山形大学紀要（工学）第32巻　平成22年２月
Bull. Yamagata Univ. (Eng.) Vol.32 Feb. 2010

MTAから見たSPAMメールの特徴

山　本　広　志

地域教育文化学部

The Features of the Spam Emails from the Viewpoint of MTA

YAMAMOTO Hiroshi

Faculty of Education, Art and Science

（平成21年９月30日受理）

Abstract

In an attempt to investigate the features of spam emails from the viewpoint of MTA, experiments were made in 2009 with five real servers installed in Japan and a domain name under .jp. It has been found that not only did a group of spam emails arrive at the server with a high MX priority in the DNS, but also another did at the one with a low MX priority in comparable numbers and the remaining mails at the other with A record in one fewer number of digits. 99.99% of these spam emails came via IPv4 whereas only 0.01% via IPv6. The countries of origin where their IP addresses were assigned to most were Brazil, Korea, Vietnam, China and India in descending order. These five countries made up 42 % of all. 19% of envelope from domains in the spam emails were identified as the destination domain. 51% of them had their existent domains, but were not reachable email addresses. Reachable addresses were 27% of all. Although SPF and Sender ID were registered in 31% and 35% of domains respectively, DomainKeys and DKIM were less than 0.05%. In one of the experiments in which greet pause was set, about 50% connections were disconnected within 10 seconds. After that, the frequency of accumulated disconnections increased in a stair-step fashion until almost all were eventually disconnected within 120 seconds.

１．序

インターネットの商業利用が解禁されて以降，SPAMメール（迷惑メール）は人々を悩ませ続けている．年々増大するSPAMメールはインターネットに関連する資源を浪費し，人々の時間を浪費し人々を不愉快にし，さらには莫大な経済的損失をもたらしている．そのためSPAMメールを排除する目的で様々な対策の提案や実行がなされてきた．初期には単純な方法，例えばIPアドレスやメールアドレスのブラックリストとホワイトリスト，「Viagra」のようなキーワードで排除する方法が有効だった．しかし対策が普及するとSPAMメール送信者は対策をかいくぐる方法を考え，イタチごっこになっている．[1)]

SPAMメール対策はサーバ管理者がMTAで行うもの，ネットワーク管理者がパケットフィルタや帯域制限で行うもの，利用者がMUAで行うものに大別される．ネットワーク管理者が行う対策としては25/tcpのフィルタ（OP25B）が普及しつつある．これはMTAを持つネットワーク利用者に著しい不便を強いることになるが，SPAMやワームの被害があまりにひどいためにやむを得ないと考えられている．MUAでの対策としては学習機能付SPAMフィルタが普及しつつあり，一定の効果を上げている．そして本稿ではMTAで行う対策を詳しく取り上げる．

近年普及しつつあり今のところ一定の有効性が期待できるMTAでの主なSPAMメール対策には，送信ドメイン認証，Greylisting，Greet Pause，送信者アドレス検証が挙げられる．

まず，送信ドメイン認証はIPアドレスによる方法と電子署名による方法に大別される．IPアドレスによる方法にはSPFとSender IDがある．SPF (Sender Policy Framework)[2] では各ドメイン管理者が，自分のドメイン発のメールを発信するIPアドレスを予めDNSに登録しておく．受信側MTAはその登録情報とSMTPの接続元IPアドレスをリアルタイムで比較することによって，送信者メールアドレスが詐称かどうかを判定する．

Sender ID[3] はSPFと互換性があり，違いは少ない．主な違いはSPFが発信者ドメインとしてエンベロープfromを利用するのに対し，Sender IDは発信者ドメインをヘッダから取得することである．

IPアドレスによる送信ドメイン認証は送信側の設定が極めて簡単という利点がある一方，メール転送に対応できないという弱点がある．各利用者のメール転送設定全体をMTA管理者が把握することは極めて困難と言える．

一方，電子署名による送信ドメイン認証は接続元IPアドレスを用いないため，転送による影響を受けない．各ドメイン管理者は予め電子署名に必要な秘密鍵と公開鍵を作成し，公開鍵をDNSに登録しておく．そしてメールの送信時にMTAから呼び出されたプログラムがメールの内容と秘密鍵を使って自動的に署名をヘッダに付加する．

受信側ではヘッダの署名と公開鍵を使って送信者メールアドレスが詐称かどうかを判定する．メールの内容が変更されれば当然署名も変わるため，SPAM送信者が正当な署名を流用しても詐称と判定できる．

電子署名による送信ドメイン認証のうちDomainKeys[4] は署名のないメールへの対応が規定されていなかった．この弱点を改良したのがDKIM（DomainKeys Identified Mail)[5] で，DKIMは署名のないメールについても公開鍵と共に予め登録されたポリシを参照することによって詐称と判定することが可能になった．しかし，電子署名による送信ドメイン認証はIPアドレスによる方法に比べて送信側ドメインの設定が簡単ではなく，普及しにくいという欠点がある．送信側ドメインの設定がなければ受信側で苦労して設定しても全く効力がない．

次のGreylisting[6] は認証とは全く異なる視点に基づいている．SPAMメール送信者は短時間に大量のメールを送り付けることを重視するため，通常のMTAとは異なり一時的エラー（4.x.x）で送れなくても再送しない場合が多い．この習性を利用し，最初の何回かの接続では必ず一時的エラー（4.x.x）を返すように受信側MTAを設定し，相手に再送を促す．相手が通常のMTAであれば一定時間後に再送されるため，一定の遅延でメールを受け取ることができる．欠点としては遅延の他に，善良な相手にも負荷をかけるということと，障害等の際にメールを受け取れない確率が高まるということが挙げられる．

