNEHIKO SUZUKI††

A new method of blocking spam sent from zombie PC at the MTA is proposed. As spam sending hosts send mails in non-RFC conpliant way, it is easy to tell them from normal MTA by replying 4xx (tempfailng) in SMTP session. And, usually zombie PC is assigned IP address in some range, so it is also easy to tell that the sender is zombie with its IP address. Zombie PC's will send spam for a few years, so the proposed method will continue to be effective for a few years.

1. まえがき

近年の spam の多くはいわゆるゾンビ PC から送られてくる．ウイルス感染して乗っ取られた常時接続の PC である．常時接続とはいえ IP アドレスは変化するので，単純なブラックリストは役にたち難い．一方，多くの spam 送信ホストは遅延応答や一時エラー返答をした場合の SMTP セッションでの振舞いを観察することで判別できることが知られている[1)]．

そこで，接続元 IP アドレスの DNS 逆引き情報を利用して動的割りあての IP アドレスを分別し，該当する送信元に対して遅延返答や一時エラー返答を適用することで spam 送信してくるゾンビ PC を効率よく判別できることを運用により確認する．

以下，第 2 節では spam ホストの SMTP セッションでの振舞いを説明する．
第 3 節では適用相手の選択方法について説明する．
第 4 節ではブラックリストとホワイトリストの維持管理について説明する．
第 5 節では本方式を運用したときの結果を評価する．
最後に第 6 節では今後の課題を述べる．

† 東京工業大学 原子炉工学研究所
Reseach Laboratory for Nuclear Reactors, Tokyo Institute of Technology
†† 中京大学 情報科学部 情報科学科
School of Computer and Cognitive Scie nces ,Chukyo University

2. spam 送信ホストを見わける方法

DNS ブラックリストや逆引き情報を使って spam メイル受信の拒否を試しているときに spam 送信ホストが特徴的な振舞いをすることに気づいた．

(a) SMTP セッションで一時エラーを返答して，受信不能を通知した場合，spam 送信ホストは再接続してこないことが多い．高速配送の妨げになるからだと思われる．SMTP (RFC 2821) に準拠した MTA であれば一定時間後に再接続してくるはずであるので，再接続時に受け取るようにしてやればいい．

(b) 高速配送を妨げるのであれば応答を遅延 (throttling) させることも効果があるはずである．これは tarpitting という名前でも知られている．実際，接続開始時と SMTP セッション中に遅延を入れることにより，spam 送信ホストは接続を自ら放棄することが多かった．接続を放棄してくれれば，再接続時には特別の対応をする必要がない．

(c) 短時間のうちに再接続を繰返すことがある．

(a),(b) は以前から知られていた手法である．

2.1 SMTP セッションでの対応

上記の特徴を利用して，spam ホストを判別するものとする．この方式を『一時エラー返答方式』と呼ぶことにする．

- SMTP セッションでは 20 秒程度の遅延を入れて応答する．
- SMTP セッションが進んで Data コマンドが送られてきたら，一時エラーを返答して受信不能を通知する．

- helo/from/to のパラメタに問題 (詐称など) がないときには, 再接続時に受信するように準備する.

すべての SMTP 接続で応答を遅延することはメイル受信の性能を低下させる. また, 一時エラーを返答することは spam ではないメイルの受信を遅延させることになる. そこで, spam が疑われる相手だけに適用することが望ましい. 次節ではわれわれの採用している選択の基準を示す.

3. 適用相手の選択方法と実装

再送を待つことによる遅延は通常メイルではなるべく避けたい. 『一時エラー返答方式』を適用する相手は spam が疑われるホストだけに限りたい. 接続元の IP アドレスにもとづき相手を選択することにする.

適用しない相手の一覧 (ホワイトリスト) を作成しておき, リストにないものにだけ遅延応答と一時エラー返答を適用する. これは無用な遅延を避けるためだけでなく, 無駄な DNS 検索や dnsbl 検索を避ける意味もある. ホワイトリスト以外にも適用する必要のないものがある.

3.1 遅延, 一時エラー返答を適用しないホストの一覧表や手法

ホワイトリスト: メイル交換の相手が限られている場合にはこの方法が簡単である. RFC を守らないホストなどを救済するのにも使う.

ブラックリスト: すでに spam ホストだと判明している相手の一覧である. 接続拒否, 受信拒否, 遅延適用などの対応を指定できる. あとの二つは情報収集したい場合に使用する.

インターネット上の DNS blacklist: DNS を使って公開されているブラックリストである. 信頼性と登録の遅れが問題になる. 信頼できるものを選ぶのが難しい. 負荷が集中しやすく, スケーラビリティの問題が起きることがある. 次の方法があるので, 特に使う必要はない.

