spam メール対策による遅延を低減するための whitelist 自動作成システム

松竹俊和† 金高一†† 吉田和幸†††

現在，spam 対策手法として，メール送信サーバの挙動を検査する対策手法である greylisting や throttling が広く用いられている．しかし，これらの対策手法は通常メールの受信にも遅延を強いる．この問題を解決するために信頼できるメール送信サーバをあらかじめ登録した whitelist が利用されている．しかし，この whitelist の作成はメールサーバ管理者の手で行われるため，登録数が少なく効果が限定的である．そこで，我々の研究室では whitelist の作成を自動化するシステムの開発を行った．本システムの利用により whitelist の登録数が増加し，遅延の低減が可能になった．本論文では whitelist 自動作成システムとその運用経験について述べる．

Automatic Whitelist Generating System to Decrease Delivery Delay by Spam Mail Measures

TOSHIKAZU MATSUTAKE† HAJIME KANETAKA††
KAZUYUKI YOSHIDA†††

As spam measures technique, greylisting and throttling, which are the measure techniques for inspecting the behavior of the Mail Sending server, are widely used. However, these measure techniques usually force the delay on the reception of mail. Whitelist that registers the trusted Mail Sending servers is solve this problem. However, because this whitelist is made by mail server manager manually, only small number of mail servers can be registered and the effect is limited. Then, we developed the system that generating whitelist automatically. The number of mail servers in the whitelist increase using this system, and the delay time of mail delivery can decrease. In this paper, we describe automatic whitelist generating system and 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，電子メールを始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 Bulk 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，ウイルスを検知・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イルスの有無の検査と同時にさまざまな spam 対策も行なっている[1] [2] [3].

我々が利用している spam 対策の中で sendmail [4]の throttling 機能[5]と greylisting [6]は，spam 送信サーバが，大量のメールを送ろうとするため，メール転送プロトコル（SMTP）[7]の規定とは異なる動作をすることに注目して，spam 検出を行うものである．

しかしながら，これら対策手法はメールの受信に遅延が発生してしまう. このため whitelist が利用される. whitelist とは無条件にメールの受信を行うことを許可した MTA(Mail Transfer Agent)のリストのことである. このリストを使うことで信頼できる MTA からのメールはすぐに受信できる.

自動作成システムの導入前は whitelist への MTA の登録作業は手動で行っており，登録できる数に限りがあった．そのため whitelist に登録されていない MTA からのメールには通常のメールであっても spam 対策が適用され，遅延が発生していた.

そこで本論文では，逆引きによるドメイン調査を利用した whitelist 自動作成システム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 greylisting と throttling について述べ，3 章で whitelist の必要性と問題点について述べる. 4 章で開発した whitelist 自動作成システムについて述べる．5 章で実際の運用結果について述べた後，最後に 6 章でまとめと今後の課題を述べる.

2. greylisting と throttling

現在，spam 対策手法として，spamassassin [8]等のコンテンツフィルタリングがよく利用されている[9]. コンテンツフィルタリングはメールの内容から spam 判定を行う．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る．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

† 大分大学大学院工学研究科
Graduate School of Engineering, Oita University
†† 大分大学工学部知能情報システム工学科
Department of Computer Science and Intelligent Systems, 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

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．

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（図 1）．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．

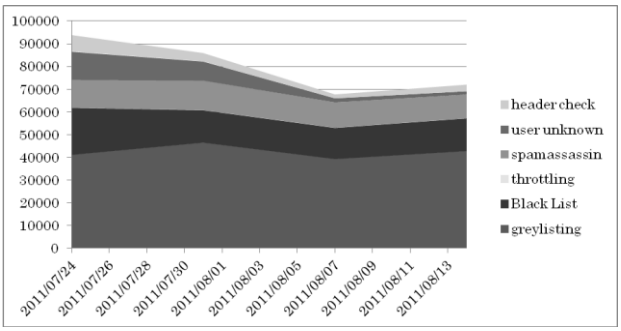

図 1　大分大学全体の spam 検出数内訳

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[10] 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. whitelist

