社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2006－ＤＳＭ－42／2006－ＱＡＩ－20（７）
2006／7／20

スパム耐性に優れた電子メールアーキテクチャ構築に関する検討
－法人向けメールシステム開発・運用現場の最前線から－

池田 和幸†

†富士通株式会社 〒261-8588 千葉県千葉市美浜区中瀬 1-9-3
E-mail: †ikeda-kazuyuki@jp.fujitsu.com

あらまし　スパムの特徴量を利用したスパム遮断手法に関しては、ベイジアンフィルタやグレイリストに代表される各種の研究成果や実用実装が報告されている。その一方、メールシステムへのスパムメール着信数は近年増加の一途であり、その種類も多様化していることから、電子メールアーキテクチャの抜本的な見直しを行わずにスパムメールによる被害を防止することには限界が生じはじめている。

本論文では、法人向け電子メール開発・運用の最前線で活動する技術者の立場から、スパムメールがメールシステムに及ぼす問題点を再定義する。同時に、筆者が現在取り組んでいるスパム耐性の高い電子メールアーキテクチャの開発について、その検討結果を報告する。

キーワード　スパムメール対策、電子メールアーキテクチャ、大規模メールシステム、アドレス検証、DHA 攻撃

Consideration on Spam-Proof E-Mail Architecture

Kazuyuki IKEDA†

†FUJITSU Limited.　1-9-3 Nakase, Mihama-ku, Chiba, 261-8588 Japan
E-mail: †ikeda-kazuyuki@jp.fujitsu.com

Abstract　There are many researches and practical implementations on spam mail protection such as bayesian filtering, graylisting and so on. On the other hand, because number of spam mails are continuing to increase and their characteristics vary, we must review e-mail architecture to avoid damages brought by spam mails.

In this paper, from the view of IT architect who works on frontline of e-mail development and system operation, the author redefine how spam mails cause problems on e-mail system, and reports a consideration of spam-proof e-mail architecture that the author is now developing.

Keyword　spam mail protection, e-mail architecture, enterprise messaging system, address verification, DHA

1. はじめに

電子メールシステムは、現在企業活動や学術研究を進める上で欠かすことのできないインターネットインフラ基盤となっており、電子メールシステムの停止や遅延が業務に致命的な影響を与えるケースが増えている。一方でスパムメール送信手法の多様化、広帯域インターネット接続の普及により、スパムメールの受信数が急増しており、スパムメールが電子メールシステムの継続性や安定性を侵害するケースも出始めている。

米 Messaging Lab 社のレポート[1]によると、2005年に全世界で受信したメールの 68.6%、日本国内に限定しても 36.1%がスパムメールであるという統計が出ている。これらの調査からも、スパムメールの受信はメールシステムの安定運用において無視できない問題となっていると考えることができる。

スパムメールの問題点は、以下の２点を中心にこれまで議論されてきた。

1. 望まないメールを受信することによる業務の生産性低下。重要なメールの見落し、メールボックスの容量超過や過負荷によるメール受信障害に起因する機会損失の発生。
2. 望まないメールを受信することによるメールサーバ負荷の増大に起因した、サーバ資源への継続投資の発生。同時に、ユーザからのクレーム処理やサーバの安定稼働を維持するための運用要員増員が不可欠となり、組織の予算を圧迫。

最近の傾向として、エラーメールの返送やスパムメールの自動転送、スパムメールに対する自動応答機能の利用により、スパムメールを受信している被害者自身が逆に望まないメールの発信者となるケースが見られる。この場合、正常な送信メールの受付まで拒否され、業務コミュニケーションに支障をきたしてしまう事態も発生している。

上記状況から、筆者はスパムメールを業務の継続性および通信の安全性を妨害するセキュリティ上の脅威と捉えており、脅威への対応には、スパムメール対策技術の適用に留まらず、電子メールアーキテクチャ全体も含めた見直しが必要不可欠であると考えている。

