spam SMTP接続を減らす手法

前野年紀

東京工業大学 原子炉工学研究所

Reduce SMTP connection from spam bots

Toshinori Maeno

Research Laboratory for Nuclear Reactors, Tokyo Institute of Technology

増加するspamメイルは受信せずにすませたい。あわせて受信拒否も軽快に行いたいという要求がある。spam送信ホストからのSMTP接続を減らす手法を実装して評価しているので現況を報告する。

SMTP接続に対して牛歩対応したり一時保留返答したりすることはspamの判別にとても有効ではあるが、受信動作中にはシステム資源が使用される。SMTP接続させる相手をできるかぎり少くするように、複数の受信サーバによるMX遷移検査に加えて、以下の手法を検討している。(1) DNS MXレコードの検索にTCPを使うように要求する。(2) greet pause時間の短縮をはかる。(3) SMTP接続のpacket OS fingerprintを判別に利用する。

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

1 はじめに

spamが増加して、spam対策にも軽量化が望まれるようになった。

spamかどうかにかかわらず送られてくるメイルはすべて受け取ってからフィルターで判別して捨てるということが行われてきたが、受信メイルの9割がspamという状況では資源負担に耐えられなくなってきている。

これまで筆者らはspamを受信しない手法を研究してきた。SMTP接続元がbotであるかを判別することにより受信保留する方法などである。受信保留は受信後に破棄するよりは各種資源の必要が少ないけれども、資源が使われる。spam判別にきわめて有効なDNS PTR情報を入手す

るにはネットワークに負荷がかかる。受信サーバでの牛歩対応は受信側システム資源の負担となる。これらの理由で大規模サーバでは採用しづらいこともある。

そこで、利用資源が少くてもすむspam送信ホスト判別法を検討してきた。すでにMX 遷移法やheloコマンドの利用などは報告した [5] が、今回はSMTP サーバへの接続そのものを減らす手法を検討している。現状を報告する。全体の構成は以下の通りである。第 2 節ではspam 判定法を資源利用面から概略の説明をする。第 3 節では今回検討したDNS TCP 接続要求と OS fingerprint の利用について説明する。第 4 節では今回の手法についての議論を行う。第 5 節は全体のまとめである。

2　これまでのspam対策

spam を受信しないために、メイル受信サーバは牛歩対応したり、保留返答をしたりすることが行われている [1, 3]。そして、spam 送信ホストの判別にはDNS PTR 情報が利用されている。これらは判別には有効な手法であるが、大量のspam メイルを扱う環境ではネットワークなどの負荷が大きくて、資源利用の面から望ましくない。

資源負荷を減らすためには対策手法を適用する場面を限定すべきであり、複数MX ホストを利用したMX 遷移検査法 [5] が分別に有効であった。

各種の対策の資源要求について反省してみる。

2.1　DNS 逆引きによる判別

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。その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。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ので、spam ホスト判定に使える [3]。

しかしながら、DNS 逆引き操作はDNS 問合せを多段に行うため、逆引きサーバやネットワーク全体の負荷となる。逆引きが設定されていない場合は検索回数も増える。また、ひとつのIP アドレスからは一回限りの接続が多いため、DNS キャッシュは効果が少い。DNS 設定の不良でネットワーク全体の負荷を増大させる。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 対策は勧められない。

代わりに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ことを提案した [5]。

2.2　牛歩戦術と必要な資源

牛歩対応はSMTP セッション中に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が増えることがシステム資源を占有するという問題をひきおこすので、短かいことが望ましい。接続元が接続をあきらめたことを早期に発見することが求められる。

2.3　一時保留返答

SMTP で受信保留返答をするものである。spam 送信プログラムは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に接続記録の保持 (30分から数時間) が必要となる。

