環境に優しいspam対策

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

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

spam 送信ホストを少ない資源で精度よく判別する手法を評価した。

これまでの spam 対策法で問題とされた受信遅延 (一時エラー返答法) やシステム資源の消費 (牛歩戦術、throttling) などを軽減する方法を検討した。DNS 逆引きを使って対象を選択する方法もあるが、それも避けることにして、複数 MX を利用した spam 判別法と SMTP helo コマンドを利用する判別法を評価したところ、有効性が確認できた。複数 MX 判定法の効率よい実装法が課題である。

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

1　これまでの spam 対策

メイル受信サーバにおいて、メイル受信時に牛歩対応したり一時エラー返答することは spam を受信しない有効な手法である [1] が、利用環境によっては問題点となることもある。

牛歩対応 (throttling) では 1 分程度の sleep であっても受信プロセスにシステム資源が占有されるので、大規模メイルサーバでは資源を圧迫することになる。spam を受信処理することに比べれば、負荷は減っているはずではあるが。

一時エラー返答 (tempfailing) は多くの spam 送信プログラムが再送してこないことを利用している。通常サーバからのメイルは再送を待つことになる。この 15 分から 30 分程度の再送による遅延を避けるためには、即時受け取りのためのホワイトリストの充実が欠かせない。一方で初

見のホストからのメイルについても遅延発生を嫌う人達がいる。送信側の立場からは再送の手間そのものを非難する声もあるが、インターネットメイルというものの性質なので、これは相手にしない。

これらの理由から、上記の対策を適用する相手を限定できるとよい。ホワイトリストやブラックリストを使うなら、その充実が欠かせない。リスト作成を容易にする道具や代替手段が望まれる。

さて、多くの spam は spam 送信者にあやつられるゾンビ PC (bots) 経由で送られていて、それらでは動的割りあての IP アドレスが使われていることが判明している。そして、動的割りあての IP アドレスはそれと分かる名前をもつ DNS 逆引きレコード (PTR) が設定されていたり、逆引きレコードがなかったりするなどの特徴があることが知られている。このことから、送信ホスト選別手段として送信元の IP アドレスを DNS 逆引きして判定材料にする方法が有効であり、すでに広く使われている [3]。ただし、逆引きは逆引き DNS サーバやネットワーク全体の負荷となる。また、逆引き DNS サーバはきちんと管理されているとは言えない状態であることにも注意すべきである。つまり、安定しているとは言えないサービス、セキュリティ面での心配のあるサービスである DNS に依存することになる。

DNS ブラックリストも有効な手法であるが、逆引き同様の問題がある。

これらの理由から、DNS に頼らない判断方法を検討すべきだと考えた。

以下に、複数の受信サーバを使いわける MX 遷移検査法と SMTP helo コマンドを利用する判別法を説明し、これまでの方法と比較してみる。

2 MX 遷移検査方式

ドメインのメイル受信サーバ (DNS MX レコード) は優先順位つきで複数個設けてもよいことになっている。SMTP (RFC 2821) によると、送信ホストの振舞いは以下のようになる。

> 受信ドメインに複数のメイル受信サーバが存在するとき、最初に最優先の受信サーバに接続して、メイル送信を試みる。接続あるいは送信に失敗したときには次の優先順位の受信サーバに対して送信を試みる。複数の受信サーバがあるときには、最低でも二つのサーバを試すことになっている。

しかしながら、多数のメイルを送信したい spam は上記の振舞い (以降 MX 遷移と呼ぶ) をしないことが多いことが知られている。[4, 6]

これを追実験し、spam 対策に有効なことを確認した。

MX 遷移検査は以下のように行う。

1. 対象ドメインに複数の受信サーバと DNS MX レコードを用意する
2. primary MX では packet filter(ipfw, pf 等) で SMTP 接続を拒否する
3. secondary MX で接続状況を観察する

1 台のサーバにすべての MX の IP アドレスを割り当てておき、接続ログをみると、結果を比較観察しやすい。

2.1 複数 MX 設定実験

テストには筆者が管理する reflec.to ドメインを用いた。サーバ ns.reflec.to で運用中の 16 ドメイン のすべてに 2 つの MX レコードを登録し、primary サーバの SMTP ポート宛のパケットは ipfw で drop した。山口・鈴木 [4] は primary MX での接続拒否方法として TCP Reset を用いているが、遷移検査をするまでの時間的余裕をとるために、パケットをドロップ (接続に応答しない) する方法を試した。

