Vol.2010-IOT-8 No.14
2010/3/1

spam 対策用 whitelist を一元管理できるメールシステムとその運用について

飯田隆義[†]　松竹俊和[††]　吉田和幸[†††]

spam 対策として，メール受信時にゆっくり応答をする throttling や，一時エラーにより再送をうながす greylisting といった相手のメールサーバを検査する spam 対策がよく利用される．一部のメールマガジン送信者等は送信プロトコル(SMTP)を守らないことがあり，throttling, greylisting 等により誤検知されることがあるため，whitelist の利用が欠かせない．spam 検出の精度向上のため，複数の spam 対策を組み合わせて実施すると，各対策で利用する whitelist の内容は共通であっても，各設定ファイルでの文法が異なるため，個々に whitelist の管理が必要となる．しかし，登録件数が多い whitelist を異なる文法で管理することは，誤りを起こしやすく正常なメールの受信拒否につながりかねない．そこで，我々は whitelist を一元管理するために，whitelist によってメールの振り分けを行う分別装置を利用したメールシステムを構築した．本論分では構築したメールシステムとその運用経験について述べる．

Consolidating the Management of Whitelist for Spam Mail Control and Its Operation

Takayoshi Iida[†]　Toshikazu Matsutake[††]
and Kazuyuki Yoshida[†††]

As a measure against spam, "throttling" which answers slowly at the time of e-mail reception, and "greylisting" which causes an error temporarily and is made to force into resending are used well. Since some mail magazine sender do not follow SMTP exactly, it may be detected as a spam incorrectly by throttling and greylisting. Therefore, whitelist of mail servers is indispensable. In order to raise the accuracy of spam detection, when some measures against spam are implemented, whitelist is needed in each measure. Although the contents of whitelists are the same, since the each descriptive grammars of whitelists differ in other measures, whitelists are need to manage separately. However, to manage two or more whitelists with many entries may be a cause of an error. In this paper, we describe the system which is able to manage whitelists uniformly. We also describe the operational experience of the system.

1. はじめに

近年，インターネットの急速な発展と普及に伴い，電子メールを初めとするネットワークを介したコミュニケーションは不可欠な物となっている．これに伴い spam が大きな社会問題となっている．spam とは受信者の意図を無視して無差別かつ大量に一括して送信される電子メールを指し，UCE (Unsolicited Commercial E -mail), UBE (Unsolicited B ulk E -mail)とも呼ばれる．電子メールは通常の郵便と比べると，送信者側が容易にメールを多くの相手に対して送信でき，送信者側の負担が金銭的にも時間的にも労力的にも極めて少ないといった特徴が挙げられる．

現在，大分大学学術情報拠点情報基盤センターでは，ウィルスを検知・除去するためのメールゲートウェイを導入し，学内 LAN とインターネットとの間を行き来するメールについてウィルスの有無の検査と同時に spam 対策も行なっている [1) 2) 3)]．

いろいろな spam 対策を組み合わせて利用する際にそれらを適用する順序によって，spam 検出時に発生するエラーが異なるため，適用順を考慮する必要がある．適用順を決定するにあたり，以下の 2 つの用件を設定した．(a)「User unknown」といった，メールアドレスの有無に関するエラーは，なるべく発生させないようにする．(b)コンテンツフィルタリングのように CPU パワーを必要とするものはなるべく後回しにする．また throttling，送信元 MTA(Mail Transfer Agent)の IP アドレスの検査，各ヘッダの検査等，実施するタイミングが固定されているものもある．そのため，現在のところ送られてきた電子メールに対して，以下の順序で spam メール対策をしている．使用している MTA は sendmail[4)]バージョン 8.14 である．

