spam 判別における資源節約

前野年紀　　鈴木常彦

東京工業大学 原子炉工学研究所, 中京大学 情報理工学部 情報システム工学科

Ecological Anti-spam measure

Toshinori Maeno and Tsunehiko Suzuki

Research Laboratory for Nuclear Reactors, Tokyo Institute of Technology
School of Information Science and Technology, Chukyo University

spam 送信ホストをより少ない資源で精度よく判別する手法を提案する。

spam を送ってくる bot の識別には DNS PTR レコード情報が有効である。[3] しかし、PTR レコードを入手するには ネットワークに負荷がかかる。そこで、SMTP helo コマンドのパラメタを DNS PTR レコードの代わりに使うことで bot を判定する。判別結果は受信拒否、受信受入れ、再送要求などの決定に利用できる。

spam 判別に有効なもうひとつの方法である牛歩戦術は受信システムの資源に対する負荷が大きい。そこで、一時エラー返答に対して、spam ホストが再接続を試みるまでの経過時間によって spam 判定する。これらを合わせて利用する資源消費の少い spam 判別法を提案する。

Keywords: MTA spam blocking, SMTP helo parameter, tempfailing, throttling

1　はじめに

受信するメイルの大半をしめるようになった spam を判別することが重視されて、spam 対策のために使われる資源にはあまり注意が払われていない。例えば、メイル送信元の IP アドレスから DNS PTR 情報 (逆引き) を入手するために使われるネットワーク資源である。

そこで、資源利用をできるだけ節約して spam 送信ホストを判別することを検討した。第 2 章ではこれまでの spam 判定法とその資源利用について述べ、第 3 章で SMTP helo コマンドを利用した資源節約法を説明する。第 4 章では保留返答に対して再接続してくるまでの経過時間から spam ホストが判別できることを示す。第 5 章では spam 対策においてはポリシーが重要であることを説明する。第 6 章は全体のまとめである。

2　これまでの spam 対策

かつて提案してきた spam を受信しない方法 [2, 3] ではメイル受信サーバにおいて牛歩対応 (throttling) したり、保留/一時エラー返答 (tempfailing) したりする。そして、動的割りあて IP アドレスホストの判別には DNS PTR 情報を利用していた。これらはいまも判別に有効な手法であるが、大幅に spam が増えた現在では、大量のメイルを扱う環境にはネットワークに対する負荷が大きく、資源利用の面からは望ましくなくなってきた。

その理由は牛歩対応は受信プロセスにサーバシステム資源が占有されるため、システム資源を圧迫するからである。spam を受信処理することに比べれば負荷は十分小さいはずだが、資源を要求していることは間違いない。

保留/一時エラー返答は通常のメイルクライアントは再送してくるのに spam 送信プログラムは再送してこないことを利用している。再送を識別するために接続のあった IP アドレスなどの記録が必要である。

メイル送信サーバの逆引き設定は義務ではないので、設定されていないホストもある。なにより、再接続してこないホストの逆引きはネットワーク負荷になるだけである。

これらの資源消費を減らすには、対策手法を適用する場面を限定するのがよい。複数 MX ホストを利用する MX 遷移検査法 [1] は分別に有効であった。

ローカルホワイトリストやブラックリストなども有効であるが、しかし、リストを維持する負担が発生する。共用ブラックリストには負荷集中という問題がある。

まずは各種の対策の資源要求について検討する。

2.1　spam 送信ホスト名による判別

多くの spam は spam 送信者にあやつられたゾンビ PC (bots) から送られてくる。bots はウイルス感染して乗っ取られた PC であり、多くはプロバイダから動的に割りあてられた IP アドレスを使っている。

動的割りあての IP アドレスには逆引きレコードが設定されていなかったり、それと分かる名前の DNS PTR レコード (逆引き) が設定されていたりする。このことが spam ホスト判定に使える。

送信元 IP アドレスから DNS 逆引きによりホスト名前を求めて、判定する方法はとても効果的であり、すでに広く使われている [3]。そして、メイルサーバに DNS PTR レコードを設定することは広く行われるようになってきている。