2.2 遷移の観察結果

表 1 は各サーバへの接続ホスト (IP アドレス) 数である。

表 1: MX fallback (Nov 14 0:00 ～ Nov 18 17:15)

全 IP アドレス数	1329
1) primary で drop したもの	738
2) secondary に接続したもの	1020
3) 両方に接続したもの	429

(いずれも IP アドレス数)

2.3 遷移時間の観察

両方のサーバに接続した 429 の送信元 (表 1 の 第 3 項) を spam とそうでないものとに分離して、遷移時間を分析したものが表 2 と表 3 である。

受信サーバでは tempfailing と throttling による spam 対策を使っていて、受信したものを ham (非 spam)、 受信しなかったものを spam と分類した。

遷移は primary への複数回の接続の試み (SYN) に対する drop のあとに生じる。遷移時間は primary への最後の接続に対して、その後の最初の secondary への接続までの時間とした。

表 2: MX fallback (ham)

遷移時間	ホスト数	積算比率
1) 5 秒以内	25	13%
2) 10 秒以内	102	52%
3) 20 秒以内	154	79%
4) 30 秒以内	154	94%
5) 60 秒以内	192	98%
6) 90 秒以内	194	99%
非 spam ホスト	195	

(いずれも IP アドレス数)

表 3: MX fallback (spam)

遷移時間	ホスト数	積算比率
1) 5 秒以内	65	28%
1) 10 秒以内	104	45%
2) 20 秒以内	178	78%
2) 30 秒以内	186	81%
3) 60 秒以内	203	89%
4) 90 秒以内	208	91%
spam ホスト	229	

(いずれも IP アドレス数)

送信ホストの 99%が 90 秒以内に primary から secondary へ遷移している。そこで、90 秒以内を MX 遷移とみなすと、spam 送信ホストは (表 3 の 4) 208 個ある。これは **spam 送信ホスト全体数 1134 (=1329-195) の 18% (=208/1134)** にすぎない。

(60 秒を境界としても約 18%である)

jp.qmail.org で同様の実験を行なった結果では MX 遷移した spam ホストは全接続ホストの 10 %程度であった。

MX 遷移検査が 有効な spam 判定法であることが確認できた。

2.4 検査の実装法

MX 遷移 検査は以下のように実装できる (実装中)。

1. primary および secondary MX で SMTP を接続拒否
2. primary MX で drop を記録
3. 上記記録の IP アドレスから seondary MX への接続を許可する期限付ホワイトリストを動的に生成
4. secondary MX で 受信
5. 複数回受信した IP アドレスは primary MX で接続許可するよう恒常的ホワイトリストを作成／更新

2.5 運用の注意点

非 spam サイトに毎回 MX 遷移を強いることのないように、primary ではホワイトリストを使って受信する仕組みを導入すべきである。

MTA によっては primary MX に接続できないことをキャッシュして、secondary MX を送信先とするものがある。1 度 MX 遷移 検査をパスしたものについてはしばらく許可を続ける考慮があると良い。

山口･鈴木 [4] の報告や運用経験では MX 遷移しない MTA(主にアプライアンス製品) が存在することがわかっている。これらは RFC に従っていないので、その時々でホワイトリストに登録することで受信することを試みる。

2.6 Reset か Drop か

パケットフィルタのルール生成にかかる時間を考慮して今回は drop を評価した。secondary 側で SMTP の会話中に切断する実装を行なうなら、時間的な余裕を持たせられるので、TCP Reset も使えなくはない。Reset なら通常送信者には MX 遷移時間を短縮できるメリットがある。

しかし、spam に対して、SMTP の会話を行なうコストはそれよりも大きいだろう。drop ならば spam ホストに対して、余分の負荷をかけることもできる。実装の際にさらに考察を進めたいところである。

3 helo 検査

接続元の IP アドレスを DNS 逆引きすることは動的割りあて IP アドレスを使うゾンビ PC(以下 bots) を判別する有効な方法ではあるが、DNS に頼ることになる。しかし、一回限りの接続しかしてこない相手が多くて、DNS のキャッシュはあまり効果的ではない。逆引きの大規模な利用はネットワークに負荷となるので、できれば避けたい。