3.1 whitelist による遅延低減

全てのメールに greylisting および throttling 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．したがって，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 greylisting および throttling で誤検知されてしまうような，規定された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．このように信用できるメールサーバの whitelist の作成は，重要である．

大分大学の場合 whitelist 作成は管理者が，greylisting によって動的に作成される auto-whitelist[11]から大学，政府機関，地方公共団体，医療機関，メールマガジン等のメール数の多い送信元を選び，MTA のドメイン名と送信元メールアドレスのドメイン部分が一致することを確認した後，whitelist に追加登録することで行われる．whitelist 自動作成システムを開発する以前の whitelist 更新頻度は半年に一回ほどであり，1500 件ほどの whitelist を管理していた．

3.2 whitelist の問題点とその解決策

whitelist は通常手動で作成され，その登録のために管理者は前章で述べたような確認作業を行う必要があり手間がかかる．そのため，whitelist の登録数はある程度の数に限られる．その結果，多くの通常メールに遅延が発生してしまう．大分大学の場合，システム運用以前の 2011 年 1 月の時点で通常メールの 56%に spam 対策による遅延が発生していた（表 2）．

そこで我々は whitelist の作成を自動で行うシステムを開発した．

4. whitelist 自動作成システム

4.1 大分大学のメールシステム

自動作成システムの説明の前に，大分大学のメールシステムの構成を述べる（図 2）．spam 対策を行うメールゲートウェイサーバの前に whitelist を持った分別装置が設置されている．メールゲートウェイには，主に spam メールを処理する MTA プロセス 1，主に通常メールを処理する MTA プロセス 2 が独立して起動している．メールが分別装置に送られると分別装置は whitelist を参照し，登録されたメールサーバからのメールならプロセス 2 へ振り分ける．登録されていないメールサーバからのメールは全てプロセス 1 に送られる．プロセス 1 ではさまざまな spam 対策が行われる（図 3）．一方，プロセス 2 ではほとんど spam 対策は行われない．

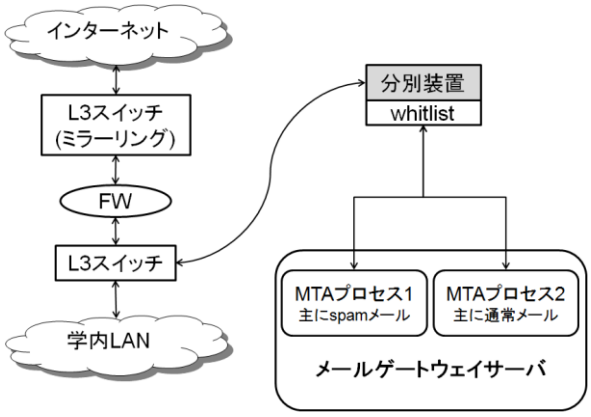

図 2　メールシステム構成

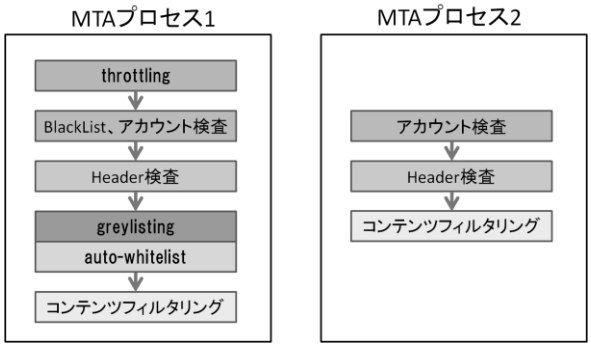

図 3　プロセスの spam 対策

プロセス 1 で行われている対策の内，最も遅延が大きい対策は greylisting である．この対策手法は他の対策手法の遅延が数秒であるのに比べて，平均で 1 時間 25 分 22 秒かかる（調査期間 2010 年 10 月 3 日～12 月 26 日）．プロセス 2 はほとんど対策が行われないので，遅延は数秒程度である．