2.2　DNS 逆引きの問題

しかしながら、DNS 逆引き操作は手間がかかる (DNS 問合せを多段に行う) ため、spam 対策としての逆引きが世界中で利用されると逆引き DNS サーバやネットワーク全体の負荷となる。

また、特定の IP アドレスからは一回限りの接続が多いという spam 送信の特徴のため、DNS のキャッシュは効果が少い。逆引きが設定されていない場合は検索回数も増えることになる。

あわせて、DNS 設定の不良もネットワーク全体の負荷を増大させる。逆引き DNS サーバにかかる負荷は受入れるとしても、DNS サーバ一般の管理が不十分な現状で、大規模利用することは望ましくない。安定しているとは言えない、セキュリティ面でも心配のあるサービスである DNS に依存する spam 対策は勧めたくない。

そこで、DNS PTR レコード相当の情報が得られる SMTP helo コマンドを利用する判別法をのちほど検討する。

2.3　牛歩戦術に必要な資源

牛歩対応とは疑わしい接続に対して、数秒から数十秒の間プロセスをスリープさせ、接

続を諦めさせたり、大量送信を妨害したりする方法である。大規模メイル受信サーバでは受信プロセスにシステム資源が長時間占有されるため、資源を圧迫するという問題を引きおこす。サービス不能攻撃に使われる心配もある。

最近は spam 送信も特定のターゲットを狙うものに変化しているようで、簡単には諦めてくれず、スリープの効果も少くなってきている。今回の提案では牛歩戦術は使わないこととする。

2.4 一時エラー返答のための記録

一時エラー返答は SMTP 保留返答に対して spam 送信プログラムが多くの場合再送してこないことを利用する対策である。再送を識別するためは接続の記録が必要となる。

通常のメイル送信プログラムの再接続までの時間は 30 分程度から数時間であり、その間は記録として保持する必要がある。この記録は大部分が使われないということにより目的を果すものであるので、ホスト名などから spam だと判定できる接続は記録に残さないことにして、記録資源を節約する。

3 helo 検査

DNS PTR 検索で入手できる接続元ホスト名は SMTP [4] helo/ehlo コマンドのパラメタ (以下、helo パラメタと略記) でも入手できることになっている。こちらの方が DNS 逆引きよりも手軽に入手できるので、この helo パラメタにより送信元の bots 判定に使えるかを調べる。

3.1 SMTP セッションの始まり

SMTP セッションの開始時のやりとりは以下のようになっている。

1. 送信ホスト (クライアント) は受信サーバの TCP port 25 への接続を試みる。
2. 受信サーバは TCP port 25 への接続を受けつける決定をしたら、greeting と呼ばれる返答をクライアントに送り返し、クライアントからの SMTP helo メッセージなどを待つ。
3. greeting を見て、送信ホストは ehlo または helo コマンドを送る。ehlo は SMTP 拡張仕様を使うためのコマンドである。RFC 2821 [4] では helo コマンドではなく、ehlo コマンドを使うことが奨励されている。

ehlo が奨励されている現在では helo を使うホストは spam の可能性が高い。helo を使い続ける通常ホストはホワイトリストに登録することとし、その他は spam として、保留返答をする。

ehlo コマンドを使うホストに対しては、helo をサービスしていないサーバを装って「コマンドエラー」を返答してみているが、最近の spam はこれにも対応しており、効果は小さい。

3.2 helo パラメタ

helo パラメタは送信ホストの DNS 完全修飾ドメイン名 (FQDN) とすることになっている。これが本当に送信ホスト名を表わしていれば、DNS PTR の代わりとして、動的割りあて IP アドレスの判定に利用できる。実際、helo パラメタを spam 判定に使っているメイルサーバソフトウェアも存在する。

しかし、過去には間違った設定のホストが存在したために、通常のメイル受信サーバ (プログラム) は helo パラメタを検査しない。RFC でも「間違いを理由に受信拒否はしないように」と書かれている。そのためいい加減な設定が残っている。

設定ミスの可能性があるものをどう扱うかは受信サイトのポリシーの問題である。送信

側でも外部にメイル送信するときに、正しい名前が設定できていないようなサーバから送るのかどうか、送っても受け取ってもらえると考えるのかどうか、spam と間違われないように送信方法をよく考えるべきである。

spam ホストの多くも本来のホスト名 (FQDN) を送ってこない。bots はホスト名の取得が困難という理由もあり、簡単に spam だと分るパラメタを送ってくるものもある。この状況を利用すれば、spam ホストの識別が可能である。