また，Greet Pause[7] もSPAMメール送信者の性質を利用している．SMTP接続の冒頭で応答（Greeting）を送るまでに受信側MTAがわざと遅延を挿入する．SPAMメール送信者は短時間に大量のメールを送り付けることを重視するため，比較的短い時間であきらめてしまう場合が多い．これに対して通常のMTAは数分間待つことができ，若干の遅延でメールを受け取ることができる．欠点としては，善良な相手にも負荷をかけるということと，SPAMメール以外でも短時間しか待てないSMTP接続が現実には来るということが挙げられる．

最後に送信者アドレス検証[8] は，SMTP接続を受けた際にリアルタイムでMAIL FROM:アドレスにSMTP接続を行い，送信者メールアドレスの有効性を確認する方法である．メールアドレスの詐称を判定することは原理的にできないが，無効なメールアドレスを検出することはできる．欠点としては，ドメインを詐称された相手に執拗なプローブを行う結果となりがちであることが挙げられる．

以上のようにSPAMメール対策の技術にはそれぞれ一長一短があり，複数の方法を組み合わせて試行錯誤を重ねているのが現状である．[9] SPAMフィルタリングの研究は多くある[1] が，MTAからの視点で見たSPAMメールの特徴を分析した研究は数少ない．[10]

2．実験方法

序で述べた幾つかのSPAMメール防止技術を意識しつつ，SPAMメールの特徴を分析するため，以下の実験を行った．

PCを５台用い，Linux 2.6をOSとする実験用サーバを構築した．実験用サーバはIPv4およびIPv6インターネットの双方に日本国内で接続し，５台がそれぞれ異なるパブリックIPアドレスを有する．

SPAMメールを受信する実験用ドメイン名は，末尾が.jpで10年以上前に使用停止されたものを再利用した．ドメインが使用されていた当時は多数の利用者がいて，使用停止後１年間はメール転送の措置が講じられた．その後は全利用者宛メールが「user unknown」で拒否される状態のまま放置されていた．この実験用ドメイン宛には現在も世界から毎日数千通のメールが届くが，そのほぼ全てがSPAMと考えられる．届いたメールのうち1,000通を抽出して目視検査したところ，1,000通全てがSPAMであった．今回の実験にあたってSPAMを誘引するための新たな行為は何も行っていない．

５台のサーバには，それぞれ「NS1」「NS2」「A」「MX10」「MX20」という名前を付けた．これは実験用ドメイン名のDNS設定で，それぞれのサーバが登録されたレコードに由来する．実験用ドメイン名のネームサーバは「NS1」サーバと「NS2」サーバとし，NSレコードに登録した．「A」サーバは，AレコードとAAAAレコードの双方に登録した．「MX10」サーバと「MX20」サーバはMXレコードに登録し，優先順位の値をそれぞれ10および20とした．どのサーバもDNSでIPv4アドレスおよびIPv6アドレスの両方が参照できるようになっている．

実験は2009年８月～９月に行った．

2.1　実験１

sendmail 8.14.3，sid-milter 1.0.0およびdkim-milter 2.8.3を用いて５台のサーバでSPAMメールを14日間受信し，SPAMメールそのものと伝送上の特徴を記録した．通常のMTAの設定とは異なり，５台とも実験用ドメインを自己宛と認識しメールを中継せず自己のメールスプールに格納するよう設定した．また，ユーザ名と無関係に実験用ドメイン宛のメールは全て１つのメールボックスに格納されるよう設定した．従って「user unknown」は発生せず，それぞれのサーバで到達した実験用ドメイン宛メール全てが宛先ユーザ名と関係なく同じメールボックスに格納される．

この際，sid-milterがSender IDとSPFを確認しログに記録する．ただし，受信拒否は行わない．さらにdkim-milterがDomainKeysとDKIMの検証結果を記録する．こちらも受信拒否は行わない．なお，dkim-milterはdk-milter 1.0.2のライブラリを組み込んでコンパイルした．実験用ドメイン自身では，全てのIPアドレスでSPFにfailを設定した．("v=spf1 -all") 実験用ドメインからメールの発信は行わない．実験用ドメイン自身のDomainKeysとDKIMには何も設定しなかった．

また，自己認証の電子証明書を作成し，STARTTLSを受け付けるようにもした．

2.2　実験２

postfix 2.6.3を用いてGreetingを出す前に240秒のGreet Pauseをかけ，接続の様子を７日間ログに記録した．postfixはpostfix-sleep.patchを当ててコンパイルしたので，接続が切断されるまでの時間を記録することができる．切断されるまでの時間が正しく記録されることはtelnet localhost 25を使って手動で確認した．

2.3　実験３

postfix 2.6.3を用いて７日間SPAMメールを受信し，送信者アドレス検証を行った．postfixはエンベロープfromのメールアドレスが到達可能かどうかを自動的に検証し，ログに記録する．「user unknown」等の理由でエンベロープfromが到達不能の場合でも受信拒否は行わなかった．

3．結果および検討

3.1　SPAMメールの数と大きさ

実験１で到達したSPAMメールの数と大きさをTable 1にまとめた．通常の電子メールであれば「MX10」サーバに集中して到達するはずであるが，実際には「MX10」サーバに匹敵する数が「MX20」サーバにも到達していた．これはMXで優先度の低いメールサーバはSPMA対策の遅れている場合があり，その弱点を狙っていると考えられる．また，MTAの通常の振る舞いとは異なり「A」サー