(1) throttling (greet_pause) [5)]
(2)外部の Blocking List を用いた送信メールサーバの IP アドレスの検査 [3)]
(3)メールヘッダの形式検査 [3)]
(4)LDAP[6)]を利用した学内各メールサーバのユーザアカウントの有無の検査 [2) 3)]
(5)greylisting[7)]による送信メールサーバの検査 [8)]
(6)spamassassin[9)]によるメール内容の検査 [10)]

sendmail の greet_pause 機能と greylisting は，spam 送信サーバが，大量のメールを送ろうとするため，通常のメールサーバとは異なった動作をすることに注目して，spam 検出を行なおうとするものである．greylisting は，一旦一時エラーを送って，再送を待つ方式である．このため，再送されるまで 30 分程度配送に余分な時間がかかる．

† 大分大学大学院工学研究科
Department of Computer Science and Intelligent System, Oita University
†† 大分大学工学部
Department of Computer Science and Intelligent System, Oita University
††† 大分大学学術情報拠点情報基盤センター
Center for Academic Information and Library Services, Oita University

さらに，再送されたメールであることを確認するために，送信元メールサーバの IP アドレス，送受信メールアドレス，時刻のデータベースを作成する必要があり，そのデータベースのためのメモリ領域を必要とする．このデータベースの保持期間等，設定すべきパラメータが多い．一方，greet_pause では，最初の応答まで待つ時間は，今のところ 60 秒までで十分効果があがっている．設定すべきパラメータは，待ち時間のみである．ただし，TCP コネクションを保ったまま待つので，sendmail のプロセス数，TCP セッション数が増えやすい．そのため，プロセス数があらかじめ決めた上限に達し，通常のメールの配送に影響が出ることもある．

これらの問題を軽減するために whitelist を利用することが多い．whitelist とは無条件にメールの受信を行うことを許可した MTA のリストのことである．ただし，本大学で運用中である greet_pause と greylisting では whitelist の記述方法が異なるため，2 つの whitelist を異なる文法で維持・管理しているといった現状がある．

そこで本論文では，whitelist によってメールを振り分けるメール分別装置 [11) 12) 13)]を利用した，whitelist の一元管理方法とその運用結果について報告する．

本論文の構成は以下の通り．まず，2 章で throttling と greylisting について論じ，3 章で whitelist の必要性について述べる．4 章で whitelist の問題点と解決策ついて述べ，5 章でメール分別システムについて述べる．6 章で実際の運用結果について述べた後，最後の 7 章で結論を述べる．

2. throttlingとgreylisting

spam が近い将来に無くなることは無いと考えられるため，spam の被害をいかに防ぐかは受信側での対策が重要となる．spam 対策で重要なのは spam メールを受け取らないことではなく，spam でないメールを確実に受け取ることである．さまざまな spam メール対策方法が考案されているが，spam の検出漏れ(false negative rate)よりも，通常のメールの誤検知(false positive rate)という観点で評価すべきである．多少の見逃した spam メールは単に削除すればよいだけだが，重要なメールが spam と判定されるとその影響は大きい．また，spam メールが大量に送られてきた場合，対策手法の計算機にかかる負荷が多いと，メール配送に遅延が発生するだけでなく MTA の運用自体が困難になってしまうことも起こりうる．

現在の主な spam 対策手法として，spamassassin[9)]等のコンテンツフィルタリングがよく利用されている．コンテンツフィルタリングはメールの内容からフィルタリングを行う．そのため，CPU 負荷は他の手法と比べて大きい．そして，spam メールの手口が多様化していくにつれて，フィルタを行うためのルールも肥大化する傾向にあり，どうしても誤検知(false positive)の問題がつきまとう．そこで，greylisting や throttling といった手法を用いることにより，メールの中身を見ずに spam 送信者かどうかという観点で判定することで，受け取る spam の数を減らすことが考えられた．これらは spam を見分ける効果(false negative rate) も高く，一般的な MTA からのメールを失う可能性(false positive rate) も低いとして現在注目されている．

greylisting と throttling は，spam メールを送り出す専用ツールの挙動に注目し，一般の MTA との違いから spam の選別を行う手法である．以下 greylisting と throttling についてさらに詳しく述べる．