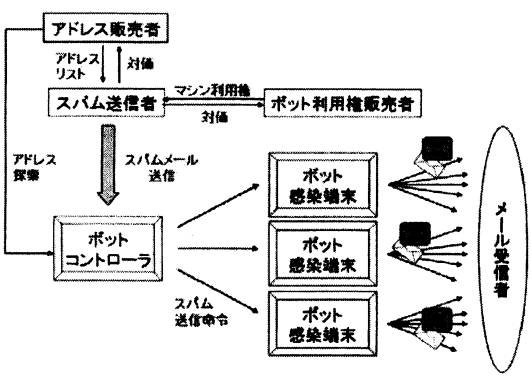
本論文では、スパム対策の主要技術を振り返った上で、スパムメールの振舞とそのメールシステムへの影響について最新の運用実績を基に報告する。その上で、スパムメールから電子メールシステムを保護するために筆者が検討している電子メールアーキテクチャ開発の取り組みを概説し、その検討内容と課題、今後の方針について論述する。

2. スパムメールの定義と対策技術

2.1. 「スパムメール」の定義

米国において 2004 年から試行されているスパム送信者の規制を目的とした法律である”CAN-SPAM ACT of 2003[2]”では、「unsolicited commercial electric mail（頼みもしない商用電子メール）」を法規制の対象としている。本論文では範囲を広げ、「受信側の都合を考慮せず一方的に送られてくる無差別に大量配信されたメール」をスパムメールと呼称することとする。

スパムメールは、かつてはスパム送信者が 1 台のマシンを使って送っていたが、現在ではインターネット上のワーム感染 PC（ボット）を遠隔操作して送信されるケースが一般的となっている。代表的なスパムメール送信方式の一例を【図１】に示す。



【図１】ボットを用いたスパム送信方式の一例

スパムメールが大規模に送信される理由は、以下の３点に集約されると考えられる。

1. スパムメール送信の仕組みを販売する者とメールを送信したい者の需要と供給が一致している。
2. 郵便や電話と比べて、情報を送信するためのコストと作業量が圧倒的に低い。
3. 電子メールが、アカウントを自由に作成・公開できるという文化の下運営されており、組織の内部に存在するメールサーバやアカウントの情報管理が杜撰な状態で運用されている場合が多い。

2.2. 主要なスパム対策技術とその課題

現在提唱されている主なスパム対策技術の特徴と利害得失を以下に記載する。

① ベイジアンフィルタ

個々の単語が過去に学習したスパムメール及び非スパムメールの文面に登場する確率を辞書に格納し、新規に到着したメールの類似性を評価関数から算出することでスパムメールか否かを判断する方式である。2002 年に Paul Graham の論文[3]にて提唱された。

到着メールによる学習を繰り返すことでスパム精度の向上が期待できる半面、文章の短いメール、多くの一般語や業務用語を含むメールの判定は困難である。辞書の作り方、評価関数に用いる変数の抽出方式の最適化が課題となっている。

② ヒューリスティックフィルタ

メール本文やヘッダからスパムメールの特徴を抽出して得点付けを行い、得点の総和と閾値を比較してスパムメールか否かを判定する方法である。オープンソースの SpamAssassin[4]が本方式を採用した代表例である。

特徴量を増やすことで高い検出率と低い誤検知率を達成できる一方、時々刻々変化するスパムメールの特徴に追従するためのコストと、オプトインメールのような、外見上はスパムメールと酷似したメールの判定に課題が残る。

③ グレイリスティング

グレイリスティングとは、一旦全てのメールに対して RFC2822 で規定された Transient Negative Completion Reply（一時拒否エラー：コード 4YZ）を返し、同一の差出人／受取人／接続元 IP アドレスを持つセッションを受け付けることで、スパムメールを効率的に遮断する方式である。ボットから送信されるスパムメールが再送処理を行わない性質を利用している。

特に大規模環境において、再送遅延による業務影響の回避するためのホワイトリストの維持管理が運用上の課題として残る。

④ IPアドレスベースのブラックリスト

不正中継を許可するメールサーバ、スパム送信の多いメールサーバの IP アドレスやネットワークアドレスを列記したブラックリストをメール受信時に参照し、レコードに一致する送信元 IP アドレスからのメールを遮断する方式である。

リストをサイト管理者が作成管理することは事実上不可能であるため、spamhaus[5]など有志によってメンテナンスされているものを購読することが一般的であるが、ブラックリストの内容の正確性を保証することが難しいため、誤認識を回避する方法を運用側で考慮する必要がある。