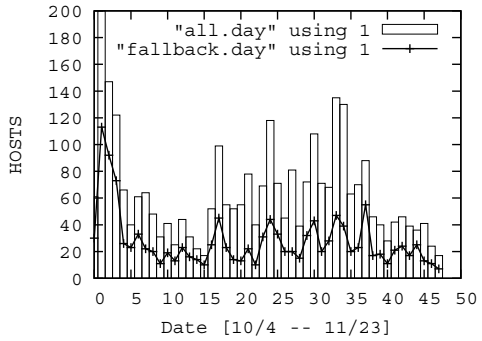

図 1: 全接続と MX 遷移/日別ホスト数

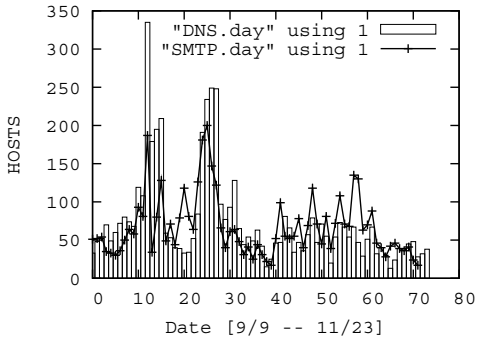

図 2: DNS 検索,SMTP 接続/日別ホスト数

この記録は大部分が使われないことで目的を果すものであるので、別途 spam だと判定できた接続は記録に残さないことにして、記録のための資源を節約する。

2.4 MX 遷移検査

spam 送信ホストが RFC の手順通りに複数の MX を順に試すということをしないことを利用した判別法である [5]。2 年前には有効であったが、最近では遷移する spam ホストが約 40%に増えていて、単独では十分な効果があるとは言えない (図 1)。

3 今回の検討案

spam ホストからの SMTP 接続を減らすべく、以下の手法を実装して評価中である。

(1)DNS MX レコードの検索要求 (UDP) に対しては DNS/TCP query を使うように要求する。

(2)greet pause 時間の短縮をはかる。

(3)SMTP 接続開始時の packet の OS fingerprint を調べて、OS 種別を接続判定材料とする。

3.1 DNS/TCP 検索要求

メイル送信の前段階として、送信側は受信サーバの IP アドレスを知るという作業がある。この作業として DNS MX レコードの検索を行うが、通常の DNS 検索には UDP が使われる。しかし、我々は DNS MX レコードの検索に TCP を使うように要求する[1]ことで、UDP だけで DNS 検索するような bot は排除できると考えた。spam 送信ホスト自身が MX 検索を行っているとしたら、DNS/TCP 接続は実装していそうもないから。

実際、UDP と TCP への DNS MX query 送信元ホストを集計、比較してみると、UDP を使っているホストのうち、約 25% は TCP を使ってこなかった。

また、これらの DNS 検索を行っているホストと spam を送ってくるホストの集合を比較したところ、DNS 検索ホストと SMTP 接続ホストは分離されているようであった。

[1]UDP で MX レコード検索要求を送ってくるホストに対して truncated bit を ON にした返答を返すなどして、TCP を使うようにうながす。

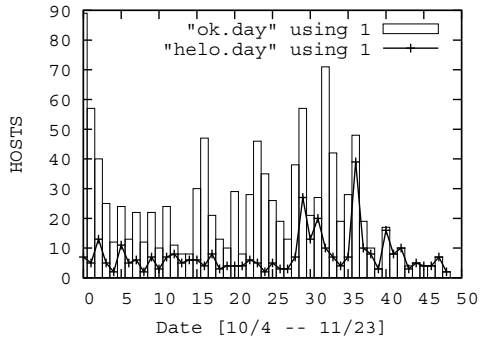

図 3: 接続と helo 送信/日別ホスト数

3.2 greet pause の再評価

greet pause 手法では SMTP 接続開始時にサーバから最初に送るメッセージを 5 秒から 30 秒程度遅延させてから送信するものである。この遅延により接続をあきらめる spam ホストがある。効果があることが報告されているが、遅延の間、受信サーバの資源が占有されるという短所もある。そのため効果が維持できる範囲で、遅延は短かいことが望ましい。

遅延時間を短縮していき、3 秒程度でも効果があることを確認した。図 3 は MX 遷移検査を通過して SMTP 接続させたホスト数と 3 秒の greet pause を通過して helo を送ってきたホスト数とを日別に数えたものである。接続したが helo を送ってこないもの、つまり greet pause 中に切断したホスト数は 3 秒の遅延の場合でも全体では約 75%をしめている。

最近まではかなりの効果を示していたことが分る。しかしながら、接続ホスト数の急減した 11 月半ば過ぎから効果がなくなっている。これは spam 業者が ISP から締め出されたというニュースと関係がありそうだ。

3.3 OS fingerprint の利用

spam 送信にどのような OS が使われているか、SMTP 接続開始時に送られてくる TCP packet OS fingerprint (os fp) を調べてみた。p0f というプログラムを使って 10 月 3 日から 11 月 22 日までの間に記録されたものを分析する。受信サーバは三台 (3 MX レコード) 用意してあり、MX1 に接続したあと、MX2 に接続してくる (MX 遷移) ものだけを受入れるように設定してある。

OS 種別毎の MX 接続と遷移の IP アドレス数である (表 1)。

表 1: MX 遷移と OS 種別ホスト数

OS	MX			遷移	全体
	1	2	3	1-2	
不明	962	963	978	543	1160
Windows	681	640	752	393	989
Linux	130	124	50	110	142
Solaris	6	6	180	5	184
FreeBSD	16	16	5	16	18
全体	1788	1747	1963	1059	2484