Table 1 Quantitative data of received spams (Experiment 1)

サーバ名	到達数 [通]	総分量 [バイト]	平均分量 [バイト／通]	１通の最大分量 [バイト]	１通の最小分量 [バイト]
NS1	0	0	–	–	–
NS2	0	0	–	–	–
A	2,295	5,803,399	2,528.71	23,090	833
MX10	18,464	118,365,068	6,410.59	579,535	389
MX20	17,817	79,672,785	4,471.73	59,108	392
全体	38,576	203,841,252	5,284.15	579,535	389

バに到達したSPAMメールもあった．敢えてＡレコードに登録されたIPアドレスに送る理由はDNS検索の時間を節約するためであろうか．SPAMメール送信は短時間に如何に大量のメールを送り付けるかが勝負であり，送信者がDNS検索の時間を少しでも節約したいと考えたとしても不思議ではない．Ａレコードに登録されているアドレスにメールを送っても正しく届くとは限らないが，実際には届く場合が多いということを反映した結果と考えられる．それに対して「NS1」サーバと「NS2」サーバに到達したSPAMメールは皆無であった．これは，ネームサーバはメールサーバと分離されている例が多いため，わざわざネームサーバにSPAMメールを送り付けるのは効率が悪いことの現れと考えられる．

SPAMメールの大きさに着目すると，「MX10」サーバに到達したSPAMメールの1通の平均の分量（ヘッダを含む）が6.4kBであるのに対し，「MX20」サーバは4.5kB，「A」サーバは2.5kBと差が出た．MXで優先度の低いメールサーバは弱点が狙える一方，大きすぎるメールの中継が拒否される場合がある．従って大きなSPAMメールはMXで優先度の高いメールサーバに直接送り付けた方が有利であるということの現れと考えられる．事実1通の最大分量を見ると，「MX10」サーバに到達したSPAMメールは最大で580kBもあり，他のサーバの最大値より１桁大きい．「A」サーバに到達したSPAMメールは前述のように時間の節約を強く意識した送信者によるものと推測されるので，その結果として１通当たりの分量が最も小さくなっていると考えられる．なお，全体の平均は5.3kBだった．

到達したメールのうちMAILER-DAEMONからのものが全部で５通あった．これらを目視で確認したところ，４通はバックスキャッタと判断できたが，残り１通は発信元メールアドレスを詐称したSPAMメールだった．

また，DATA部が空のSPAMメールが「MX10」サーバに387通，「MX20」サーバに368通到達した．DATAコマンドが完了しないとメール送信が有効にならないことから，空のDATA部が送られて来ていることになる．わざわざ広告効果のないSPAMメールを送る意図は明らかではないが，メールアドレスの有効性を確かめている可能性がある．

3.2 IPのバージョン

実験１でSPAMメールに使用されたIPのバージョンは，ほとんど全て（99.99%）がIPv4で，IPv6は４通（0.01%）しかなかった．IPv6で到達できる宛先がまだ少数派である現状から考えれば当然の結果と言える．IPv6で到達した４通のヘッダを目視で確かめたところ，欧州のインターネット接続業者と北米の大学のメールサーバが接続元だった．SPAMメール送信者が自らIPv6を使用した訳ではなく，IPv6に対応した善意のMTAが悪用された結果と考えられる．４通は全て「MX10」サーバに到達しており，他のサーバにIPv6で到達したSPAMメールはなかった．

ただ，今まではIPv6がほとんど悪用されていないためIPv6に対するセキュリティの甘いサーバがあり，今後IPv6が普及する過程でIPv6を使用したSPAMメールや攻撃が一気に流行する可能性がある．

3.3 接続元IPアドレス

実験１で接続元IPアドレスの配布先を国別に集計した．その結果をTable 2に示す．この集計はドメイン名とは無関係に，各RIRが公表している割当国を元にした．

接続件数上位５ヶ国のブラジル，韓国，ベトナム，中国，インドで全体の42%を占めている．日本は36位で全体の0.5%に過ぎなかった．

3.4 HELO

実験１でSPAMメール送信者（クライアント）からHELOまたはEHLOで送られて来た「SMTPクライアントのFQDN」を分析した．結果をTable 3に示す．規則通りにクライアントFQDNが送信さ

Table 2 Connection origins（Experiment 1）

順位	国名	接続数	比率 [%]
1	ブラジル	4,628	12.0
2	韓国	3,137	8.1
3	ベトナム	3,007	7.8
4	中国	2,923	7.6
5	インド	2,405	6.2
6	アメリカ	1,919	5.0
7	ポーランド	1,718	4.6
8	ロシア	1,144	3.0
9	コロンビア	921	2.4
10	ルーマニア	907	2.4
11	トルコ	904	2.3
12	スペイン	819	2.1
13	アルゼンチン	809	2.1
14	ドイツ	732	1.9
15	イギリス	701	1.8
16	ウクライナ	614	1.6
17	イタリア	600	1.6
18	チリ	596	1.5
19	タイ	544	1.4
20	チェコ	501	1.3
21	エジプト	467	1.2
22	フランス	458	1.2
23	イスラエル	439	1.1
24	メキシコ	429	1.1
25	パキスタン	364	0.9
26	ブルガリア	336	0.9
27	インドネシア	319	0.8
28	ペルー	285	0.7
29	ポルトガル	272	0.7
30	ギリシア	246	0.6
31	モロッコ	237	0.6
32	スロバキア	236	0.6
33	オランダ	213	0.6
34	ハンガリー	210	0.5
35	セルビアモンテネグロ	200	0.5
36	日本	196	0.5