4.2 システムの構成

自動作成システムは greylisting によって動的に作成される auto-whitelist を利用して whitelist を作成する．auto-whitelist とは greylisting の再送要求に応え，再送を行った MTA が一定期間登録される whitelist である．大分大学の場合再送してきた MTA は 2 週間 auto-whitelist に登録される．auto-whitelist に登録される情報は MTA の IP アドレス，送信元メールアドレス，宛先メールアドレス，受信時刻（登録時刻）の 4 つである．ここに登録された MTA の中から，後述する条件に合うものだけを whitelist に登録する．greylisitng による再送を行った送信者のみをシステムの対象とすることで，spam 送信者を誤って whitelist に登録しないようにしている．

メールシステムの構成上，一度 whitelist に登録されると，その MTA からのメールは分別装置により，MTA プロセス 2 へ送られるため，greylisting による再送処理が行われない．そのため，auto-whitelist にも記録されないため，再び同じ MTA が登録されることはない．

また，このシステムは greylisting の auto-whitelsit を利用しているため，通常の MTA であっても再送処理を行わなければ whitelist に登録できない．そのような MTA を登録するためには，手動で whitelist に登録する必要がある．

4.3 whitelist 登録の条件とその根拠

auto-whitelist の MTA を whitelist にそのまま登録すると再送処理を行う spam 送信者まで登録してしまう可能性があるため，登録の条件を考案した．条件は「MTA の IP アドレスから逆引きした FQDN（Fully Qualified Domain Name）と MTA の送信元アドレスのドメイン部分が後方一致するなら登録」というものである．後方一致させる理由としては，MTA の IP アドレスを逆引きすると，FQDN が得られるが，これはホスト名とドメイン名を含む．よってそのままでは FQDN と送信元メールアドレスのドメイン部分は完全には一致しないためである．

この条件は「spam 送信者は ISP や所属組織の正規のメール中継サーバを利用しない」という考えを根拠にしている．現在 spam の大半はボットに感染したエンドユーザ PC や spam 送信者自営のメールサーバから直接送信される[12]．それらには以下の特徴がある．

① 逆引き FQDN が設定されていない
　正規のメール中継サーバのほとんどは，逆引きで得られる FQDN を持つ．IP アドレスが逆引き FQDN を持たなければ spam 送信者の可能性が高い．

② 送信元ドメインを詐称する
　spam 送信では，正規のメールサーバを経由せずにエンドユーザ PC からメール送信した場合でも，正規のメール中継サーバからの送信のようにドメイン名を詐称する手法が一般的になっている[13]．一方で，ドメイン名は詐称できるが，IP アドレスは詐称できないので逆引き結果は，その送信者自身の FQDN となる．

これらの理由から逆引き FQDN と送信元ドメインが後方一致した場合は，所属するネットワークの正規メールサーバを使って送信していると考えられるため，whitelist に登録する．

4.4 whitelist の自動作成

前述の条件を利用した whitelist 自動作成は以下の流れで行われる（図 4）．

① 自動作成システムは受信メールサーバとは別の場所に存在する PC（Linux）上で動作する．システムは，受信サーバ上の auto-whitelist を PC にダウンロードする．その後 auto-whitelist に登録された MTA の IP アドレスと送信元メールアドレスを読み込む．
② MTA の IP アドレスを逆引きし，FQDN を取得．
③ 逆引きの結果得られた FQDN に送信元メールアドレスのドメイン部分全体が後方一致するかチェックする．一致するなら，保持している whitelist に追加登録する．この処理は auto-whitelist に載っている MTA（greylisting による再送を行った MTA）全てに行う．
④ IP アドレスが記述された whitelist を元に iptables 設定ファイルを作成する．iptables とは Linux に実装されたファイアウォール機能であり，DNAT 機

能を持っている[14]. whitelist によるメールの振り分け動作（4.1 節）は実際には iptables が行うため[15]，whitelist を iptables 設定ファイルに変換する．