3.2 DNS 逆引き情報を利用した判別

ゾンビ PC の多くが ADSL などの常時接続回線を使っていることから, 逆引き情報を利用することにより, ゾンビ PC が簡単に判別できることが知られている.

no PTR: IP アドレスの逆引き情報が設定されていないことが多い. 過去半年の経験では SMTP 接続してきたもののうち 40 % 弱が該当する. 韓国や中国のプロバイダが管理する IP アドレスに多い. ネットワークや DNS サーバのトラブルにより逆引きできないこともあるので, 逆引きできないことだけを理由に恒久的拒否してはならない. 接続エラーか一時エラーを返答して, 正常な状態に回復するのを待つべきである.

動的割りあて: 逆引きにより動的割りあて IP アドレスを示す名前がえられるもの.

(例) 111.197.8.67.cfl.rr.com [67.8.197.111]

簡単なパタンマッチやサフィックスの検査で判定できるものが多い. ポリシーにより恒久的エラーを返してもよい. 動的割りあて IP アドレスの範囲を公表しているプロバイダもある.

3.3 一時エラー方式の適用相手の選択

throttling/tempfailing を適用する相手は原則として spam が疑われるホストであり, 逆引き情報によって適用を決めるのがよい. spam ホストの振舞いを調査するために, わざと適用する場合もある.

どの方法を使う場合でも, DNS 検索によるネットワーク負荷を軽減すべきであり, ホワイトリストとブラックリストを併用することが望ましい.

3.4 実装例 : お馴染さん方式

筆者らのドメインで運用中の方法を説明する. ホワイトリスト, ブラックリスト, DNS 逆引き, DNS ブラックリストを併用する.

3.4.1 tcpserver による接続制御

- ホワイトリストは受信する相手である. 再送してこない相手や, 接続の度に異なる IP アドレスから送信してくる場合も登録することになる.
- ブラックリストは接続拒否, SMTP 受信拒否, 一時エラー返答する相手である.
- DNS 逆引き情報により, 接続拒否, SMTP セッションで返す返事などを決めている.
- どれにも該当しない場合には DNS ブラックリストを検索して決定する.

3.4.2 SMTP セッションでの対応

- 返答を遅らせる (throttling).
- プロトコルを守って必要なコマンドのやりとりを行ったものには「一時エラー」を返答する. 再接続してきたら受信するように記録しておく. コマンドの引数の検査も取りいれている. SMTP セッションを途中で放棄した相手についてはなにもしない.
- 不応期 : 一定時間内に接続してきたものは spam 送信と見なす.
- 受容期 : 一定時間後の一定期間に接続してきたら, 受信する.
- 一定期間内に再接続してこなかったものの情報は取り消す.

3.5 リストへの反映

動作記録を別途処理して, ホワイトリストとブラックリストに反映する. 方式を検討しながら運用しているので手作業で行っている部分も多い. リスト作成の自動化の検討と実験について次節で述べる.

4. リストの効率的管理方法

すべての SMTP セッションに対して遅延返答や一時エラー返答をするのは非効率である. ブラックリストを作成して, 受信拒否返答 (553) や tcpserver での接続拒否を用いることを検討すべきである. ホワイトリストを充実させることも望ましい. これらのリスト

は手作業で作成・変更することも可能であるが，手作業によるミスやスケーラビリティを考慮して，自動更新する仕組みが欲しい．

4.1 ホワイトリストの更新

本方式で重要なのはホワイトリストの整備である．ポリシーによるが，以下のような基準による管理手法が考えられる．

(1) 再送を行って来たものは全て登録する (ISP，大規模サイト向き．リフレクション[5] で商用運用中)
(2) 自 MTA が送信した相手を事前に登録 (TIC で運用)

また，登録するデータとしては，以下のものが考えられる．

- 相手の IP アドレス
- 相手のプレフィックス (長さは要検討)
- 相手のドメイン
- 相手の SPF[8] をパース

特に 4 点目に挙げたように，送信 MTA の IP アドレスの公表の促進が有効的である．

4.2 ブラックリストの更新 (遷移)

ブラックリストでは接続拒否 (tcpserver での deny)，恒久拒否 (5xx 応答)，一時拒否 (4xx 応答) という対応の使い分けが肝要である．spam 送信の疑いの濃いセッションで一時拒否が延々と続く場合があり，リソースの無駄を生じさせる．受信の要不要を判断し，ある段階で接続拒否または恒久拒否へ遷移させるものとする．TIC[2),4)] で行っているように，個々のセッションを判定するスクリプトを使ってブラックリスト (tcpserver のルール) を書き換えるという実装が考えられる．判定に用いる材料としては以下のものが考えられる．