2.1 greylisting

greylisting は「spam 発信 MTA は再送をしない」との仮説に基づく対策手法であり，一時的に受信を拒否し，再送されれば受信するといった動作を行う．殆どの spam 発信 MTA は仮説通りに動作し，高い効果を挙げている．ただし，この方法は配送遅延が大きく，場合によっては 1 時間以上かかるときもある．さらに，再送されたメールであることを確認するために，MTA の IP アドレス，送受信メールアドレス，時刻のデータベースを維持する必要がある．そのため，データベースのためのメモリ領域を必要とする．また，正当なメールの送信元 MTA にも再送を強いる点や，spam 送信者でない一部の MTA に再送しないものも存在するため，whitelist の管理が必要となる．

2.2 throttling

throttling は「spam 発信 MTA は timeout が短い」，「spam 発信 MTA は SMTP の確認応答手順を無視してメールを送る」との仮説に基づく対策手法である．具体的には，コネクション確立後の応答を遅延することで，spam 発信者の MTA がこちらの応答を無視してメールを配送してくるか，メールの配信をあきらめて接続を切断することを期待するものである．設定すべきパラメータは遅延時間のみであり，設定が簡単である．また，再送かどうかの判定が不要なので greylisting より適用範囲が広く，greylisting と比べると配送遅延の時間が数十秒と非常に小さい．ただし，throttling では拒否できないが，greylisting では拒否できるものがあり，対策としてどちらか一方に集約できるものではない．throttling は TCP コネクションを保ったまま待つため，プロセス数，TCP セッション数は増えやすいといった問題もある．MTA によっては，プロセス数の上限などを考慮の上，IP アドレスの逆引き，DNS Black List 等のブラックリストサービス [14) 15)]等を利用して遅延時間を調整することで，通常のメールになるべく影響が出ないようにしている．

3. 想定外の動作をするメールサーバとwhitelistの作成

全てのメールに throttling および greylisting を適用すると，遅延や再送が必要になり，メールの受信までに時間がかかることになる．spam の可能性があるメールは，なるべくゆっくり受信し，spam でないと確信がもてるメールはすぐに受信したい．そのため，信用できるメールサーバの IP アドレスを列挙し，その信用できるメールサーバから来

るメールに関しては，throttling 処理および greylisting 処理をスキップする．そうすることで多くのメールをほとんど遅れ無しに受信できるようになる．

さらに，whitelist は throttling および greylisting で誤検知されてしまうような，規定の動作をしないメールサーバからのメール受信を許可する．以下にその動作例を示す．

(1) **確認応答を待たずにメールを送る(throttling)**

一部のメールマガジン送信者等，大量のメールを送信する際に，SMTP プロトコルを守らずメールを送ってくるメールサーバが存在する．メール配信の挙動が spam メールとほぼ同じであるため誤検知されてしまう．

(2) **再送処理を行わない(greylisting)**

ウィルス検査のためのメールゲートウェイと sendmail のような MTA との組み合わせ方によっては，再送処理をしなくなる．マニュアルにはそのような設定例が載っている．

(3) **再送するたびに MTA が変わる(greylisting)**

大手 ISP の中には，大量のメールを処理するため，複数のメールサーバを持ち，再送のたびに，異なったメールサーバから送ってくる ISP がある．数回再送されるうちに，偶然最初のメールサーバと同じサーバから再送されると受信できるが，それまでに相当な時間を要する場合がある．

(4) **再送するたびに送信者アドレスを変更するメーリングリストサーバがある(greylisting)**

spam メール対策であると思われるが，再送のたびに送信者のメールアドレスを変える ISP がある．この場合，こちら側が再送されたことを確認するために保持しているメールアドレスといつまでたっても一致いないため全く受信できない．