れて来たのは半数以下に過ぎず，SPAMメールに特徴的な傾向が現れた．

「MX10」サーバと「MX20」サーバの傾向は比較的似通っていて，規則通りのクライアントFQDNはそれぞれ36%と42%だった．次に多いのが「.」（ドット）のない名前で，それぞれ26%と32%あった．この中には「localhost」が全体の2%前後含まれる．３番目に多いのがクライアントIPアドレスで，それぞれ11%と５%あった．４番目に多いのはサーバIPアドレスの４%で，これは明らかに趣旨に反する．送信者メールアドレスや宛先メールアドレスのドメイン部分の流用は少なく，どれも１%未満だった．

一方，「A」サーバに到達した分は大きく傾向が異なり，クライアントIPアドレスが実に81%を占めた．２番目に多いのが規則通りのクライアントFQDNで，11%だった．敢えて通常と異なる動作でAレコードのアドレスにメールを送るシステムは種類が限られ，特定のものが高頻度で使用されているのではないかと推測される．

Table 3 Client names by means of HELO or EHLO（Experiment 1）

HELO/EHLO クライアント名	MX10 サーバ	MX20 サーバ	A サーバ	全体
クライアント FQDN（逆引）	6,612 (35.8%)	7,480 (42.0%)	254 (11.1%)	14,346 (37.2%)
クライアント IP アドレス	2,099 (11.4%)	825 (4.6%)	1,862 (81.1%)	4,786 (12.4%)
エンベロープ from ドメイン	158 (0.9%)	5 (0.0%)	0 (0.0%)	163 (0.4%)
ヘッダ From：ドメイン	58 (0.3%)	5 (0.0%)	0 (0.0%)	63 (0.2%)
「.」（ドット）なし うち、「localhost」	4,883 (26.4%) 332 (1.8%)	5,604 (31.5%) 444 (2.5%)	29 (1.3%) 11 (0.5%)	10,516 (27.3%) 787 (2.0%)
サーバ FQDN	0 (0.0%)	1 (0.0%)	0 (0.0%)	1 (0.0%)
サーバ IP アドレス	795 (4.3%)	775 (4.3%)	0 (0.0%)	1,570 (4.1%)
宛先ドメイン	8 (0.0%)	8 (0.0%)	71 (3.1%)	87 (0.2%)

一般にはHELO・EHLOで名乗られるクライアントFQDNが真正でないからと言ってメール受信を拒否してしまうことは勧められないが，組織の外部からサーバIPアドレスを名乗るような接続はSPAMメールと断定して差し支えないと思われる．これに対して「localhost」やその他の「.」（ドット）のない名前は，善意のMTAの設定誤りという可能性を排除できない．

3.5 TLS

実験１で到達時にTLSが使用されたSPAMメールは67通（0.2%）に過ぎず，ほとんど（99.8%）はTLSが使用されなかった．TLSを使用したことが点数化されSPAMメールフィルタにおいて有利に働く可能性はあるものの，実際にはTLSがほとんど使用されないことから，有利な効果はほとんどないと推測される．

なお，TLSの使用された67通全てがMXで優先度の高い「MX10」サーバに到達していた．ヘッダを目視して確認するとTLSに対応した善意のMTAが悪用されたと判断できる例もあったが，SPAMメール送信者が自ら設置したMTAではないかと疑われる例も多かった．現時点でSPAMメール送信プログラム（いわゆるSPAM ボットを含む）にTLSを実装する利点はほとんどないと考えられる．

3.6 送信者メールアドレス

実験３で行った送信者アドレス検証の結果をTable 4に示す．まず，エンベロープfromが宛先ドメインである実験用ドメインを詐称しているSPAMメールが全体の19%あった．バウンスが出ないためSPAM送信者にとっては面倒がない．た

Table 4　Sender mail adrresses（Experiment 3）

エンベロープ from	MX10 サーバ	MX20 サーバ	A サーバ	全体
総数	17,476 (100.0%)	15,136 (100.0%)	26 (100.0%)	32,638 (100.0%)
宛先と同じドメイン	3,866 (22.1%)	2,388 (15.7%)	2 (7.7%)	6,256 (19.2%)
ドメイン不存在	695 (4.0%)	233 (1.5%)	2 (7.7%)	930 (2.8%)
ドメインは存在するがメールは届かない（user unknown 等）	7,437 (42.6%)	9,201 (60.8%)	22 (84.6%)	16,660 (51.0%)
メールが届く	5,478 (31.3%)	3,314 (21.9%)	0 (0.0%)	8,792 (26.9%)
ヘッダ From:と一致	15,739 (90.1%)	12,536 (82.8%)	22 (84.6%)	28,297 (86.7%)
ヘッダ From:と不一致	1,737 (9.9%)	2,600 (17.2%)	4 (15.4%)	4,341 (13.3%)

だ，以前はエンベロープfromのドメインが宛先ドメインと同一の場合は拒否されにくい可能性があったが，最近は逆に外部からのそうしたメールを拒否する設定が増えている可能性はある．