⑤ iptables 設定ファイルをサーバにアップロード．

⑥ iptables を再起動させる．これによりファイルの設定が反映される．

これらの流れの内⑥以外はシェルスクリプトにより，自動で実行される．システムの根幹となる②と③のプログラムは Java でコーディングして，シェルスクリプトから実行するようにしている．⑥を自動化しないのは，iptables はフィルタリング機能を持っているため，iptables ファイルに問題があるとメールが届かなくなる可能性があるためである．iptables を再起動する前に，管理者が iptables ファイルを確認するようにしている．

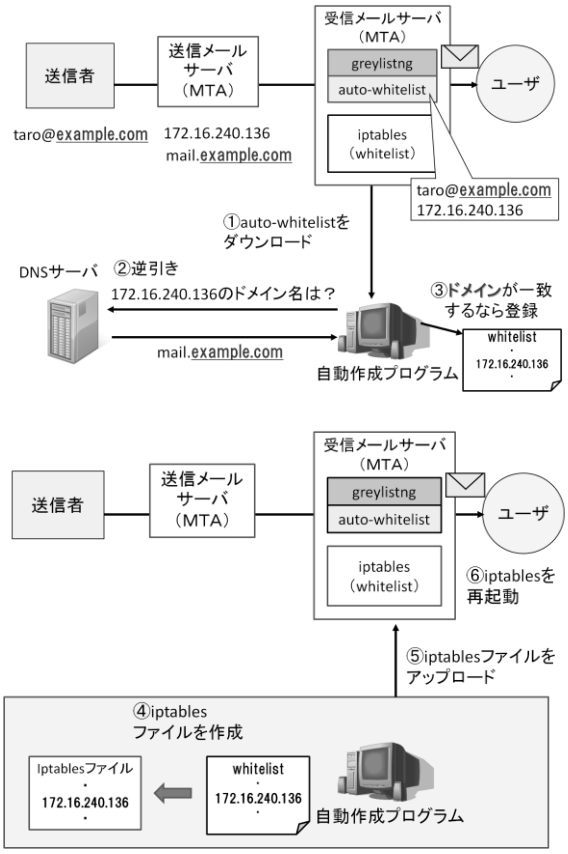

図 4　whitelist の自動作成

5. 運用結果

5.1 遅延の低減効果

システムの導入以前は通常メールの約 56%が MTA プロセス 1 に送られ，さまざまな spam 対策を受けていた．その中でも特に greylisting によって発生する遅延が大きく，平均して 1 時間 25 分 22 秒の遅延が発生する(調査期間 2010 年 10 月 3 日～12 月 26 日)．また，greylisting の遅延の中央値は 22 分 23 秒である．

システムの運用を開始した 2 月 2 日から 8 月 21 日までの間に whitelist 登録数は 1546 件から 9282 件まで増加した．自動作成システムによる whitelist の追加登録はリアルタイムではなく，auto-whitelist の更新間隔を考えて，数週間ごとに行った（表 1，図 5）．

表 1　whitelist 追加登録日時，件数

追加登録日時	追加登録件数
2011/2/2	1944
2011/2/25	882
2011/3/14	736
2011/5/2	759
2011/5/26	643
2011/6/13	629
2011/6/27	574
2011/7/8	553
2011/7/22	518
2011/8/4	498

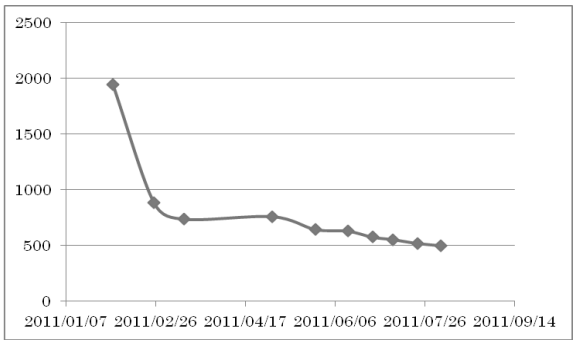

図 5　whitelist 追加登録件数の推移

whitelist 登録数の増加により，spam 対策を受ける通常メールの割合は，システム導入前の，56%から 27%まで低下した（表 2）．

表 2　spam 対策された通常メール数の変化

	spam 対策された通常メール	全通常メール	spam 対策された割合
2011/1/9～2011/2/6	390506	688633	56.71%
2/6～3/6	278261	621707	44.76%
3/6～4/3	189185	878979	21.52%
4/3～5/1	217808	825921	26.37%
5/1～5/29	225046	769981	29.23%
5/29～6/26	275180	913266	30.18%
6/26～7/24	243638	853923	28.53%
7/24～8/21	207447	761760	27.23%