3.3 helo パラメタの分類

送られてきた helo パラメタを調べてみると、間違いを含め、以下のように分類される。分類に応じて、対応方法を決める。

前提：受信サーバはインターネットからのメイルを受信するものとし、ローカルネットからのメイルは別途処理するものとする。

1. 受信サーバの IP アドレスやドメイン名:
   例：`131.112.32.6, jp.qmail.org`
   前者は構文誤りでもある。これらは spam であることが経験的に分っている。受信拒否してよい。これらを使う spam プログラムは近いうちに消滅するだろう。

2. IP アドレス風の文字列：
   例：`192.168.3.1`
   これも構文誤りである。spam である。

3. IP アドレスを '[]' で囲んだもの：
   例：`[192.168.3.1]`
   構文としては正しいが PTR レコードが存在しないことをしめす。

4. 動的割りあてらしき名前 (FQDN):
   例：`236-22-236.ded.tie.cl`
   逆引き名であると仮定して対応をきめる。動的割りあてアドレスを示す名前らしければ、受信拒否あるいは一時エラー返答をする。
   adsl, ppp などの文字列が含まれていれば、動的割りあてである。

5. ピリオドを含まないか、末尾だけにピリオドをもつ名前:
   例： `pc48, localhost, star-wars.`
   設定ミスの可能性もあるが、逆引きはできない。多くは spam である。

6. 存在しないトップレベルドメインをもつ名前:
   例:`seby.localnet.localdomain` 設定不良であり、逆引きはできない。多くは spam である。

7. ピリオドがひとつの名前:
   例: `yahoo.com, hotmail.com, mail.com, xxx.info`
   ホスト名を含まないドメイン名の多くは spam である。info TLD のドメイン名の多くは spam である。著名なドメインのドメイン名だけを使うものは詐称である。

   DNS を利用して A レコードを検索することは、DoS 攻撃に加担する恐れがある。PTR 検索で確認する。

8. 逆引き名であることを隠そうとしている生成した名前:
   例： `sguq.chello.pl`
   chello087206124055.chello.pl が使っていたもの。著名なネットワークプロバイダ下の動的割りあて IP アドレスを疑うべきである。

9. その他のメイルサーバ風の名前など
   例： a.mx.jp.qmail.org
   正規のメイルサーバらしいときは再送を待つ。

3.4 helo パラメタの観察例

spam しか送られてこないあるドメインに対して、2380 の IP アドレスから送られてき

た helo パラメタを分類してみる。

1. サーバの IP アドレスかドメイン名 (FQDN ではない) 1496 (63%)
2. 動的割りあてを示唆する長い名前 330 (14%)
3. ピリオドを含まない (最後のピリオドは除去) 136
4. その他 418

分類 1 のものの多くは DNS 逆引き設定されていない。

3.5 helo で spam 判別してよいか

helo パラメタの設定不良を理由に受信拒否しないようにという RFC の規定は helo パラメタがまともに設定できないサイトが多かった過去の規定である。また、ローカルネットワークでの中継サーバを意識した記述であると考える。

現在では、インターネット接続されたメイル送信サーバにおいて helo パラメタを正しく設定することはメイルを受信してもらうための必要条件である。

4 再接続の状況

多くの spam は再送を試みないことが分っている。かつては再送するものは 2 割に満たなかった。このことから、一時エラー返答が有効な対策であることがわかる。しかし、最近では一時エラー返答を通り抜けるためなのか、再送するものが増えてきている。ある受信ドメインでのこの半年間の (約 170 万件の接続) 記録では一度きりの接続ホストは 7 割弱であった。(図 1)

ただ、spam 送信は「あきらめがいい」という傾向は変っておらず、再接続までの時間間隔も短かくて再送する回数も少い。(図 2、3)