最も多いのは，ドメインは実在するものの「user unknown」等の理由で到達不能のアドレスで，全体の51%だった．これもバウンスが宛先か詐称したドメインのpostmasterに落ちるのでSPAMメール送信者にとっては面倒がない．送信者アドレス検証を実施すればSPAMメールを半減できることになるが，実運用での送信者アドレス検証には弊害があり，簡単には実施できない．しばしば詐称される特定のドメインに対して数多くの検証を行うと，不正なアクセスとみなされる可能性がある．また，SPAM以外で送信専用メールアドレスが利用される場合もある．そのため実運用で送信者アドレス検証を行っている例はあまり多くないと思われる．

次に多いのはエンベロープfromメールアドレスが到達可能なもので，全体の27%だった．SPAM送信者がわざわざ自分の身元を明らかにする手掛かりを増やしたりバウンスを処理する資源を割くとは考えにくいので，無関係な第三者のメールアドレスを詐称しているか，あるいはフリーメールのアドレスを使い捨てにしている可能性がある．どちらの場合も１つのアドレスを多用すればバックスキャッタでバウンスメールが殺到することになる．バウンスでメールボックスが溢れ，短時間で「ドメインが実在するものの到達不能のアドレス」になる可能性が高い．実際，実験３で到達不能だったアドレスには，かつて有効だったもののバウンスによって無効になったアドレスが含まれると考えられる．

Table 5　Sender domains（Experiment 3）

順位	ドメイン名	通数	比率 [%]
1	gmail.com	148	0.5
2	yahoo.com	138	0.4
3	hotmail.com	109	0.3
4	yahoo.co.jp	87	0.3
5	yahoo.cn	86	0.3
6	nifty.com	84	0.3
7	msn.com	79	0.2
8	odn.ne.jp	78	0.2
9	excite.co.jp	76	0.2
10	so-net.ne.jp	69	0.2
	宛先ドメインと同一	6,256	19.2
	総数	32,638	100.0

最も少なかったのはドメイン自体が存在しない場合で，全体の３%だった．エンベロープfromのドメインが存在しない場合は簡単に受信拒否の設定ができるため，SPAM送信者が存在しないドメインを詐称する割合が低くなっていると考えられる．

次に，使用されたエンベロープfromのドメインをTable 5にまとめた．宛先ドメインである実験用ドメインが詐称された19%を除けば最も多いgmail.comでも0.5%に過ぎず，使われたドメインが広く分散していることが分かる．gmail.comのように世界で広く使われているドメインの他，10位までに日本のドメインが５つ含まれている．簡単に受信拒否されないよう，SPAM送信者が宛先国に応じて工夫している様子が伺える．

なお，Table 4に示したようにエンベロープfrom全体の87%はヘッダFrom:と一致していた．

3.7　送信ドメイン認証

3.6で述べたように，送信元メールアドレスを実験用ドメインに偽装したメールがかなり多かった．そこで実験１の結果から送信元メールアドレスが実験用ドメインのものを除外した後，各送信ドメイン認証の結果を分析した．

まずSPFの判定結果をTable 6に示す．SPFは送信元メールアドレスとしてエンベロープfromを判定に使用する．SPFはDNSへの設定が手軽なため，送信側設定はかなり普及が進んでいるということが分かった．送信元メールアドレスのドメインにSPF設定のあった，判定結果がnone以外のSPAMメールは全体の31%あった．そしてfailとなったのが全体の５%，softfailとなったのが全体の７%だった．今回の実験ではfailとsoftfailは送信ドメイン詐称と判断して差し支えないと考えられ

Table 6　SPF results（Experiment 1）

SPF 判定結果	「MX10」サーバ	「MX20」サーバ	「A」サーバ	全体
fail	829 (8.1%)	466 (3.4%)	120 (5.2%)	1,415 (5.4%)
softfail	882 (8.7%)	768 (5.7%)	234 (10.2%)	1,884 (7.2%)
neutral	2,035 (20.0%)	1,660 (12.3%)	222 (9.7%)	3,917 (15.1%)
pass	586 (5.7%)	171 (1.3%)	0 (0.0%)	757 (2.9%)
permerror	52 (0.5%)	70 (0.5%)	0 (0.0%)	122 (0.5%)
none	5,811 (57.0%)	10,376 (76.8%)	1,719 (74.9%)	17,906 (68.9%)
計	10,195 (100.0%)	13,511 (100.0%)	2,295 (100.0%)	26,001 (100.0%)

Table 7　Sender ID results（Experiment 1）

Sender ID 判定結果	「MX10」サーバ	「MX20」サーバ	「A」サーバ	全体
fail	545 (5.6%)	641 (5.1%)	121 (5.3%)	1,307 (5.3%)
softfail	1,053 (10.8%)	1,274 (10.2%)	235 (10.3%)	2,562 (10.5%)
neutral	2,040 (21.0%)	1,598 (12.8%)	222 (9.7%)	3,860 (15.8%)
pass	573 (5.9%)	157 (1.3%)	0 (0.0%)	730 (3.0%)
permerror	53 (0.5%)	51 (0.4%)	0 (0.0%)	104 (0.4%)
none	5,473 (56.2%)	8,744 (70.1%)	1,708 (74.7%)	15,925 (65.0%)
計	9,737 (100.0%)	12,465 (100.0%)	2,286 (100.0%)	24,488 (100.0%)

るが，一般には利用者がメール転送を行っている場合が多々あり，直ちに受信拒否する訳にも行かない．MTA段階で受信拒否するのではなく，ヘッダAuthentication-Results: を付加して判定結果の活用を利用者に委ねるのが適切だろう．