また，spam 対策を受けた通常メールの割合の推移は図 6 のようになっている．4 週間ごとの推移を載せている．1～3 回目の whitelist 追加登録により，spam 対策を受ける通常メール数が著しく減っている（2011 年 2 月 6 日～2011 年 4 月 3 日）．それ以降は，whitelist 登録数が増加しても減少数は少なくなっている．ただし，whitelist に新しく追加登録した週は spam 対策を受ける通常メールの数が減少しており，随時 whitelist に MTA

を追加したことによって, spam 対策を受けた通常メールを低い水準に抑えることができたと考える.

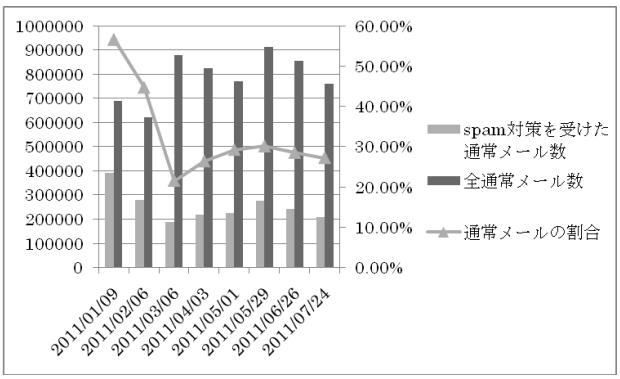

図 6　spam 対策された通常メールの推移

次に, spam 対策による遅延値の低減について述べる. 遅延値の計算方法は,（プロセス 1，2 の両方を含んだ全ての spam 対策によって発生した遅延の合計）÷（受信した通常メール数）とした．つまり通常メール 1 通あたりに発生する平均遅延を求めている．こうした理由は週ごとにメールの受信数が違うので，単純に遅延の合計値は比較できないためである．この平均は実際に 1 通のメールにそれだけの遅延がかかっているのではなく，全通常メール数における遅延の比率として提示している．以下の表 3 にシステム導入前（2011/1/23～1/30）の遅延値と導入後（2011/8/14～8/21）の遅延値を載せている.

表 3　通常メールの遅延平均

	spam 対策による遅延合計	通常メール数	平均
2011/1/23～1/30	105516795 秒	182901 通	577 秒
2011/8/14～8/21	63024182 秒	168499 通	374 秒

表より，通常のメールにかかる遅延が大幅に減少していることが分かる.

また，このシステムは greylisting の auto-whitelist を利用しているため, whitelist 登録数が増えて, greylisting に通常メールが来なくなると, 新規登録数が減少する. よって現状の水準からさらに遅延を低減させるには, 新しい whitelst 登録条件が必要となる.

5.2 whitelist に登録されたサーバの調査

whitelist は自動で作成されるため, 誤って spam 送信者を登録してしまう可能性がある．よって，whitelist に登録されたメールサーバの調査を行った．具体的には，whitelist に登録されたメールサーバから送られたメールがコンテンツフィルタリング（図 3 のプロセス 2）で spam 判定されたかどうか調査を行った．調査期間は 2011 年 7 月 24 日～8 月 21 日の 4 週間である.

調査の結果，この期間に whitelist に登録されたメールサーバからのメールで spam と判定されたのは 7623 通（1905 通／週）であり，その期間に受け取った全メール数は 1080958 通（270239 通／週）と分かった.