(1) 動的割当 IP アドレス (PTR 等で判断)
(2) PTR が付与されていない
(3) SPF 等の宣言範囲外
(4) 再送間隔が異常に短い
(5) 利用者からの申告[6]
(6) RCPT TO が存在しないメイルアドレスを指している (恒久拒否して良い)

恒久拒否については慎重を期さなくてはならないが，受信するサイトのポリシーで決める問題である．受信 (拒否) ポリシーを明確に公開することが重要であるが，「RFC を守っていること」までポリシーで述べる必要があるとするなら，議論の対象はインターネットではないのかもしれない．我々が最適と考える更新のフローは以下のようになると考えられる．

(1) 一度目の接続は tcpserver で deny (再接続時には allow)
(2) 再接続してくれば，throttling + 一時エラー返答
(3) data コマンドまで到達したら，パラメータを分析してブラックリストまたはホワイトリストを更新し再送信を待つ
(4) さらに再接続してくれば，受信または 5xx での拒否

現実的にはサイトに応じた管理方法を作ることになり，TIC[2),4)]，渥美[6]，リフレクション[5] 等で実装が進められている．

5. 運用記録の分析と評価

本方式は実装が簡単で，負荷も軽く，著者らのサイトでは spam 拒否の効果が得られているが，spam 対策の効果を客観的に示すことは難しい．

つぎの理由で再現性がなく，比較が難しいからである．spam を送信してくる相手は受信サイト (メイルアドレス) によって異なる．時間とともに変化する．対策をしたことによる影響もあり得る．

はっきり言えるのは対策後は受信する spam はホワイトリストにあるホストから送られるものだけになっていることである．そして，拒否した相手は SMTP (RFC2821) を守っていないということである．

5.1 hpcl.titech.ac.jp メイルサーバへの接続の状況

- 期間：2 月初めから 9 月 10 日まで
- 総アクセス件数は 48888，IP アドレスは 12821 であった．
- うち 26424 件 (10991 IP アドレス) は tcpserver で接続拒否した．(逆引きできないもの，動的アドレス検出用パタンにマッチするもの)
- tcpserver 接続を受入れたものは 22464 件 (1969 IP アドレス) であった．
- SMTP 接続を受入れたもののうちで，送信側から切断したものを含め，一時エラー，恒久的エラー返答したものは 5197 件 (1885 IP アドレス) であった．

5.2 遅延応答 (throttling) の効果

応答を遅らせてもめげずに SMTP セッションで data コマンドまで送ってきたものは接続数が 2090 件 (40%)，IP アドレスは 573 (30%) であった．(最近は 5xx 返答には throttling していない．) 遅延応答は判別に有効であると言える．

5.3 逆引き設定の分析

全ての接続元 IP アドレス 12821 のうち，逆引きできないもの (PTR なし) は 4756 (37%) であった．kr，cn の ISP に属するものが多い．

ホワイトリストにある件数は 120 件 (1%) にすぎず，PTR レコードがあった 60 % 強は大部分が動的割りあて IP アドレスである．

5.4 接続回数の分析

tcpserver で deny した IP アドレスの接続回数を見ると，

- 1 回だけの IP アドレスが 8457 (78%)
- 5 回以下のものの累計は 10386 (96%) であった．

spam ホストは短時間のうちに 5 回連続して接続してくることがよくある．5 分以内の再接続は受信しな

いというポリシーが有効である.

接続回数の多い順で上位の 100 個中の 72 が spam ホストであった. (最高回数は 624 回である. 存在しない宛先を指定していることから, spam と判定できた.)

5.5 受信拒否の効果

送られてくる spam はホワイトリストに登録されたメイリングリスト経由のものばかりという状態になっている. 接続してくる spam ホストの数には特に変化は見られない.

整備されたホワイトリストがあれば, 遅延応答と一時エラー返答により満足できる spam 撃退が可能である.

再送時に一群の送信ホスト集合から不規則に選んだホストから送ってくるドメインがあるのでこれらはホワイトリストに登録して対応した. SPF などの送信サーバを宣言する方法の普及が期待される.

6. 課　　題

現在の spam 送信元の大半をしめるゾンビ PC を判別することができた. しかし, 特定のサイトにおける結果であり, 定量的な評価は難しい. 例えば, 受け取らなかった spam の数はわからない. 分かっているのは RFC に従っていない送信方法を使う相手を接続拒否した数である. spam であることを確認するために受信してみることは状況を変えることになる.

spam を受信していないという事実で, この方法を採用するサイトが増えれば, 評価もしやすくなるはずである.