そこで再接続のパタンを判別することで、遅延を増すことなく判別可能である。以下の判別法を提案する。

4.1 再接続させるための判断

1. 短時間での再接続を繰返すもの、11 分以内の再接続は spam の可能性が高いので、接続を認めない。そして、接続の度に起点時刻をリセットする。

   これにより、短期間での再接続は拒否し続ける。23 分間以上は接続動作を継続しない spam が多い。

2. 11 分を越える間を置いて接続したもの、あるいは最初の接続から 3 時間以内に同じ IP アドレスから再度接続してきたものは接続させる。

4.2 再接続時の処理

再接続させたものについては対象が少くなっているので、DNS PTR を検索してホスト名を取得し、動的割りあての IP アドレスでないことを確かめてもよい。helo パラメタを使うのでもよい。再接続時に最初の helo と異なる名前を返すものは spam ホストであるが、確認のためには helo パラメタを記録しておく手間が生じる。

長い名前 (27 文字程度) や 数字を多く含むホスト名は動的割りあて IP アドレスの可能性が高いので、再接続であっても保留返答しつづけてよい。

helo 検査をしたあとは、mail from コマンドの構文検査など、SMTP セッションでプロトコルを遵守しているか調べて問題がなければ受信する。牛歩対応して相手が切断しないことを確かめるのも有効な手法である。

spam の疑いがあるときは、保留返答しつづける。このとき、メイル管理者に通知して判断を求めるのが親切である。

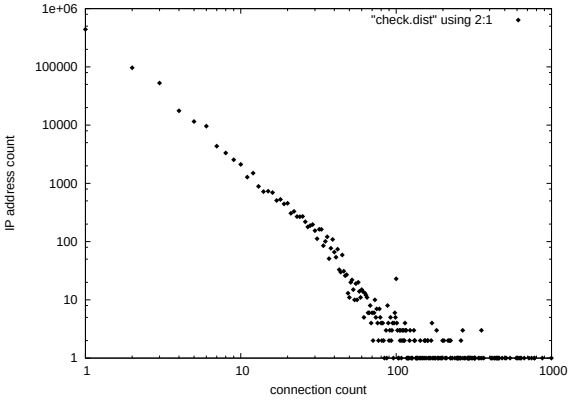

図 1: 接続回数：ホスト数

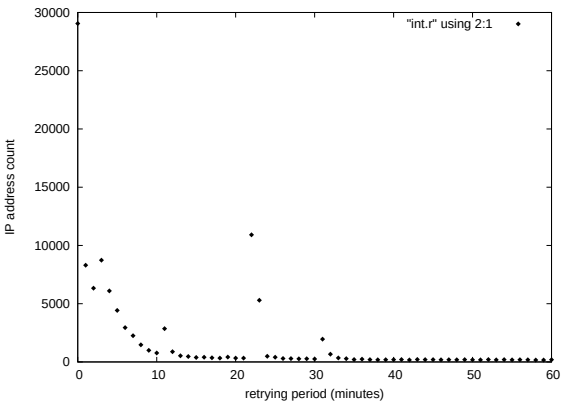

図 2: 接続期間：ホスト数

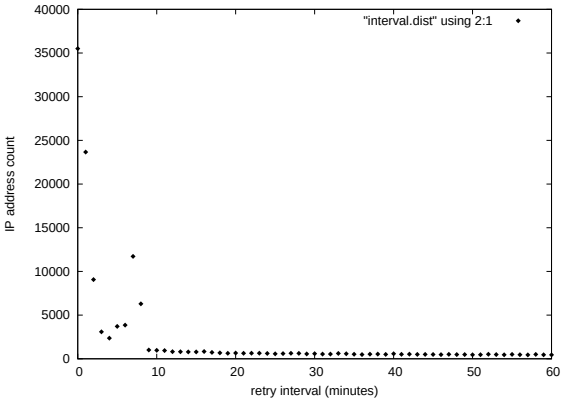

図 3: 接続間隔：ホスト数

5 討論

5.1 spam の定義について

spam でないものを spam だと誤判定する危険を抱える方式なのに spam と判断した根拠を説明しない (できない) 方式が存在する。利用者には spam の定義が明かでないので、対応に困る。受信側で spam に分類されたメイル群が保存されているなら、誤判定されたメイルを利用者自身で捜し出すしかない。これだと spam を自動分別した効果も小さくなる。

