LL-003

メールゲートウェイにおけるバウンスメール発生の抑制法とその評価
An evaluation of the suppressing method for suspicious bounce mails on the mail gateway

桝田 秀夫 †
Hideo Masuda

落合 優 ‡
Yu Ochiai

1. はじめに

近年，迷惑メール (spam) が急増しており，管理者がその対応に追われるようになってきた．迷惑メールの送られ方にはさまざまな手口が用いられているが，代表的なものとして，Web などからメールアドレスと思われる文字列を抽出したもの，ランダム文字列や辞書上の単語をユーザ名としてメールを大量に生成するハーベスト攻撃など，大量のメールアドレスに対して無作為に送信を行うパターンがある．このような大量のメールアドレスは，全てが有効であるとは限らないので，宛先不明となり受信拒否の通知 (バウンスメール) が大量に発生してしまう．迷惑メールでは，差出人情報も詐称されていることが多く，バウンスメールがさらにバウンスメールを引き起こす場合や，実際の差出人ではないユーザに誤ったバウンスメールを返してしまうといった問題が顕在化してきている [3]．正規の利用において，タイプミスなどによる宛先不明メールにはバウンスメールを届ける必要があるので，バウンスメールの生成自体を抑制することは難しい．

一方，大学などではウィルス調査などを大学内で一括したメールゲートウェイ上で実施した上で，調査済みのメールを部局 (サブドメイン) 毎のメールサーバに配信する体制が多い．この構成では，メールゲートウェイ (MGW) 上でサブドメインのメールアドレスが存在するのか確認が容易ではないため，迷惑メールに対する不要なバウンスメールが多発する傾向にあり問題となっている．この問題を解決するために，MGW 上で各サブドメインのメールアドレス情報を集中管理する方式が提案されている [4]．しかしこの方式は，各サブドメインの管理者が自ら管理するメールアドレス情報を整合性を保つように登録する必要があり，管理の複雑さが増すという問題がある．

本論文では，メールアドレス情報を集中管理することなく，不必要なバウンスメールを発生させない方式の効果とその方式の性能の評価を行う．

2. バウンスメール抑制法

2.1 メールゲートウェイ型の構成の問題点

通常，バウンスメールはメール配送セッションの最中に，宛先不明のような再送しても解決することが不可能である問題 (相手のサーバが 5xx のステータスを返す) が判明した際に，クライアント側が生成し (特に指定が無ければ) 差出人アドレスに向けて送信する．従って，メールスプールを持つサーバにハーベスト攻撃元から直接セッションが張られる場合は，配送セッション中に宛先アドレスが存在しないことが判明するので，自組織のメールサーバがバウンスメールを生成することはない．しかし，図 1 のようなシステム構成の場合，MGW は外部からメールを受信する際に，ドメインパートさえ一致すれば，実際には存在しないメールアドレスであってもメールを受け取ってしまう．一旦受け取ってしまったメールは，対応するサブドメインのメールサーバに対してメールを送信する際に宛先不明が判明するので，MGW 上でバウンスメールを生成することになる．

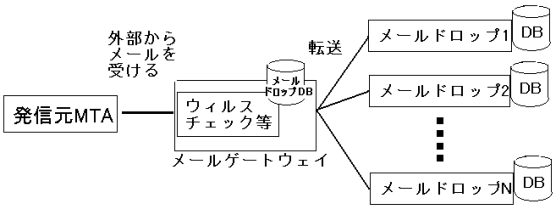

図 1: 想定するメールゲートウェイを配置する構成

2.2 本研究のアプローチ

2.1 節のような迷惑メールに対するバウンスメールを抑制するためには，MGW が外部の MTA とのセッションの最中に宛先アドレスの有効性を判断できれば良い．

本論文では，外部 MTA からのセッション中の受信者アドレスの通知の段階で，ドメインパートのみをチェックして応答を返すのではなく，サブドメインのメールサーバに対して宛先アドレスが存在するかの検証を実施してから，その検証結果に応じて応答を行う方法を採用する．

RFC2821[1] によれば，セッションの最中において相手の応答を少なくとも 5 分は待つべきであるとされており，この検証にかかる時間がこれを越えないことも必要となる．また，サブドメインにメールサーバに対して宛先アドレスの存在の検証は，負荷をかける行為となるので，できるだけ回数が少ないことが望まれる．