SPFの判定結果がneutralとなったものが全体の15%もあり，SPAM判定に活用したい立場から見るとかなり物足りない設定が多いと言える．各ドメイン管理者も利用者の動向全体はなかなか掴みきれず，まずは無難な設定を行った結果と考えられる．また，SPAMメールであるにも関わらずpassとなったものが「MX10」サーバで６%もあった．これは善意のMTAが悪用されている場合と，SPAMメール送信者自身が送信元メールアドレスのドメインを管理している場合が考えられる．近年はレジストラ間の競争が促進される制度になり，ドメイン１つあたり年間千円以下で手軽に利用できる．SPAMメール送信者が安価なドメインを次々と使い捨てにして送信ドメイン認証を正しく設定すれば，送信ドメイン認証はSPAMメール排除の役には立たなくなる．

SPFがpassと判定されたSPAMメールは「MX10」サーバが６%であるのに対し，「MX20」サーバは１%，「A」サーバは０%と差が大きい．善意のMTAが悪用されている場合は当然MXが最優先のサーバに到達する．SPAMメール送信者自身が送信元メールアドレスのドメインを管理している場合は，SPAM判定を回避する自信があるためわざわざ通常のMTAと異なるサーバに送りはしない，ということがあるかも知れない．

次にSender IDの判定結果をTable 7に示す．Sender IDは送信元メールアドレスとしてヘッダのFrom:等を利用する．序で述べたようにSPFとSender IDは互換性があり，違いは判定対象とするドメイン名の検出方法程度で大きな差はないことから，Sender IDの判定結果もSPFと類似している．

一方，DomainKeysとDKIMはヘッダに電子署名を埋め込む方式で，IPアドレスで認証するSPFやSender IDとは異なる．実験１におけるDomainKeysの判定結果はpassが11とfailが２のみで，合わせて全体の0.05%だった．序で述べたようにDomainKeysは電子署名のないメールについては判定を行わないという欠点がある．その欠点を改良したのがDKIMであるが，実験１におけるDKIMの判定結果はpassが３とneutralが２のみで，合わせて全体の0.02%に過ぎない．DKIMと言えども各ドメインのDNSに設定がなければどうにもならない．DomainKeysとDKIMは現状ではほとんど全く普及していないということが分かった．

DKIMが普及すれば発信元メールアドレスが詐称かどうか判定できるようになるはずであるが，前述したようにSPAMメール送信者が安価なドメインを次々と使い捨てにしてDKIMを正しく設定すれば，DKIMもSPAMメール排除の役には立たなくなる．

3.8　Greet Pause

実験２の結果をFig. 1に示す．横軸は接続開始からの時間，縦軸はGreetingを待ち切れずに切断された接続およびGreeting以前にコマンドが送られて来た接続の累計割合を表す．最初の10秒で50%前後まで急峻に立ち上がり，その後は20秒，30秒，50秒，60秒，80秒，120秒と階段状に累計割合が上がって行く様子がはっきりと読み取れる．120秒で累計割合はほとんど100%に達する．「MX10」サーバと「MX20」サーバでは「MX10」

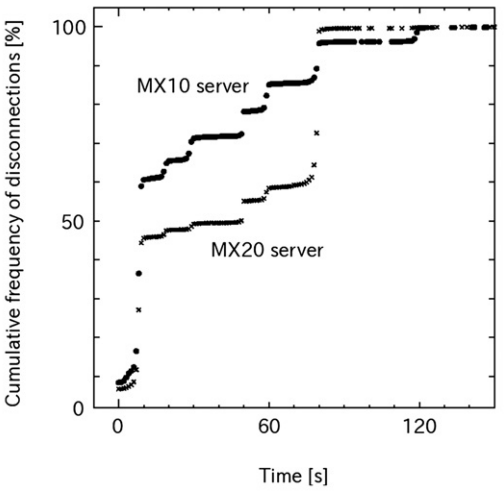

Fig. 1　Greet Pause（Experiment 2）

サーバの方が立ち上がりが早いが，80秒の時点で「MX20」サーバに逆転されている．これらのパターンは使用されているSPAMメール送信プログラムの設定が色濃く反映していると考えられる．

Fig. 1から，Greet Pauseを120秒以上に設定すればSPAMメールを100%近く排除できることが分かる．しかし通常のMTAでは問題ないとは言え，実際には顧客の承諾に基づいて大量のメールを送信するプログラムが120秒のGreet Pauseを待てない例が多々あり，実運用でそこまで長いGreet Pauseを設定することは難しい．

3.9　Content-typeとcharset

MTAの動作と直接の関係はないが，SPAMメールの性格を知る上で実験１によって到達したSPAMメールのヘッダからContent-typeとcharsetを分析した．まずContent-typeの集計結果をTable 8に示す．Content-typeはヘッダ中のものだけを集計し，htmlタグに記述されたものやmultipartの一部分に記述されたものは除外した．「MX10」サーバではtext/htmlが59%と過半数を占めたのに対し，「MX20」サーバではmultipartが59%と過半数だった．さらに「A」サーバではmultipartが94%でtext/htmlが０という際立った特徴が現れた．SPAMメール送信プログラムはmultipartを好むという可能性が考えられる．

次にcharsetの集計結果をTable 9に示す．charsetはヘッダ中のものとmultipartの一部分のものを集計し，htmlタグに記述されたものは除外した．「MX10」サーバと「MX20」サーバの結果