6.1 本方式の持続性

今後も spammer の送信方法が同じままであるとは考えにくい. しかしながら, 遅延につきあい, 一時エラー返答には再送するということであれば, それは spam 送信により多くの資源が必要になるということであり, spam を抑制することになる. また, 長期間, 同一のホストからの送信を続けることはブラックリストに登録されることになり,DNS ブラックリストの効用が増すということであり, 受信拒否につながる.

6.2 ISP での対策

簡単に実装できるとは言え, すべての受信サーバで spam 受信拒否対策するのは手間がかかる.

spam の発信元は ISP 管理下の動的割当て IP アドレスを使っている. この ISP 配下の IP アドレスから ISP 外の port 25 への直接接続を制限あるいは禁止すれば, ゾンビ PC からメイルサーバへ直接送られてくる spam は激減するはずである. すでに実施して, 効果をあげている ISP もある.

IP アドレスの判定の有効性を高めるためには ISP のアドレス空間に関する情報公開が必要であり, SPF や RMX の普及も望まれる.

6.3 通常のメイルサーバ からの spam

通常の MTA から送られてくる spam についても対策を考えなくてはならない.

- ISP は spam の中継をすべきではないし, spam 送信を禁止する手段を用意すべきである. 実施には社会的なサポートが必要になる.
- メイリングリスト管理者は spam を配布することがあってはならない.
- 詐称送信者アドレスへの送り返し (bounce) はやめるべきである. 宛先が存在しないからと言う理由でいったん受けつけたメイルを送信者に送り返すと, 詐称に使われた無関係な相手に spam を転送することとなり, 被害を拡大する. ウィルスの警告メイルも余計なことである.

これらの spam の多くはフィルタリングを利用して排除できるものだが, 送信側での抑制が望ましい.

6.4 バウンスを抑制することの必要性

発信者アドレスが詐称された spam を宛先不明であるとか, spam であるとかの理由でバウンスさせると詐称された送信者へ spam 被害を拡大することになる. 場合によっては DDoS 状態を生み, インターネット全体に害を撒き散らすこととなる.

自らが不要なバウンスを生成することだけは避けなくてはならない. TIC では「バウンス対策のために spam 対策を行う」というのが真の動機であった. spam を徹底して受信拒否することは不要なバウンスを発生させないことと同義となる.

7. ま と め

spam 送信ホストに使われているゾンビ PC を見わけるための手法の有効性を検証した.

SMTP セッションでの振舞いによる判別では, SMTP 応答を遅らせて spam 送信をあきらめさせるのと一時エラー返答を利用してメイルが再送されるのを待つことにより, 通常ホストを選別する方法は有効であった. RFC に従わない送信方法の通常ホストも排除されるが, 必要ならホワイトリストに登録することで受信可能とできる.

一時エラー返答によるメイル受信の遅延を避けるためには, ホワイトリストを作成するのがよいが, 接続元の DNS 逆引き情報の有無による判別や逆引き名にもとづく動的割りあて IP アドレスの判別も有効である.

簡単に実装できて, 負荷も軽く, 大きな効果がえられるので, MTA での spam 撃退法として広く使われることを期待する.

参 考 文 献

1) 前野 年紀 : MTA でできる spam 撃退術, 情報処理学会, 第 45 回プログラミング・シンポジウム報告集 pp. 135-145, (2004).
2) 東海インターネット協議会: MTA における spam 対策, http://www.tokai-ic.or.jp/spam (2003-2004).
3) 後藤 邦夫: 迷惑メイル対策の現状と課題, 東海

インターネット協議会 2003 年度会報, pp. 52–60 (2004).
4) 鈴木常彦,et.al:MTA による spam 対策の実践報告,DSM-34
5) http://www.reflection.co.jp/spam/
6) 渥美清隆: アクセス制御と SPAM フィルタを組み合わせた動的 SPAM 拒否システム,dsm-33
7) Danisch, H., The RMX DNS RR and method for lightweight SMTP sender authorization, Internet Draft, draft-danisch-dns-rr-smtp-04.txt, (May 2004).
8) Sender Policy Framework, `http://spf.pobox.com/`.
9) 前野年紀: spam 関連の解説記事 http://moin.qmail.jp/
10) Brad Templeton: Why "Bulk Mail from a Stranger" is the best definition for spam. http://www.templetons.com/brad/spam/define.html
11) The Economics of Spam http://cc.uoregon.edu/cnews/summer2003/spameconomics.html