そこで，本論文では，同一の宛先アドレスの検証をなんども行う必要がないように，検証結果をキャッシュする方法も検討する．

2.3 宛先アドレスキャッシュ機能の特徴と問題点

宛先アドレスの検証結果をキャッシュする機能を導入すれば，検証回数の削減が可能であるが，逆にサブドメイン側での有効宛先アドレスの増減に対してリアルタイムに対応ができない可能性が生じる．

宛先アドレスのキャッシュには，宛先アドレスが有効であった場合のアドレス情報 (ポジティヴキャッシュ) と宛先アドレスが無効であった場合のアドレス情報 (ネガティヴキャッシュ) がある．

†京都工芸繊維大学情報科学センター
‡京都工芸繊維大学工芸科学部電子情報工学科

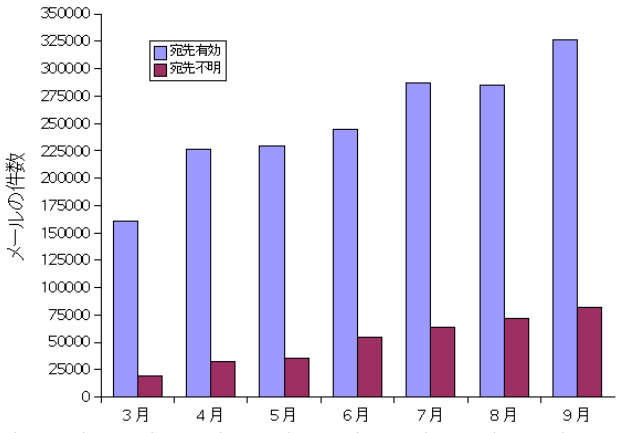

図 2: 受信メールの内訳

ポジティヴキャッシュでは，キャッシュの有効期間内にそのメールアドレスが無効になる場合に問題が生じ，宛先が有効であると誤認して MGW がメールを受信した後に，バウンスメールが発生してしまう．ネガティヴキャッシュでは，キャッシュの有効期間内にそのメールアドレスが有効になる場合に問題が生じ，実際には有効な宛先を無効と誤認して MGW がメールの受信を拒否してしまう．

これらの問題点とキャッシュの有効期間はトレードオフがあるため，キャッシュ機能のパラメータ調整が重要となる．

3. シミュレーション評価

本学で実際に稼働している MGW のログデータを解析し，本アプローチの有効性の評価を行う．

3.1 ログ取得環境

本学のシステム [5] では，MGW 上でアンチウィルスゲートウェイが導入されており，到着したメールに対してウィルスチェックをしたのち，ローカルで稼働している sendmail に渡し，その sendmail が必要なサブドメインにメールを配送する構成となっている．

解析に使用したログデータは，この sendmail が出力したログデータのうち，2006 年 3 月 5 日から 2006 年 10 月 1 日までのものを使用した．

3.2 宛先不明メールの件数

まず，ログデータから sendmail のログ解析ツールである fromto[6] を用いて受信日時，宛先アドレス，差出人アドレスを抽出し，受信できたメールと宛先不明であったメールの総数と調べた．図 2 は，正常に受信できたメール数，及び，宛先不明であったメール数の推移を表している．

これにより，全体のうち宛先不明であったメールの割合は 5 月末までにおいて，約 12 %，それ以降で約 19 %であった．メールの件数では, 期間内で正常に受信できたメール数 176 万 2750 通に対して，宛先不明メールは 35 万 9029 通 (約 17 %) であった．

また，送信されたバウンスメールのうち宛先が存在したものは 13 万 6741 通で，全体の約 38 %であった．つまり，少なくとも残りの約 62 %は差出人が詐称されていると考えられる．配送が出来た約 38 %のバウンスメールも詐称された差出人アドレスが存在する場合を含んでいるので，実際に必要であったバウンスメールの数は更に少ないと考えられる．

3.3 ポジティヴキャッシュの効果の推定

図 3 より，ポジティヴキャッシュはキャッシュの有効期間 1 日で全体の検証回数の約 28 %，宛先有効アドレスに関する検証回数だけでみると約 34 %を減らせることが分かった．

また，検証を行う頻度については，解析したログデータの期間中，最もメールの到着が多かった時間帯では 10 分間に 1280 通の到着が記録されていた．このとき，ポジティヴキャッシュを導入すれば，キャッシュの有効期間が 1 日である場合，10 分間に 392 回まで下げることが可能となる．