そこで、SMTP セッション内の情報を活用して送信元が bots かどうかを判定できるか helo コマンドを使うことを評価した。

3.1 SMTP ehlo/helo コマンド

SMTP[5] 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。

- TCP port 25 への接続を受けつけると、受信サーバは greeting と呼ばれ

る返答を送り、クライアントからの helo などを待つ。

- これを受けて、送信ホストは ehlo または helo コマンドを送る。ehlo は helo の拡張仕様である。

ehlo/helo コマンドのパラメタには送信ホストの完全修飾ドメイン名 (FQDN) をつけることになっている。このパラメタが正しく送信ホスト名を表わしていれば、DNS 逆引きをするまでもなく動的割りあて IP アドレスかどうか推定できる。

しかし、過去には間違った設定のホストが多く存在したために、通常のメイル受信サーバ (プログラム) は ehlo/helo パラメタを検査していない。最新の RFC でも「間違いを理由に受信拒否をしないように」書かれている。

spam ホストの多くも正しいホスト名を送ってこない。この状況を逆に利用して、spam ホストを判別する。

3.2 コマンドパラメタの検査

spam ホストが送ってくる ehlo/helo コマンドのパラメタを調べてみると以下のように分類される。この分類に応じて、対応方法を決める。

(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。理由は不明ながら、これらは spam であることが経験的に分っているので、受信拒否してよい。

(2) 逆引き名らしき名前 (FQDN):
例：`236-22-236.ded.tie.cl`
逆引き名であると想定して対応をきめる。動的割りあてアドレスを示す名前であれば、受信拒否あるいは一時エラー返答をする。

(3) 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, star-wars.`
設定ミスである可能性もあるが、多くは spam である。サイトのポリシーに従って対応を決める。

(4) ピリオドがひとつの名前:
例： `yahoo.com, hotmail.com`
ホスト名を含まないドメイン名らしきものも多くは spam である。DNS を利用して A レコードを検索し、送信ホストの IP アドレスと一致するかを確認するのもよい。

(5) その他のメイルサーバ風の名前など
例： a.mx.jp.qmail.org

3.3 コマンド別の試練

最新の SMTP 仕様 (RFC 2821 [5]) によるとセッションの開始時にはクライアントは ehlo を使うべきとある。ただし、互換性のために helo を送ってもよいことになっている。これらを判定材料に用いる。

3.3.1 helo ホストに対する試練

現在では helo を使うホストのほとんどが spam である。helo 送信ホストに対しては「一時エラー」を返答して、再送してくるかどうかを試す。

helo を使っていることが分っている通常ホストはホワイトリストに登録しておく。

3.3.2 ehlo ホストに対する試練

ehlo コマンドを送ってくるホストに対しては 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、helo を送り直してくるかを試す。

ehlo を前提にしている手抜きの送信プログラムを使う spam はセッションを切断するので排除できる。

ehlo のパラメタをみて、各種の対応を適用するかどうかを決めるのが妥当である。

3.4 実地試験と観測

3.4.1 試験環境

試験に使ったドメイン (jp.qmail.org) では現在はメイルの使用を停止しており、接続してくるものはすべて spam ホストである。

複数の MX を設定して、最優先サーバには接続をさせないで、次点のサーバに接続させている。

3.4.2 接続ホスト数と再接続の有無

数はそれぞれの振舞いをした IP アドレス数をしめす。再接続してくるホストが少いことが分る。単発接続は再送のない接続 (ホスト) を指す。

表 4: ehlo ホストの分析

ehlo を送ってきたもの	413
単発接続	324 (78%)
helo を送りなおしてきたもの	132
単発接続	79

表 5: helo ホストの分析

helo を送ってきたもの	2504
単発接続	2361 (94%)

3.4.3 helo パラメタの分類

2380 個の ホスト (IP アドレス) から送られてきた helo のパラメタを分類した。

1. サーバの IP アドレスまたはドメイン名 (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または最後にだけピリオドを持つ 136
4. その他 418

サーバの IP アドレスを送ってくるものの多くが DNS 逆引き設定されていない。このことからも DNS 逆引きに匹敵する情報が得られていることが分る。

3.5 ehlo/helo 判別の効用

spam ホストの推定が低コストで行なえる。一時エラー返答する場合には再送時に受信する候補を絞りこむのに使える。牛歩対応を使うのであれば、適用相手を選択するのに使えるので、システム資源の節約につながる。

helo パラメタがサーバの IP アドレスである spam は他の検査でも排除できることが分っている。それぞれの検査が spam 判定に有効であることの傍証であるが、一方で重複した検査は資源利用の上で無駄であるとも言える。