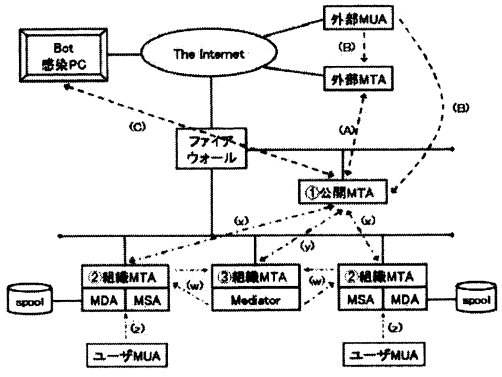
⑤ クライアント側の POP3 Proxy フィルタ

ベイジアンフィルタ技術を使用した POP3 Proxy をクライアント側に準備し、メール受信時に経由させる方式である。自身のメールのみを学習に用いるため、高いメール振り分け効果が期待できる一方、スパムメールの排除を MTA で実施しない場合には、スパムメールがサーバに及ぼす問題の解決にはならない。

現時点では、スパム対策技術の適用によって 100%の検知率と 0%の誤検知率を同時に満たすことは事実上不可能といってよい。大規模なメール環境においては検知漏れしたスパムの絶対数自体が無視できないため、検知漏れメールが引き起こす影響についても極小化するよう、電子メールアーキテクチャを設計する必要がある。

3. 法人メールシステムにおけるスパムメールの振舞と影響

【図 2】は、典型的な DMZ 構成を採用した場合の典型的なメールシステム構成であり、多くの法人で採用されていると考えられる形態でもある。



【図 2】DMZ 構成を採用したメールシステム構成

【図 2】の構成をベースに、近年のスパムメールの振舞とメールシステムに与える影響を分析する。

① 広域分散ボットからのスパムメール受信

【表 1】は、弊社でスパム対策の業務運用を委託されているお客様 A 社に届く、スパムメール数とスパムメールの送信元 IP アドレスの総数である。

【表 1】スパムメール送信 IP アドレス数
(2006/05/22～2006/05/28 の 7 日間で測定)

日付	メール遮断数	送信元 IP アドレス数
5/22	221498	55926
5/23	224003	59892
5/24	196517	50493
5/25	195396	52721
5/26	222594	46716
5/27	181673	36497
5/28	192495	43856
合計	1434176	346101
1 週間に接続した IP アドレスの総数		293120

1 週間のうち 2 日以上スパムの送信に使用されたボットが全接続 IP アドレスの僅か 15.3%、各ボットが 1 日当たりに送信したメール数の平均は僅か 4.14 通/日、最も大量のメールを送信したボットであっても、860 通/日の送信に留まっているという統計が得られた。

このことから、スパム送信者は大多数のボットをその日限りで使い捨てし、ボットを渡り歩きながら少量ずつターゲットにスパムメールを送信する「広域分散」的な送信を行うことで、ブラックリスト作成の難易度を上げて遮断を回避し、同時に証跡を追うことを困難にしていると推測できる。

② 宛先ユーザが存在しないメールの急増

ボットからの DHA(Directory Harvest Attack)攻撃や無差別に送信されるスパムメールが増加し、メールシステム内に宛先ユーザが存在しないメールが増加している。

【表 2】は、弊社でメールシステムの業務運用を委託されているお客様 B 社に届く、全メール数とメールボックスが存在しないメール数の統計である。

【表 2】宛先が存在しないメール数
（2006/05/22～2006/05/28 の 7 日間で測定）

日付	メール着信数	宛先ユーザの存在しないメール数
5/22	17500	3372
5/23	18251	3139
5/24	17401	3016
5/25	16046	2737
5/26	16203	3383
5/27	5534	1848
5/28	3602	1399
合計	94537	18894

着信したメールの約 20%が、実際に存在しない宛先を少なくとも 1 個持っている。

【図 2】において、(A)の経路で公開 MTA にてメールを受信し、(x)または(y)の経路で内部 MTA に配送を行い、メールボックスがないことが判明してエラーメールが(A)の経路で戻る場合、外部 MTA で中継しなければならないメールの総数は以下の数式で表現される。

$$\text{増幅率} = 2(1+r_u) + r_u r_b \left(\frac{t_e}{t_b} - 1\right) \quad \cdots \text{（式 1）}$$