さらに，ポジティヴキャッシュを導入することでどれだけ問題 (検証すれば防げたバウンスメールの発生) が生じるかをシミュレートを行った．図 3 より，ポジティヴキャッシュは有効期間 1 日でおよそ 200 通程度のバウンスメールが発生し，有効期間が増えると比例的にバウンスメールが増加している．しかし，およそ半年間のバウンスメールは約 36 万通であったので，ポジティブキャッシュを導入したとしても想定されるバウンスメールは抑制可能であると考えられる．

3.4 ネガティヴキャッシュの効果推定

図 4 より，ネガティヴキャッシュはキャッシュの有効期限が 1 日でも，全体の検証回数を約 13 %，宛先不明アドレスに関する検証回数だけでみると約 76 %を減らせることが分かった．またキャッシュの有効期限を延ばしてもそれほど検証回数が削減できないことから，宛先不明アドレスに向けられるメールについて，同一アドレスを使用するメールは短期間にしか集中しない傾向にあることが分かる．

次に，ネガティヴキャッシュを導入することで，どれだけ問題 (有効アドレスに向けられたメールの破棄) が生じるかシミュレートを行った．図 4 より，ネガティヴキャッシュの有効期間が 2 時間でも，有効アドレスに向けられたメールを破棄してしまう場合があることが判った．従って，例えば運用として「サブドメインで有効にしたメールアドレスは，有効にしてから 2 時間は届かない可能性がある」ということを周知したり，キャッシュ情報を手動で操作できるような構成にする必要があると考えられる．

4. 負荷実験と評価

本論文でのアプローチを，システムとして構築し，負荷や遅延について検証する．

4.1 実験方法

実験には表 1 のマシンを 3 台使用した．負荷試験には，postfix に付属する smtp-source をベースに，以下の時刻を gettimeofday 関数で記録するように改造したものを使用した．

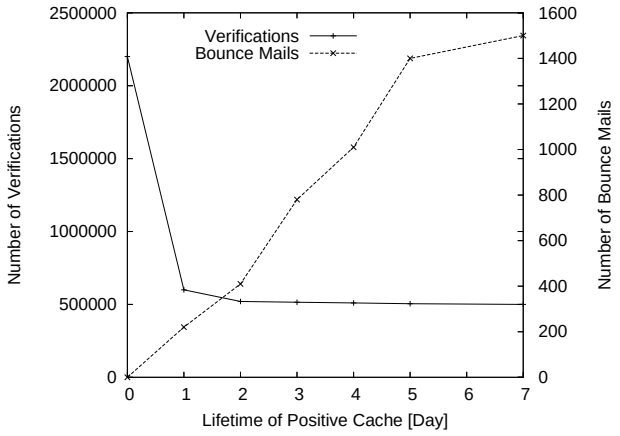

図 3: ポジティヴキャッシュの有効期間と検証回数，バウンスメール発生数

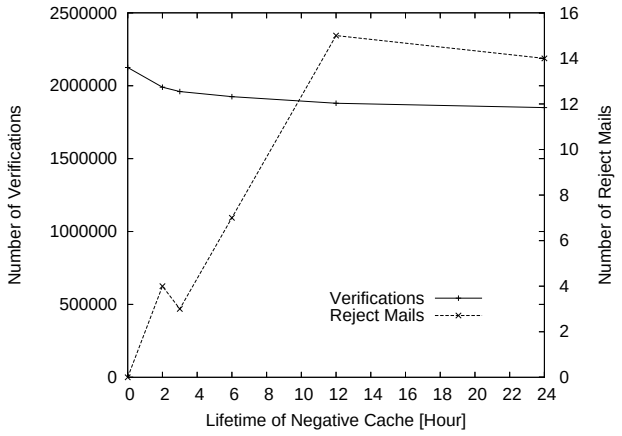

図 4: ネガティヴキャッシュの有効期間と検証回数，誤拒否件数

T_1 ：コネクションの開始の時刻

T_2 ：セッションの開始の時刻

T_3 ：RCPT TO コマンドの送信時刻

T_4 ：RCPT TO コマンドの応答時刻

T_5 ：メール本文の送信完了時刻

計測には，同時に張るセッションの数を 1, 10, 20, 30, 50, 100, 150, 200 と変化させ，それぞれ 10 回の平均値とした．

4.2 計測結果

まず T_1〜T_5 の記録から各区間 ($P_i = T_{i+1} - T_i(i = \{1, 2, 3, 4\})$) にかかった時間についてセッション数ごとに着目する．最初に各方式ごとにセッション数とかかった時間を比較すると図 6,7,8,9 のようになった．