MX 遷移検査を通過したホストは再送してくる比率も高く、helo 検査を通過す

る比率も高いようである。MX 遷移検査を実装したシステムでの helo 検査の効果を検証する必要がある。

4 討論

4.1 spam 対策法の比較

今回の方法とこれまでの方法とを資源負荷と遅延の大きさで比べてみる [表 6]。

表 6: spam 対策比較

方法	負荷	遅延
tempfailing	小	大
throttling	大	中
DNS blacklist	中	?
DNS PTR	中	中
MX fallback	小	小
helo/ehlo	小	小

4.2 判定法の利用

spam であると判断できるものはその時点で拒否 (切断) するとしても、疑わしいという状況では spam 判別の目的 (ポリシー) によって、(途中および最終の) アクションが異なるのは当然である。

まずは疑わしい場合の抽出ができたとして、それに対してどういう反応をするかのポリシーについて検討する。

ポリシーをふたつの立場にわけてみる。

(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。単純に再送を待つのと、再送のパタンを判断材料にする方法がある。再送を行う spam が増えてきても、現状では多数回の再送を行う spam はまれであるので、一度の再送では受信せず、複数回の再送を行うかを監視する方法が考えられる。

(2) spam であると判断できなかったものはすべて受信する。

ブラックリスト作成は送信ホストを取りかえて送ってくる spam に対しては有効な手段とはいえず、今後はホワイトリストを充実する方向に進むであろう。

それならば、ホワイトリストを整備する方法を研究する方がより建設的である。

ホワイトリストがある程度整備された段階においては、(1) の立場をとるサイトが増えるであろう。

4.3 推奨案

1. 受信ドメインに対して複数の DNS MX レコードを設定する。2 MX あるいは 3 MX を推奨する。
2. MX 遷移検査で接続を制御する。
   - 最優先サーバではホワイトリストにあるホストだけを接続させる。
   - 次点のサーバで SMTP セッションを受けつけ、処理する。
   - 次点より優先度の低いサーバは囮である。動かさない。

4.4 誤判定などメイル受信のポリシーについて

世間には spam でないものを spam だと誤判定する危険を抱える方式があって、しかも、なぜ spam と判断したのかを説明できない (つまり spam の定義ができていない) ものが存在する。spam の溜りか

ら利用者が false positive を捜し出さなければならないとしたら、spam を分別する意義もうすれるであろう。

一時エラー返答や MX 遷移検査方式では受信しなかったものについてはポリシーにより明確に理由を示すことができる。

メイル受信者がポリシーを明確にし、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して自らが spam 送信者ではないことを明確に示すことが, 現在のメイルシステムを延命させる正しい対策であろう。

5 まとめ

spam 送信ホストを判別する手段を環境にやさしいかどうかの観点から評価して、複数 MX を設定し、MX fallback を検出する方法と helo コマンド情報を使った判別法の二つを評価、提案した。

メイル送信サーバが SMTP インターネット標準 (RFC) に従ってメイル送信するかを検査し、従っていないことの多い spam 送信ホストを判別するものである。

今回の手法により spam が疑われるホストを DNS 検索に頼らずに絞りこめ、システム資源やネットワーク資源に大きな負荷をかけることがないので、遅延返答や一時エラー返答などがより使い易くなる。これらが利用できる場面が拡大できれば、spam 受信が減らせるだろう。

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] 鈴木常彦・後藤邦夫・山口榮作・石川雅彦: MTA による spam 対策の実践報告, 情報処理学会研究報告, 2004-DSM-34, pp.61-64, 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] 山口榮作・鈴木常彦：TCP Handshake 制御を利用した spam 対策システム, 国公立大学センター情報システム研究会, 大学情報環境研究 Vol.8 pp. 60-67,2005

[5] SMTP RFC2821 http://www.ietf.org/rfc/rfc2821.txt

[6] Unlisting : http://www.joreybump.com/code/howto/unlisting.html

[7] 前野年紀: spam 対策の解説記事, http://moin.qmail.jp/spam

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

[9] 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).