保留/一時エラー返答や MX 遷移検査方式 [1] では受信しなかったメイルは明確に理由を示すことができる。発信者にも通知が届くはずである。

メイル受信側 (サイト、サーバ) が受信のポリシーを明確にしておけば、メイルを受信して欲しい送信者は相手のポリシーを尊重するであろう。そして、自分が spam 送信者ではないことを明確に示そうとする。これが、現在のメイルシステムを延命させるための対策である。

5.2 メイル受信のポリシーについて

spam と判断したものは受信拒否するが、疑わしい状況では spam 判別の目的 (ポリシー) により、行動が異なる。一時エラー返答による遅延は受容するとして、どう行動するかのポリシーについて、ふたつの立場にわける。

1. spam であると判断できなかったものはすべて一時エラーを返し、再送を待つ。

   再送を行う spam が増えているので、再送のパタンを判断材料に加える。多数回、長時間の再送を継続する spam はまれであるので、一度の再送では受信せずに、複数回の再送を待つこともあり得る。ホワイトリスト作成が必要となる。

2. spam であると判断できなかったものはすべて受信する。受信後の spam 判定

ツールに期待する。受信するメイルが増え、サーバの負荷が大きい。

この場合はブラックリストの作成が重要になる。しかし、送信ホストを取りかえて送ってくる (bots) spam に対してはブラックリストは有効ではない。

ホワイトリストを充実する方が建設的である。第一の立場をとるサイトが増えるであろう。

5.3 spam 対策の比較

今回の方法とこれまでの方法について、資源負荷と受信遅延の大きさを相対的に比較してみる。

方法	負荷	遅延
保留返答	小	大
牛歩対応	大	中
DNS blacklist	中	?
DNS PTR	中	中
MX 遷移検査	小	小
helo parameter	小	小
再送間隔検査	小	大

表 1: spam 対策の比較

再送間隔検査は一時エラー返答 (保留返答) を前提にしているが、保留返答と比べて遅延が増える訳ではない。通常のメイルサーバが再送するまでに待機している時間を監視しているだけだからである。

5.4 spam 送信方法の変化に対して

spam 送信ホストの特徴をとらえて spam であるかを判断する方法はメイル本文を検査しない方法であるために、spam 送信方法が変化して、対策の効果が低下することはあり得る。

その場合には、永続的な spam 対策が見つかっていない現在はあたらしい対策をいろいろ提案することになる。

6 おわりに

資源負荷の小さい spam 対策手法を提案した。一時エラー返答に対する再接続間隔を観察する手法と SMTP セッションの helo 情報による判定を組み合わせて、DNS 逆引きや牛歩戦術などの手法で使われる資源負荷を回避した。

spam ホストを判別するのに有効な DNS PTR 情報は検索の負荷が大きい。そこで、SMTP セッション中にメイル送信ホストが送ってくる helo コマンドのパラメタが SMTP プロトコル (RFC) に従っているかを検査することで、spam ホストを推定できることを確かめた。helo パラメタを判定に使うことの正当性についても議論した。

サーバに負荷をかける牛歩戦術は使わないこととし、受信サーバには負荷とならない一時エラー返答を検討した。一時エラー返答に対して再接続してくるまでの経過時間が判定材料に使えることをしめし、これを helo による判定と組み合わせて用いる手法を提案とした。

DNS PTR 検索に頼らずに spam ホスト (候補) を絞りこめるので、ネットワークや外部の DNS サーバに負荷をかけずにすむ。また、受信サーバでは牛歩戦術による負荷を回避できた。より使い易い spam 対策ができあがった。

普及が課題である。

参考文献

[1] 前野年紀・鈴木常彦：環境に優しい spam 対策, 情報処理学会, 第 48 回プログラミング・シンポジウム報告集, pp. 49-56, (2007).

[2] 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).

[3] 前野年紀・鈴木常彦：spam 送信ホストの見分け方, 情報処理学会, DSM シンポジウム 2004 年度論文集,pp.25-29, 2004.

[4] http://www.ietf.org/rfc/rfc2821.txt

[5] http://www.reflection.co.jp/spam/

[6] 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).

[7] Spam Filtering for Mail Exchangers: 2.3. SMTP checks http://www.linux.com/base/ldp/howto/Spam-Filtering-for-MX/smtpchecks.html