以上の問題を回避するためにも，信用できるメールサーバの whitelist の作成は，重要である．現在，1500 行ほどの whitelist を管理しており，上記(2)をカバーするためなどに/24，/16 のネットワークを丸々登録している場合もある．そのため，信用するメールサーバの数は相当多くなっている．

4. whitelist管理における問題点と解決策

3 章で述べたように，whitelist は throttling および greylisting を行う際には，必須の対策となる．ただし，whitelist の管理には 1 つの問題点がある．

4.1 whitelist管理における問題点

本大学で運用中の spam 対策の中で，whitelist を利用している対策は throttling および greylisting である．この 2 つの対策で利用している whitelist の内容はほぼ共通であるにも関わらず，それぞれの設定ファイルでの記述方法（文法）が異なるために，別々に whitelist の管理が必要となってしまっている．エントリーが 1500 を超える whitelist を二通りの異なる書き方で維持・管理することは，誤りや設定漏れの危険をいたずらに増やすだけであり，それによって正常なメールの受信拒否につながりかねない．

4.2 解決策

上記のような理由により，whitelist は同じ記述方法で一元管理できることが望ましいと考えられる．そこで，whitelist によってメールを振り分けるメール分別装置と，2 台の MTA を用いた whitelist を一元管理するシステムを設計し運用することで問題の解決を図る．この管理方法では，本大学で利用している MTA である sendmail だけでなく，他の MTA を利用している場合でも汎用的に利用可能な管理方法である．また，2 台の MTA を用意せず，1 台の MTA だけで whitelist を一元管理するシステムも構築し実験的に運用を行ったので，その構成と運用についても 5 章, 6 章で詳しく述べる．

5. メール分別装置

まず事前準備として，複数の文法によって管理されている whitelist を 1 つの形式にまとめる．本提案では greylisting によって利用されている文法に従い whitelist をまとめている(図 1).

```
acl whitelist addr 202.23.64.0/24
acl whitelist addr 202.23.130.48
```

図 1. whitelist の登録例

5.1 2 台のMTAと分別装置によるwhitelistの一元管理（IPアドレスベース）

2 台の MTA を利用した whitelist の一元管理方法について述べる．この方法では，事前準備によりまとめられた whitelist を分別装置に取り込み，インターネットから受信するメールの振り分け判定として利用することで一元管理を実現する．振り分け方は，ソース IP アドレスが whitelist に一致せず spam である可能性が高いメールを MTA1 へ配送し，ソース IP アドレスが whitelist に一致し spam である可能性が低いメールを MTA2 へと配送する．すると，MTA1 には whitelist 以外の spam 対策を施す必要があるメールのみ配送されてくる．そのため，MTA1 に関しては無条件で throttling および greylisting 等，各 spam 対策を行えばよい．MTA2 へは whitelist に一致した送信元からのメールしか配送されないため，一切の spam 対策は必要ないことになる．ただし，転送処理によって spam メールが送られてくる可能性が少なからずあるため，コンテンツフィルタリングは適用する必要性がある．ここで 2 つの MTA および分別装置の構成図は図 2 のようになる．このような構成にすることによって，whitelist が分別装置内部で適用され，whitelist が構成全体において 1 つで済むようになる．

また，分別装置はファイアウォール(FW)と 2 つの MTA の間に配置し，NAT[16] サーバのように動作する．spam である可能性の高いメールは MTA1 で処理し，spam でないと思われるメールは MTA2 へ配送する．その振り分け方法としてレイヤ 3 レベル(IP

アドレス)を用いる. ある電子メール(SMTP パケット)の送信元 MTA が whitelist に含まれていた場合は，そのディスティネーションアドレスを MTA2 へ変更する.

分別装置のより具体的な動作について説明する．DNS の MX レコードには MTA1 の IP アドレスを設定している．分別装置は以下のように動作を行う.