ここで、

r_u：宛先ユーザが存在しないメールの割合

r_b：バウンスメールの受信を拒否される割合

t_e：メールキュー内におけるメール保持時間

t_b：メール再送間隔

である。一例として、宛先ユーザが存在しない着信メールが 20%、バウンスメールの受信を拒否するサイトが 1%、メール保持時間 5 日、キューの再送間隔 4000 秒（postfix MTA の標準値）とすると、着信メール 1 通に対する増幅率は 4.558 となる。理想的な状態では増幅率は 2.0 であることから、宛先ユーザの存在しないメールが着信することで、サーバに大きな負荷を掛けていることが推測される。

③ メールボックスの拡張機能を介在した二次影響

フリーメール業者や企業でスパム対策が進んだ結果、少量のスパムメールが外部に自動転送されただけで IP アドレスレベルの遮断をされてしまうケースが増加している。同様の現象は、スパムを大量に受信するユーザが自動応答設定を行っている場合においても発生する。

本現象はスパムメール受信の二次被害であるとみなすことができるが、スパムメールの受信同様正常なコミュニケーションに影響を及ぼすため、現在与えられているメールボックスの操作権限についても、再度適切な管理を見直す必要がある。

4. スパム耐性の高い電子メールアーキテクチャの開発に向けて

本章では、スパムメールからメールシステムを守るために、筆者にて開発を行っている電子メールアーキテクチャについて紹介する。

4.1. 開発目標

以下 3 点の指針で表現される「スパム耐性」能力を高めることで、電子メールアーキテクチャの Best Practice を構築することが開発目標である。

① スパムを受けにくい

アドレス調査およびスパム送信をスパマーにとって非効率にすることで、スパム送信者に狙われにくいシステムを実現する。

② スパムを出さない

意図的なスパムメールを発信しにくくする。同時に、スパムメールを受信した結果発生する外向きのメールを抑制し、接続先サイトの規制対象となることを防止する。

③ スパムを中継しない

適切な権限制御を実施し、スパムメール発信の温床となることを防止する。

4.2. 検討中のアーキテクチャ

RFC2505[6]において、MTA においてスパムを抑止するための BCP（Best Current Practice）を主に以下の観点から論述している。

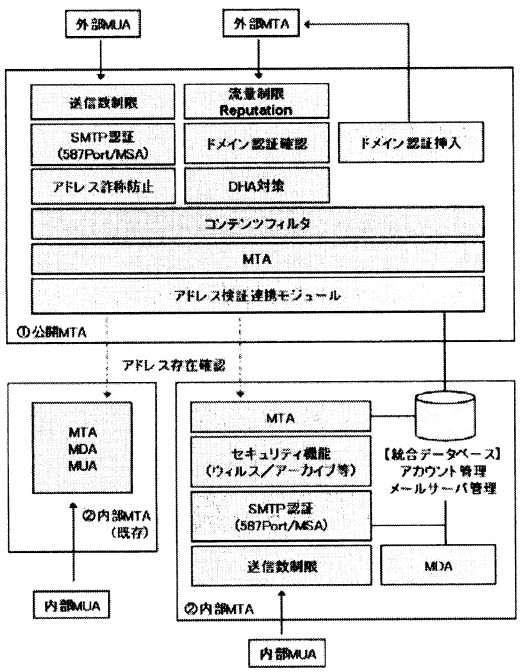
1. 許可されないメールの中継を抑止する。（2.1 節）
2. 調査のためメールの証跡を残す。（2.2/2.3 節）
3. ホスト、送信者およびそのグループをキーとしたメールの受信拒否を行い、その証跡を残す。（2.4/2.5/2.7 節）
4. レートコントロールを行う。（2.8 節）
5. 送信者アドレスの検証を行う。（2.9/2.10 節）