5. 考察

postfix の verify サーバを用いた実験では，アドレス検証にかかる時間 (P2) により，メール 1 通当たりの処理時間が 3 秒程度増加することが判った．しかし, 同時セッション数を増やすと，検証の結果を待っている時間を他のメールの処理にあてることが可能になるので，スループットは向上する．同時セッション数 150 において，スループットは verify を使用しない場合の 1/3 まで向上した．しかし，同時セッション数 200 において，全てのメールでアドレス検証が必要な場合では，一部のメールで一時エラーが返ることになった．

本学におけるメールサーバのログ解析結果から，10 分間で最大でも 1280 通のメールを受信するということであった．これは，ウィルスチェックを通ったメール数であるので実際はこの数値よりも高いことが考えられる．しかし，今回実験に使用した環境では，図 5 より全てのアドレス検証が必要な場合であっても最大で 10 分間に約 9000 通のメールの処理が可能ということになり，十分実用に耐え得ると考えられる．キャッシュを導入した場合はさらに高いスループットが出ることから，このシステムは充分実用に耐え得ると考えられる．

CPU	PentiumIII 933MHz
HDD	30GByte(5400rpm)
Memory	512MByte
NIC	Intel Pro/100
OS	Vine Linux 3.2
MTA	Postfix 2.3.5

表 1: 実験に使用したマシンのスペック

6. まとめ

本論文では，メールゲートウェイ型構成の際に発生する不必要なバウンスメールの発生を抑制するため，MGW が外部からの SMTP セッションの最中に宛先アドレス検証を行って宛先不明であるメールを拒否する方式の効