ホストの半数以上が分類された「不明」は fingerprint データベースに登録されていなかったため、判定できなかったものを示している。データベースを充実させれば、判明させられる。それ以外の OS はその名の通りであるが、判定の正確さは分らない。複数の fingerprint をもつ OS もある。

MX 遷移と OS 種別の関係があきらかなものは spam 判定に利用できる。現状は「Windows は接続させない。」という風な使い方をしている。筆者のメイル交換相手でメイルサーバに Windows を使っているケースはまれであるからだが、救済手段 (whitelist) も用意してあるので問題は起きていない。

これらの方法により選別した少数のホストだけをSMTP接続させるので、システム資源、ネットワーク資源に大きな負荷をかけることもなく、安心して他の手法を併用できている。

4 考察

spam排除に使われる資源を軽減するために、spamホストからの接続をできるかぎり減らし、接続してもできるだけ初期に切断する手法を検討した。複数MXを使ってのMX遷移検査もこの目的に沿うものであった。

今回の手法を順に考察する。

4.1 DNS/TCP 検索要求

DNS MXレコードの検索にTCP queryを使うように要求することで、UDP検索しかしないbot(約25 %)にはMX情報を渡さないですむ。TCP接続してきたホストには短期間に接続を繰返すものが目立つ。MX レコードの取り出しに失敗しているのかもしれない。

一日単位で統合してみると、一度だけの問合せのホストが多い。メイル送信のためのDNS MX問合せとしては不自然である。

SMTPホスト集団とDNS検索ホストとの間には直接の関係は見あたらない。botに指令を出しているマシンを介してつながっているからであろう。ここの関係を見つけることができれば、次の対策が考えられる。例えば、DNS/TCP接続を拒否するなどの対策である。

4.2 greet pause 時間の短縮

greet pause時間を縮めて、3秒でも効果があることを確認した。しかし、11月半ば現在は効果がない。効果を示す時間がいつどのように変化するのか、予想がつかない。監視をつづける必要がある。

4.3 OS fingerprint による判別

SMTP 接続開始時の packet OS fingerprint を調べて、Windows ホストは接続させないことにして、約半数のホストを排除できた。しかし、残る半数以上のホストがデータベースでは分類できていない。この機能を活用するには、OS fingerprintデータベースの充実が課題となる。

4.4 その他の判別法の併用

SMTP helo 情報を使う判別法は資源への負担は小さいので、いつでも併用できる。

選別した少数のホストだけをSMTP接続させることにするならば、システム資源、ネットワーク資源に大きな負荷をかける心配をせずにいろいろな手法が併用できる。DNS 逆引きの負荷も問題にしなくてすむ。

負荷が大きいという訳ではないが、併用したときに効果の大きいものとしては一時保留返答と再接続間隔の監視もある。

4.4.1 一時保留返答

helo検査などを通過した一見さんには一時保留返答をする。遅延が望ましくないというケースでも、休日や夜間に即時受信したいということはなかろう。24時間でも問題はないと考える。

一時保留返答に対する再接続があってもただちに受信処理をするのではなく、接続間隔のパタンをみて、接続させるかを判断する。spam には 6 分間隔で 6 回接続を繰返すものとか、11 分間隔で 2 回接続を繰返すものが目立つ。どちらも 15 分以内の再接続は再度保留にするという対応で排除できる。

5 おわりに

メイル受信サーバにかかる負荷を減らすために、spam 送信ホストからの SMTP 接続を減らす手法を検討した。

spam メイルが増加したせいで、spam 判別のためにも SMTP セッション中に受信サーバ側のシステム資源が使用される。このため spam ホストからの接続はできるかぎり少くし、接続してもできるだけ初期に切断する手法を検討した。

(1) DNS MX レコードの検索に TCP query を使って検索するように要求する。UDP 検索しかしない bot が排除できる。SMTP 接続ホストと DNS 検索ホストとの関連を調査して、今後活用する方法を探る。

(2) 接続時の greet pause 時間は 3 秒でも効果があることを確認した。

(3) 接続開始時の packet OS fingerprint を利用して選別することで、約半数のホストを排除できた。fingerprint データベースの充実が課題である。

選別された少数のホストだけを SMTP 接続させることにすれば、その後の処理に使われるシステム資源、ネットワーク資源を軽減できる。DNS 逆引きの負荷すら問題にしなくてすむ。一時保留返答の記録も対象が少くなり、使い易くなる。

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] 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] 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).

[6] 前野年紀: spam 解説記事, http://moin.qmail.jp/spam

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

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

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