本節では、スパマーによるスパムメールの送信傾向を踏まえて RFC2505 を拡張することで、スパム耐性に優れた電子メールアーキテクチャを検討する。

3 節で分析した影響から、現在のメールシステムにおいて新たに考慮すべき検討事項として、以下の 4 点を抽出した。

① メール受信者が実在ユーザでない場合に発生するバウンスメールを抑止し、スパムメール受信の二次被害を防止する。
② 内部のアドレスをスパム送信者がリスト化し、スパムメール送信に利用することを抑止する。
③ 大量のメール送受信およびなりすまし送信を防止する。
④ 組織内に存在するメールアカウントおよびメールサーバを把握し、管理を野放しにしない。

上記課題を解決するために、【図 3】の構成を持つ電子メールアーキテクチャを提案する。この構成は、RFC2505 で提示されているスパム抑止観点のうち、2.1 節、2.4 節、2.5 節、2.7 節、2.9 節、2.10 節を拡張したものである。

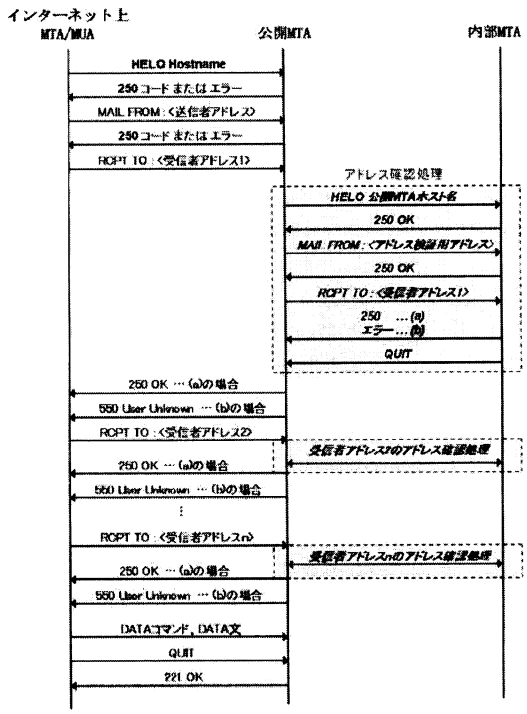


【図 3】スパム耐性の高いアーキテクチャ

各モジュールでは、以下の制御を実施する。

【アドレス検証連携モジュール】外部 MTA におけるメール受付前に、各回送先におけるメール受信者の実在確認（以下、「アドレス検証」と呼ぶ）を行うことで、実在しない宛先に対しては電文中で永続拒否エラーを返し、無差別な大量送信によるバウンスメール発生を抑止し、メールサーバの負荷を軽減する。一例として、20%の宛先が実在しない場合、(式 1)で計算される増幅率は 1.6 となり、アドレス検証を実施しない場合の僅か 35%のメールを中継すればよいことになる。

アドレス確認処理のフローを【図 4】に示す。アドレス確認処理は、RFC2505 の 2.1 節に記載されているメール中継制御を宛先メールアドレスレベルに拡張したものである。



【図 4】アドレス確認処理フロー

【DHA 対策】メールシステム内部の実在アドレス探索によるスパム送信先リスト作成を抑止するため、同一セッション中のエラー回数による切断や遅延の導入、単位時間当たりのメール処理数制限を実施する。

アドレス確認処理の結果返される永続エラーをエラー回数に含めることで、不当なアドレス探索の試みを効果的に遮断することが可能になる。

【流量制限、Reputation】メール受信時に、接続元 IP アドレス単位の流量制限を行う。接続元 IP アドレスの振舞により単位時間当たりの 0 通から制限なしまで変化させることで、スパムを送信する IP アドレスからのメールを優先的に拒否することが可能となる。

【SMTP 認証】メール送信時に ID とパスワードを用いて認証を行うことで、送信権限を制御する。また、同一認証 ID から送信できるメール数をカウントすることで、発信メール数の制限を実施し、悪意を持った発信者のスパム送信被害を極小化する。

【アドレス詐称防止】 SMTP 認証 ID とエンベロープ送信者を関連付け、両者の組み合わせが正しい送信メールのみを通過させる。

【ドメイン認証確認】 Experimental RFC および IETF Draft として提出されているドメイン認証技術を実装し、送信時の署名と受信時の署名確認を行う。現時点では各技術が確定したものではないため、認証の結果によってメールを遮断することはできない。