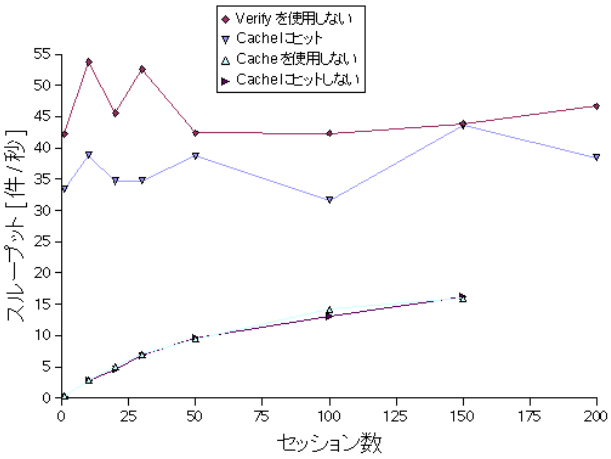

図 5: キャッシュの有無によるスループットの変化

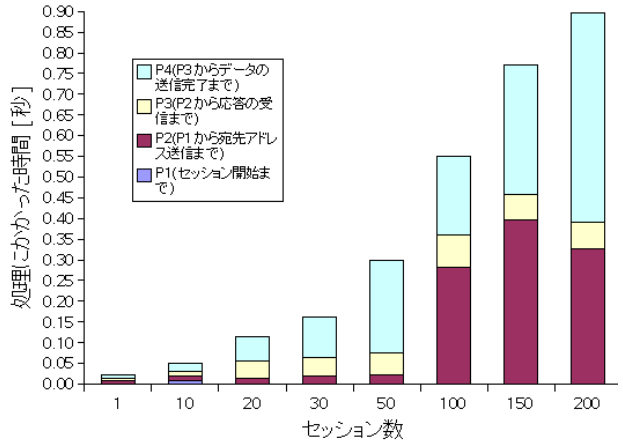

図 6: verify サーバを使用しない場合の比較

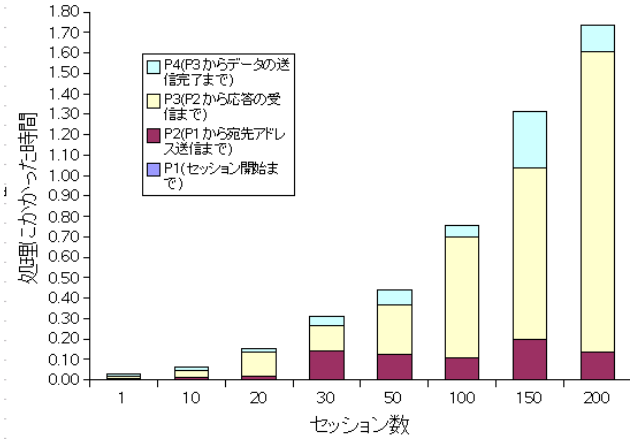

図 7: 全てキャッシュにヒットする場合の比較

果を，実際のログデータから推定した．また，キャッシュ機能の効果及び発生する問題の程度を確認した．さらに，postfix に付属する verify サーバ機能を用いて環境を構築し，実際に導入した際のオーバヘッドの測定を行い，本学規模の組織であれば十分実用に耐え得ることを示し，システムの有効性を確認した．

今後の課題としては，実際に稼働しているシステムに適用して，想定される効果が得られるかを確認することや，キャッシュのメンテナンス方法の検討が考えられる．

参考文献

[1] Klensin,J.(ed.):“Simple Mail Transfer Protocol”,RFC2821,IETF (2001).

[2] Resnick,P.:“Internet Message Format”,RFC2822 (2001).

[3] 山井成良, 繁田展史, 岡山聖彦, 宮下卓也, 丸山伸, 中村素典: “発信者詐称 spam メールに起因するエラーメール集中への対策手法”, 第 3 回情報科学技術フォーラム情報技術レターズ,pp313-316 (2004).

[4] 吉田和幸, 矢田哲二, 原山博文, 伊藤哲郎: “spam メール対策と統合メール管理システムについて”, 情報処理学会論文誌,Vol46,No.4,pp1035-1040 (2005).

[5] 桝田秀夫, 平田博章, 黒江康明, 柴山潔: “京都工芸繊維大学における新情報教育システムについて”, 2006 年 PC カンファレンス,pp.173-176 (2006).

[6] 歌代和正: “fromto” ,ftp://ftp.sra.co.jp/pub/lang/perl/sra-scripts/fromto-1.5 .

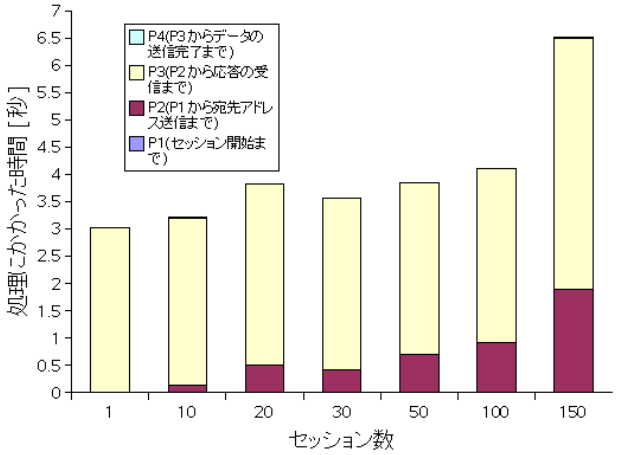

図 8: キャッシュを使用しない場合の比較

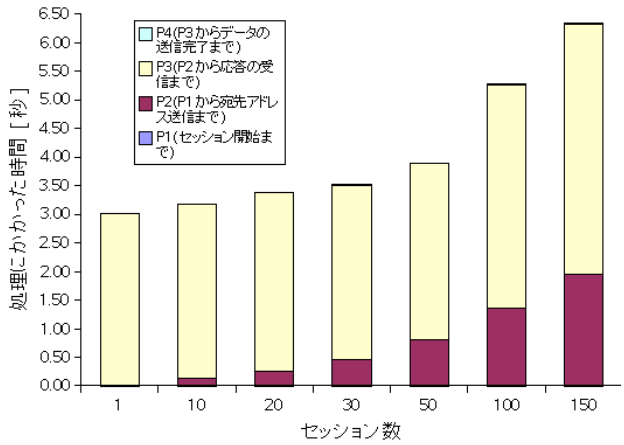

図 9: キャッシュにヒットしない場合の比較

謝辞

本研究は一部，日本学術振興会・科学研究費補助金・基盤研究 (C)（課題番号 19500069）の研究助成による．