① インターネット側から MTA1 の TCP ポート 25 番宛ての SMTP パケットが届く.
② ソース IP アドレスを抽出する.
③ ソース IP が,
A whitelist に登録されている場合，ディスティネーション IP アドレスを MTA2 に置き換える.
B それ以外の場合は，何もしない
④ IP ヘッダ及び TCP ヘッダのチェックサムを再計算する.

なお，MTA1, MTA2 からの応答パケットに対して分別装置は以下のような動作をする.

① MTA1,2 からソースポート 25 番の SMTP 通信の応答パケットが届く
② ソース IP アドレスを無条件に MTA1 に変換する
③ IP ヘッダ及び TCP ヘッダのチェックサムを再計算する

MTA1, 2 が送信元になる SMTP 通信等上記以外のパケットに関して分別装置は何もしない.

5.2 1 台のMTAと分別装置によるwhitelistの一元管理（ポートベース）

1 台の MTA による whitelist の一元管理方法について述べる．この方法では，5.1 節で説明した IP アドレスでの振り分け方法と大きな違いはないが，5.1 節でデスティネーション IP アドレスを置き換えていたところを, ディスティネーションポートの置き換えに変更し, 2 台の MTA へ振り分けるのではなく, 1 台の MTA の中にある別々の MTA プロセスへポート番号を変えることで振り分ける点が異なる．構成図は以下のようになる(図 3).

6. 運用実験

まず，2 台の MTA による運用結果について述べる．2009 年 5 月 24 日から 2010 年 1 月 23 日まで，各 MTA において「各対策手法での spam 検知数」の解析を行った.

この期間で受け取ったメールは平均で 187608 通／週である．その内 spam でないと判定されたメールは 121476 通／週であり，spam と判定されたメールは 66132 通／週であった(spam 含有率 35.3%)．各 MTA の spam メール数を見ると，主に spam が配送される MTA1 では全メール 106295 通／週に対し，spam 62253 通／週(同 58.6%)，主に正常メールが配送される MTA2 では全メール 81313 通／週に対し，spam 3879 通(同 4.8%)であった．各 MTA の spam 含有率を見て分かるように，正常メールを受信する MTA2 の spam 含有率が低く, きちんと whitelist による振り分けが行われていることが分かる(図 4).

続いて, 各 MTA で検出する spam の種類について分析する. MTA1 で検出された spam のうち，多くは greylisting，Black List により検出され，ヘッダ検査によっても検出されている(図 5)．次に検出率の高い対策はユーザアカウントの有無を調べる対策，throttling による対策である．本大学では受信前対策として throttling を行っているが，その spam 検出率は意外と低くなっていることが分かる．spam 対策の順序として，最後から 2 番目となっている greylisting では, 他の対策手法で検出できていない spam メールの多くを検出できていることが分かる.

MTA2 で検出された spam のうち, 多くの判定を上げた対策はヘッダ検査による対策, ユーザアカウントの有無を調べる対策，spamassassin による対策である(図 6)．MTA1 で効果が高かった greylisting は whitelist によって振り分けられているため，MTA2 では行っていない.

次に，1 台の MTA による運用結果について述べる．運用実験は 1 月 26 日の 18 時 15 分から 21 時 50 分までの 3 時間 35 分行い，実験開始直後から sendmail プロセスの数が上昇した(図 7). そのため，メールの配送に大幅な遅延が発生してしまい実用に耐える MTA ではなくなってしまった. 原因を調査するため, MTA の CPU やメモリ, I/O によるボトルネックの可能性も考慮し，プロセス数がピーク付近の時間帯において CUP，メモリ，I/O について調査したが問題は見つからなかった．しかし，プロセスの監視により，消えないメールプロセスが複数あることを確認した．何らかの原因によって，消滅しないプロセスが少しずつ増えることで，プロセスが溜まってしまったと考えられる．明快な原因の特定には至っていないが，恐らく sendmail における設定ファイルの調整ミスか，溜まったプロセスはコネクションが確立されたままである点から，分別装置にバグがあり FIN パケットの転送等に不具合が生じている可能性がある．また，図 8 において所々プロセス数が大幅に減少している理由は，sendmail を再起動したためである.