Table 8　Content-type（Experiment 1）

Content-type	「MX10」サーバ	「MX20」サーバ	「A」サーバ	全体
text/plain	1,206 (6.7%)	1,028 (5.9%)	131 (5.7%)	2,365 (6.3%)
text/html	10,620 (58.8%)	6,165 (35.3%)	0 (0.0%)	16,785 (44.4%)
multipart	6,248 (34.8%)	10,253 (58.8%)	2,154 (94.3%)	18,655 (49.3%)
計	18,074 (100.0%)	17,446 (100.0%)	2,285 (100.0%)	37,805 (100.0%)

Table 9　Charset（Experiment 1）

charset	「MX10」サーバ	「MX20」サーバ	「A」サーバ	全体
ISO-8859-1	8,816 (41.9%)	10,547 (42.4%)	44 (1.0%)	19,407 (38.6%)
ISO-8859-2	380 (1.8%)	287 (0.0%)	0 (0.0%)	667 (1.3%)
ISO-8859-3	21 (0.1%)	0 (0.0%)	0 (0.0%)	21 (0.0%)
ISO-8859-4	36 (0.2%)	1 (0.0%)	0 (0.0%)	37 (0.1%)
ISO-8859-5	34 (0.2%)	0 (0.0%)	0 (0.0%)	34 (0.1%)
ISO-8859-6	28 (0.1%)	0 (0.0%)	0 (0.0%)	28 (0.1%)
ISO-8859-7	35 (0.2%)	0 (0.0%)	0 (0.0%)	35 (0.1%)
ISO-8859-8	29 (0.1%)	1 (0.0%)	0 (0.0%)	30 (0.1%)
ISO-8859-9	39 (0.2%)	0 (0.0%)	0 (0.0%)	39 (0.1%)
ISO-8859-10	26 (0.1%)	0 (0.0%)	0 (0.0%)	26 (0.1%)
ISO-8859-11	35 (0.2%)	0 (0.0%)	0 (0.0%)	35 (0.1%)
ISO-8859-13	27 (0.1%)	0 (0.0%)	0 (0.0%)	27 (0.1%)
ISO-8859-14	19 (0.1%)	0 (0.0%)	0 (0.0%)	19 (0.0%)
ISO-8859-15	31 (0.1%)	0 (0.0%)	12 (0.3%)	43 (0.1%)
ISO-2022-JP	124 (0.6%)	83 (0.3%)	0 (0.0%)	207 (0.4%)
UTF-8	3,263 (15.5%)	2,845 (11.4%)	0 (0.0%)	6,108 (12.2%)
us-ascii	3,945 (18.7%)	1,803 (7.3%)	2 (0.0%)	5,750 (11.4%)
SJIS	34 (0.2%)	20 (0.1%)	0 (0.0%)	54 (0.1%)
Windows-1250	653 (3.1%)	266 (1.1%)	2,144 (49.3%)	3,063 (6.0%)
Windows-1251	259 (1.2%)	106 (0.4%)	6 (0.1%)	371 (0.7%)
Windows-1252	2,476 (11.8%)	8,118 (32.7%)	2,133 (49.1%)	12,727 (25.3%)
koi8-r	718 (3.4%)	774 (3.1%)	5 (0.1%)	1,497 (3.0%)
koi8-u	1 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.0%)
GB2312	0 (0.0%)	1 (0.0%)	0 (0.0%)	1 (0.0%)
計	21,065 (100.0%)	24,852 (100.0%)	4,346 (100.0%)	50,263 (100.0%)

は，ラテン文字であるISO-8859-1とus-asciiとWindows-1252の合計がそれぞれ72%および82%と半数を大きく超えて圧倒的に多く，日本語（ISO-2022-JP，SJIS）は１%未満だった．そして各々は少数ながら様々な言語があった．また，ユニコード（UTF-8）はそれぞれ全体の16%と11%を占めていた．

これに対して「A」サーバはcharsetでも際立った特徴を見せた．windows-1250とwindows-1252の両者が合計98%以上を占め，極めて限定されたSPAMメール送信プログラムが用いられていることを示唆している．しかしながら接続元IPアドレスは世界の広範囲に拡がっており，いわゆるSPAMボットが存在するか，あるいは極少ない種類のプログラムが広く使用されている可能性がある．

3.10　今後の動向

序で述べたSPAMメール対策のうち，送信ドメイン認証が今後５年程度のうちに普及するのか，SPF／Sender IDとDKIMのどちらが，あるいは両方が普及するのかに注目したい．Greylisting，Greet Pause，送信者アドレス検証は判別条件が確

定的ではなく，意図せずに必要なメールを排除してしまう可能性が残る．それに対して送信ドメイン認証は，送信ドメイン管理者と受信側MTA管理者が規格に基づいて明確に条件を定めることができる．その意味で送信ドメイン認証はMTAでのSPAM対策の「王道」と言える．

ただ，3.7で述べたように送信ドメイン認証も万能ではなく，SPAM送信者の使い捨てドメインに対しても認証が通ってしまう．送信ドメイン認証が普及した後には，この使い捨てドメインの扱いが焦点になろう．

今後５年というと，IPv4アドレス枯渇の影響でインターネット変革期になると考えられている．IPv6対応のために多くの機器が更新されると，それに連られてMTAのSPAMに関連する設定変更が急激に進む可能性もある．管理者にとって多忙な５年間となりそうである．