コンテンツフィルタリングによって spam 判定されたメールを送信した MTA は以下のようになる（一部抜粋）.

fc-ecn.ecnavi.jp [220.213.233.37]
mkrm1108d.rakuten.co.jp [203.190.62.108]
mkrm1091d.rakuten.co.jp [203.190.62.91]
cd018.mail-amycus.jp [210.169.223.201]
nm28-vm1.bullet.mail.bf1.yahoo.com [98.139.212.252]
fat.coara.or.jp [192.244.1.16]
mail.mingeihin.jp [61.195.65.75]
murago.biz [219.94.175.31]
out.oct-net.ne.jp [202.220.160.114]
col0-omc2-s14.col0.hotmail.com [65.55.34.88]
access2011u2.net [61.58.45.170]
10ibl21ser04.datacenter.cha.cantv.net [200.11.173.10]
no4.excelenceinfo.com [77.89.252.5]
n2.grp.kks.yahoo.co.jp [114.111.116.121]
mail-yw0-f46.google.com [209.85.213.46]

次に 7 月 24 日～8 月 21 日の期間内で spam 判定を受けたメールを送信した MTA の中で，多く spam を送っているものを表 4 に並べる．併せて，その MTA が送信した通常メールの数を記述する．これはコンテンツフィルタリングが通常のメールを spam と誤検知している可能性があるからである.

表 4　spam 判定されたメールを多く送信した MTA

MTA（FQDN）	spam 数	通常メール
fat.coara.or.jp	450	574
msvk8.travel.rakuten.co.jp	392	143
mkrm1123d.rakuten.co.jp	182	300
msvk10.travel.rakuten.co.jp	165	827
mail.sv1.nifty.com	156	36
mkrm1127d.rakuten.co.jp	141	402
msvk9.travel.rakuten.co.jp	138	663
msvk12.travel.rakuten.co.jp	133	615
msvk11.travel.rakuten.co.jp	107	566
mkrm1125d.rakuten.co.jp	103	439

上位 10 件の内，1 番目の coara.or.jp と 5 番目の nifty.com は信頼できる ISP であり，forward 等で spam が中継されたものと思われる．他は全てインターネット総合企業である楽天のドメインにある MTA となっている．こちらも信頼できる企業であるため，ここで多く検知されている spam は誤検知の可能性が高い. このことからシステムは信頼性の面で問題ないと思われる.

6. おわりに

本論文では，送信元ドメインの調査による whitelist 自動作成システムを構築し，それらの運用結果について論じた．これにより，手動よりはるかに登録数の多い whitelist を作成することができた．運用の結果，多

くの通常メールが whitelist によってすぐに受信できるようになり spam 対策による遅延が低減した．

今後の課題として，システムは現在，whitelist の追加の機能しかないため，長期間使用し続けると，whitelist の肥大化を招き，参照に時間がかかる恐れがある．よって今後は whitelist から MTA を削除する条件を考案する必要がある．他には，現在使用している条件の他に新たな条件を加えて，さらに遅延を低減させることを考えている．

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, TM2005-13, pp.69-74, May.2005
6) Greylisting.org - a great weapon against spammers: http://www.greylisting.org/
7) J. Klensin;“Simple Mail Transfer Protocol (SMTP)”，rfc5321, http://www.ietf.org , Oct.2008
8) Apache Spamassassin Project:“Spamassassin”，http://www.spamassassin.apache.org
9) 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
10) The Spamhaus Project: http://www.spamhaus.org/
11) 吉田：“greylisting による spam メールの抑制について”, 情報処理学会研究報告，2004-DSM-35, pp.19-24, Sept.2004
12)JEAG OP25B: http://jeag.jp/swg/op25b/
13)BIGLOBE の迷惑メールへの取り組み：http://security.biglobe.ne.jp/spam/taisaku.html
14) Manpage of IPTABLES：http://www.linux.or.jp/JM/html/iptables/man8/iptables.8.html , Mar.2002
15) 松竹，吉田：“iptables を利用した spam 対策用 whitelist を一元管理するためのメールシステム”，情報処理学会 インターネットと運用技術シンポジウム 2010 論文集，pp.75-80，Dec.2010