7. おわりに

本稿では，spam 対策に利用する複数の whitelist の一元管理を実現する方法を，IP アドレスの置き換えとポート番号の置き換えという 2 つの方法で構築し，それらの運用結果について論じた．これらの方法における基本的なコンセプトは，複数管理している whitelist を 1 つの形式にまとめ，分別装置内で管理を行うというものである．また，本学における spam 対策の構成に限らず，どのような構成，手法で spam 対策していても，汎用的に whitelist の一元管理が可能である．本システムを導入する際には，一般的にメールゲートウェイに設置すると効率よくメールのチェックを行うことが可

能である．もちろん，ネットワークの規模によってはメールサーバが一台のみであることも多いので，その際はメールサーバと外部ネットワークの間のどこかに分別装置を設置し，リピータハブなどで MTA を接続すればよい．ただし，運用の結果 IP アドレスベースによる構成では安定した運用を行えたが，ポートベースによる構成ではプロセス数が安定しない問題が発生した．しかし，ポートベースによるシステムでは MTA は 1 台で済み，導入のコストが低いことから安定した運用方法を確立したいと考えている．今後はポートベースでの安定した運用を目指し稼動状況を確認していきたい．

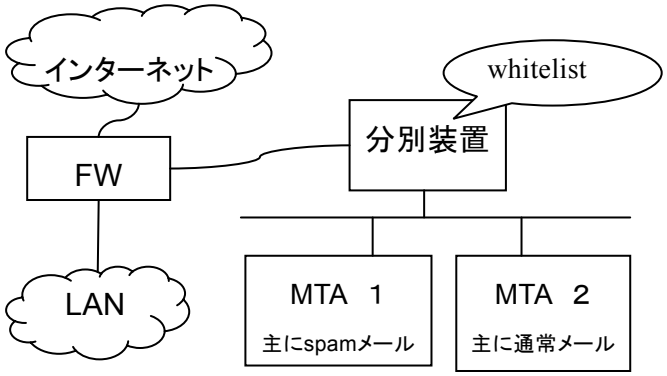

図 2. whitelist を持つ分別装置と 2 台の MTA の構成

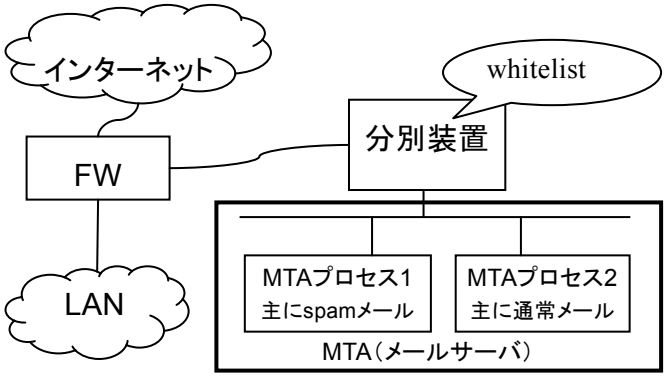

図 3. whitelist を持つ分別装置と 1 台の MTA の構成

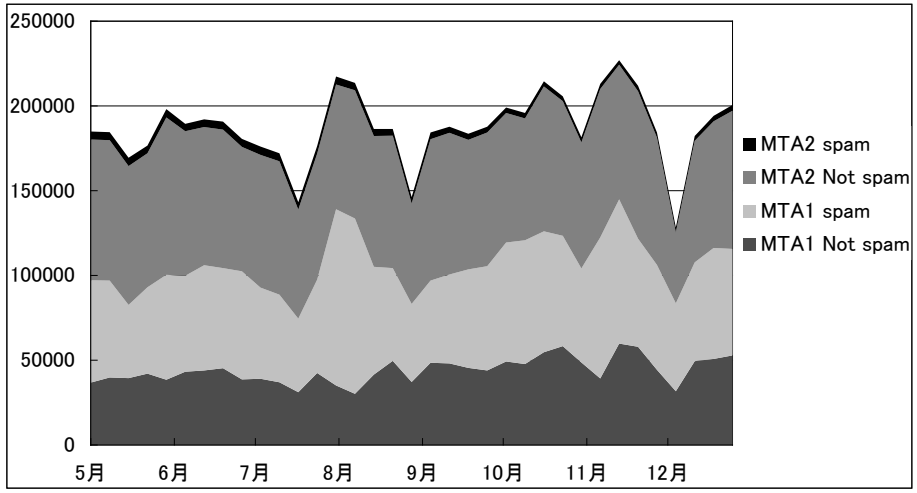

図 4. spam 検出数の推移