【統合データベース】 内部 MTA に存在するドメイン、アカウントおよび、既存 MTA でホスティングするドメイン名、IP アドレスをリスト化したものを格納する。提案するメールアーキテクチャの各モジュールから参照することにより、アドレス検証やユーザのアカウント操作権限の統一ポリシーによる制御を行う。

4.3. プロトタイプ実装

4.2 節で提示した検討結果が現実的なコストで実装可能であることを立証するため、【表 3】に示す環境においてプロトタイプを実装した。MTA としては、大規模システムへの導入が可能であり、Linux の標準ディストリビューションとしても採用されるなどメール市場における普及も進んでいる postfix を採用している。

【表 3】プロトタイプ実装環境

項目	概要
ハードウェア	Sun Fire V100
OS	Solaris9
MTA	Postfix 2.2.10
認証モジュール	Cyrus-sasl 2.1.21
統合データベース	MySQL 4.1.19
ドメイン認証モジュール	dkfilter 0.7
内部 MTA	sendmail 8.11.6 postfix 2.1.5 計 5 台

アドレス連携機能は、postfix の verify server 機能を拡張し、予め統合データベースに投入したドメインごとの中継先に対して、アカウントの実在を確認した後にメールの実データを送信する方式で実装した。内部 MTA として想定した 5 台に対するアドレス検証が、通常の SMTP 通信を阻害しない応答時間で実施可能であることを確認した。

DHA 対策は、postfix の anvil 機能と smtpd デーモンが標準で持つエラー制限機能を利用し、エラー回数が 5 回を超過した場合に通信を切断、30 分当たり 50 接続までの接続を受け入れる実装を行った。この実装では、30 分当たり最大 250 回のアドレス問い合わせしか許可されないため、辞書的な探査は実用的な時間で効果を発揮しない可能性が高い。

SMTP 認証においては、予め統合データベースにメールアドレスと認証 ID, パスワードの組み合わせを登録し、メールアドレスと認証 ID の組み合わせ一致を送信可能条件とすることで、アドレス詐称を防止した実装を行っている。同時に、リアルタイムに送信ログを解析して統合データベースの送信可能フラグを制御するバッチを組み込むことで、認証 ID あたりの送信可能数を制限している。

本システムでは、全てのアカウント権限／サーバ情報を統合データベースで一括管理しているため、統合データベースのレコード制御によりメールシステム全体の統制を図ることも可能である。

5. 今後の方針と課題、提言

スパムメールは増加の一途をたどっており、その送信方法や内容も多様化しているため、スパムメールをテクノロジーの組み合わせで 100%防御することは事実上不可能である。

筆者は、大量の電子メールを受信する環境において、遮断に失敗したスパムメールがメールシステムを用いた業務の継続性、通信の安全性を妨害するセキュリティ上の脅威になっているという認識から、アドレス検証という概念を中心に据えて既存の電子メールアーキテクチャを再構築することを提案した。また、プロトタイプ実装を通じて提案の実現性を議論した。

今後は、大規模環境向けのプロトタイプ検証および実際のトラフィックシミュレーション、ユーザ設計への適用を通じて、電子メールアーキテクチャの Best Practice を確立する予定である。

最後に、組織内のメールアカウントやメールサーバの管理を杜撰にしていることで、スパムメール被害が拡大する傾向が今後顕著になると考えられる。メールシステムのアーキテクチャ同様、メールアドレスの管理も同様に見直すことを、この場を借りて提言したい。

文　献

[1] MessageLabs Ltd, “MessageLabs Intelligence 2005 Annual Security Report”, Dec. 2005. http://www.messagelabs.com/publishedcontent/publish/threat_watch_dotcom_en/intelligence_reports/DA_114080.chp.html

[2] “CAN-SPAM ACT of 2003 (Public Law 108-187)” http://www.spamlaws.com/federal/can-spam.shtml

[3] Paul Graham, “A plan for spam” http://www.paulgraham.com/spam.html

[4] http://spamassassin.apache.org/

[5] http://www.spamhaus.org/

[6] G.Lindberg, “Anti-Spam Recommendations for SMTP MTAs”, Feb. 1999 ftp://ftp.rfc-editor.org/in-notes/rfc2505.txt