４．まとめ

MTAから見たSPAMメールの特徴を分析するため，日本国内に設置した５台の実サーバと末尾が.jpの実験用ドメインを使ってSPAMメールの受信実験を行った．５台のサーバにはDNSレコードの登録に由来する名称「MX10」「MX20」「A」「NS1」「NS2」を名付けた．実験の結果，以下のことが分かった．

⑴ SPAMメールは「MX10」サーバの他，「MX10」に匹敵する数が「MX20」に，１桁少ない数が「A」に到達した．「NS1」と「NS2」には1通も到達しなかった．

⑵ SPAMメールの99.99%はIPv4で届いた．IPv6は0.01%に過ぎなかった．

⑶ 接続元IPアドレスの割当国は，ブラジル，韓国，ベトナム，中国，インドの順に多かった．これら上位５ヶ国で全体の42%を占めた．日本は0.5%に過ぎなかった．

⑷ SPAMメールのエンベロープfromは19%が宛先ドメインと同一，51%がドメインが存在するものの到達できないメールアドレスで，メールが到達できるアドレスは27%だった．

⑸ 送信者ドメインのSPFとSender IDはそれぞれ31%と35%が登録済だったものの，failやsoftfailと判定されたのはそれぞれ全体の13%と16%だった．一方，DomainKeysやDKIMが登録されていたのは0.05%以下に過ぎなかった．

⑹ Greet Pauseの実験からはSPAMメールの接続が待機できる最大時間が分かった．最初の10秒で50%前後が切断し，その後は階段状に切断の累計割合が上がっていった．120秒でほぼ100%が切断した．

⑺ Content-typeはtext/htmlとmultipartが匹敵する割合で，最も多かった．charsetはラテン文字（ISO-8859-1，us-ascii，Windows-1252）が75%と半数を大きく超え，圧倒的に多かった．

謝　辞

英文に関する桑村昭国際化主幹の有益な助言に感謝する．本研究は山形大学教育研究基盤校費によって行われた．

文　献

1）Enrique Puertas Sanz, Jose Maria Gomez Hidalgo and Jose Carlo Cortizo Perez : Email Spam Filtering, Advances in Computers, **74**, pp. 45-114（2008）

2）Stefan Gorling : An overview of the Sender Polcy Framework (SPF) as an anti-phishing mechanism, Internet Research, **17**-2, pp. 169-179（2007）

3）J. Lyon and M. Wong : Sender ID: Authenticating E-Mail, RFC4406（2006）.

4）M. Delany : Domain-Based Email Authentication Using Public Keys Advertised in the DNS (DomainKeys), RFC4870（2007）

5）Stephen Farrell : DomainKeys Identified Mail Demonstrates Good Reasons to Re-invent the Wheel, EuroPKI 2006, pp. 145-153（2006）

6）Guillermo Gonzalez-Talavan : A simple, configurable SMTP anti-spam filter: Greylists, Computers & Security, **25**, pp. 229-236（2006）

7）Eric Allman, Claus Assmann and Gregory Neil Shapir：SENDMAIL™ INSTALLATION AND OPERATION GUIDE Version 8.708

（Sendmail, Inc., 2008）Chapt. 5.1.4.18.
８）Postfix Address Verification Howto, http://www.postfix.org/ADDRESS_VERIFICATION_README.html
９）久保田浩「特集 無料ソフトからクラウドまで迷惑メール対策最新事情」日経NETWORK,（112）, pp. 45-57（2009）
10）長谷川明生，山口榮作，鈴木常彦「迷惑メールの解析」FIT2007 （第６回情報科学技術フォーラム）, pp 377-378（2007）

附録　用語説明

AAAA

DNSに登録するレコードの種別．IPv6アドレスを登録するのに使用する．これによってドメイン名をIPv6アドレスに変換できる．

DKIM

DomainKeys Identified Mailの略．メールに電子署名を埋め込む規格．電子署名と公開鍵によって送信者メールアドレスが真正かどうかを検証できる．

EHLO

SMTPコマンド．SMTP接続後，最初にクライアントがEHLOコマンドで自己のFQDNを名乗ると，サーバは使用できる拡張SMTPコマンドを返答する．

FQDN

Fully Qualified Domain Nameの略．一切省略しない完全なドメイン名のこと．末尾にも「.」を付けると上位側が略されていないことを明示できる．

HELO

SMTPコマンド．SMTP接続後，最初にクライアントがHELOコマンドで自己のFQDNを名乗ることになっていた．しかし今ではHELOの代わりにEHLOを優先して使用することが定められ，HELOはあまり使われなくなった．

MTA

Mail Transfer Agentの略．電子メールの配送を行うデーモン．電子メール利用者が直接触れるものではないが，MTAが作動していないと電子メールの配送は行われない．

MUA

Mail User Agentの略．利用者が電子メール読み書きのために直接操作するプログラム．

MX

DNSに登録するレコードの種別で，Mail Exchangerの略．電子メールの宛先ドメイン名をIPアドレスに変換する際に使用される．

RIR

Regional Internet Registryの略．世界を５つの地域に分け，それぞれの地域内へIPアドレスを割り当て管理する５つの組織．

SPF

Sender Policy Frameworkの略．発信元IPアドレスによって送信者メールアドレスが真正かどうかを検証する規格．

TLS

Transport Layer Securityの略．暗号化・認証・改竄検出の機能を有する通信規格で，通常はTCP上で使用される．以前はSSL（Secure Sockets Layer）と呼ばれていた．