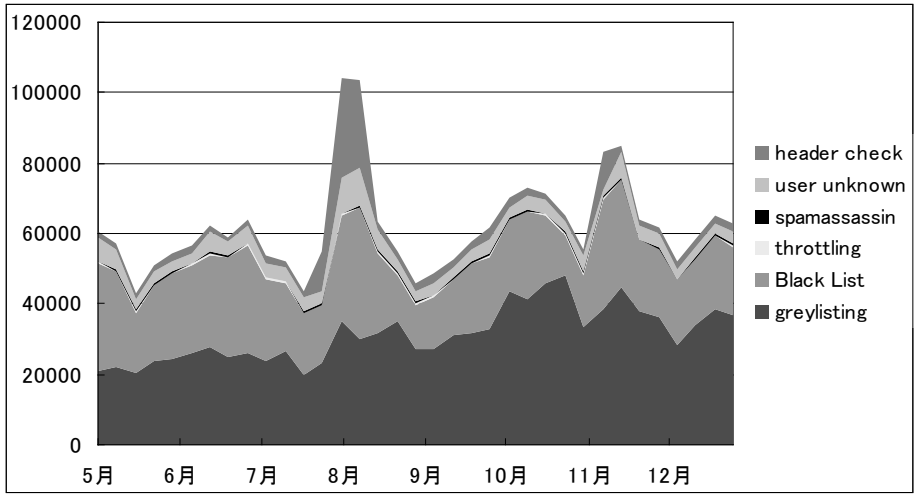

図 5. MTA1 における spam 検出数の推移

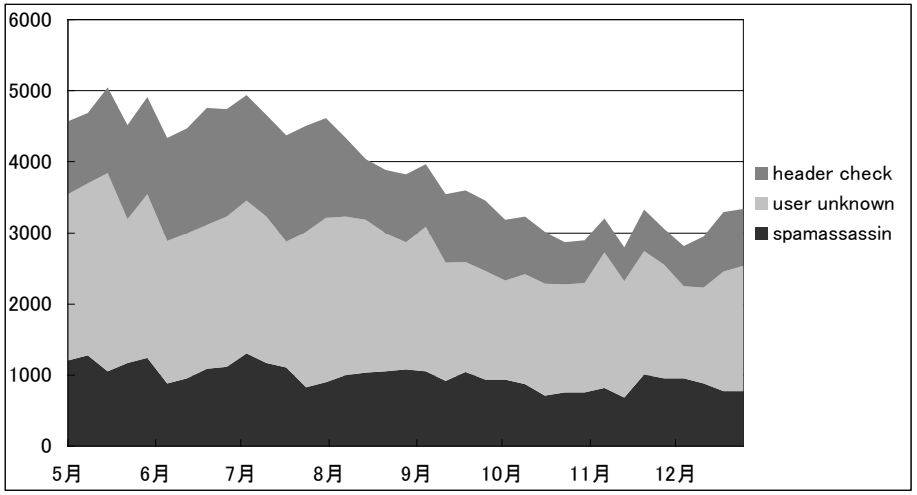

図 6. MTA2 における spam 検出数の推移

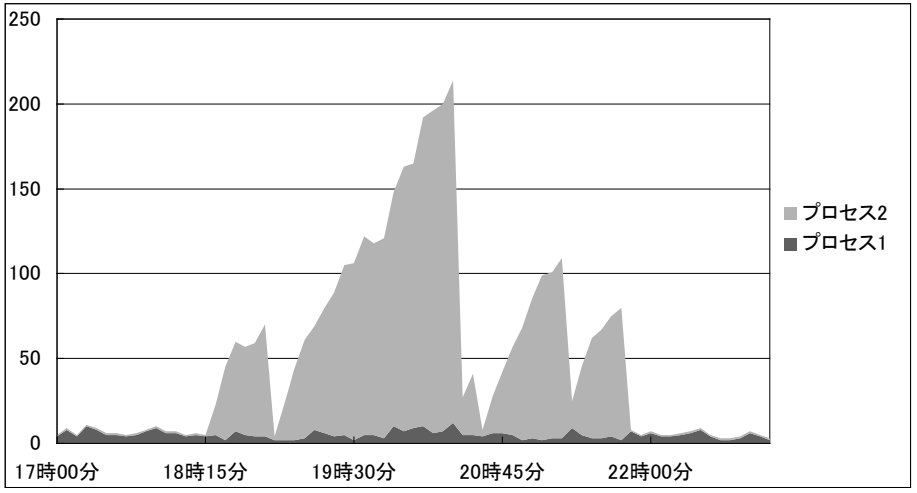

図 7. MTA における sendmail プロセス数の推移

謝辞

本研究の一部は，戦略的情報通信研究開発推進制度(SCOPE)の助成を受けている(課題番号：092310005).

参考文献

1) 吉田，矢田，原山，伊藤：“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,Vol.46, No.4, pp.1035-1040, Apr.2005.
2) 吉田:“LDAP を用いた統合メール管理システムについて”， 学術情報処理研究 No.7, pp.55-59, Spt.2003
3) 吉田：“ 統合メール管理システムとその使用経験について”， 大学情報システム環境研究, Vol.7，pp.47-52，Mar.2004
4) Sendmail Home Page
http://www.sendmail.org/
5) 吉田：“throttling による spam メール抑制の効果について” 情報処理学会研究報告，Vol.2005, No.39, pp.69-74，May.2005
6) M.Wahl, T.Howes, S.Kille: ”Lightweight Directory Access Protocol (v3)”, rfc2251, http://www.ietf.org, Dec.1997.
7) Greylisting.org - a great weapon against spammers
http://www.greylisting.org/
8) 吉田：“greylisting による spam メールの抑制について”，情報処理学会分散システム／インターネット運用研究，2004-DSM-35，pp.19-24，Sept.2004
9) Apache Spamassassin Project: “ Spamassassin ”
http://www.spamassassin.apache.org
10) 吉田：“メールゲートウェイにおける spam メールの検出について” 情報処理学会 DICOMO2004 シンポジウム論文集，pp.493-496，Jul.2004.(2004)
11) 三原，吉田:“メールゲートウェイの負荷分散による spam 対策について”，情報処理学会 分散システム／インターネット運用技術シンポジウム 2006，pp.67-72，Nov.2006
12) 三原，吉田：“Throttling による spam 対策のためのメールサーバの分別について”，電子情報通信学会 信学技報，pp.43-48，July.2007
13) 飯田，吉田：“spam メール対策のためのメールサーバの分別について”， 情報処理学会 DICOMO2009 シンポジウム論文集, pp.1291-1296, Jul.2009
14) The Spamhaus Project
http://www.spamhaus.org/
15) Distributed Sender Blackhole List
http://dsbl.org/main
16) K.Egevang, P.Francis;”The IP Network Address Translator (NAT), rfc1631, http://www.ietf.org, May 1994.