電子メイル配信方法とメイルヘッダ分析によるスパムメイル対策

2001MT020 服部 一徳　　2001MT107 若森 孝志
指導教員　後藤 邦夫

1　はじめに

近年，ネットワークの世界で spam の急激な増加が問題視されている．そして多くのメイルユーザは，spam が悪影響を与えていると報告している．

本研究では，2 つの手法を提案し実験によりそれらの効果を明らかにする．1 つ目はシステム管理者が行うフロントエンド MTA での対策であり，2 つ目はエンドユーザが行うメイルヘッダ分析を実装した POP proxy での対策である．

ここでのフロントエンド MTA の役割は，一見 (いちげん) さんお断り方式 [1][2][*1]を応用した Greylisting[3] に独自の情報を追加した方式により，spam 対策用 MTA として spammer からの SMTP 接続を防ぐことである．この方式は，SMTP セッションにおいて MTA が一時エラーを返すと，spam ホストは再送してこないという特徴を利用し，初めての接続に対して一時エラーを返しある間隔をあけて再送してきたもののみを受け取るものである．

POP proxy での対策は，まず中継 MTA でつけられる Received フィールドに着目し偽りがないかチェックして，spam 確率を求めヘッダに書き込むことである．ユーザは判断材料やメイルソフトでの振り分けの材料とする．この 2 つを同時に稼働させることで，より効果的な spam 対策になると考える．また，本研究では実験のために spam 受信用の専用のアドレスを用意し，そのアドレス宛に故意に大量の spam メイルを spammer に送信させた．

服部は主に配信方法による対策を，若森がメイルヘッダ分析による対策を担当した．

2　提案する対策

この節では，現状の対策と課題を明らかにし本研究の位置付けを示す．

2.1　現状の対策

図 1 に示すように，現状の対策は大きく分けて 3 つある．配信方法による対策，メッセージ内容による対策，また spam 問題の抜本的解決となる発信者認証である．

配信方法による対策では，受信拒否されると再送しない，DNS の逆引き情報が定義されていないといった spammer から発せられる SMTP セッションの特徴を利用する．一見さんお断り方式，Greylisting では，初めての接続に対して一時エラーを返しある間隔をあけて再送されたメッセージのみを受け取る．spam メイルを受信する前に拒否できるが，再送を待つのでメイルを受け取るのは少し遅れる．

メッセージ内容による対策では，Subject:やメッセージ本文に含まれる spam メイル特有の単語を検出し，その有無によって spam かどうかの判定を行う．配信方法について制限する必要はないが, MTA が一旦メッセージを受信しなければならない．また，Bayesian filtering[4] ではメイルを処理すればするほど spam のデータベースは肥大化するため，フィルタにおけるデータベースの管理も必要となる．したがって，spam の件数，メッセージ長が非常に大きくなると MTA や MUA の資源の浪費が顕著になる．

発信者認証は，SMTP にユーザ認証がない点を補うために，各ドメインが運営する MTA において利用者認証に基づく利用者の真正性に関する情報を追加することで実現でき，いくつかの方法が提案されている．この発信者認証による対策は，広く普及すれば大きな効果が期待できるが，統一基準がないことや普及に時間がかかるなどの問題があり，まだ実用的でない．

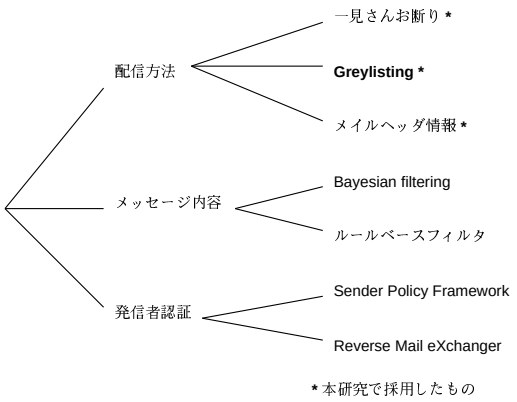

図 1　現状の対策の分類

2.2　提案するシステムの構成

本研究では，抜本的対策ができない現状における best current practice として，spam 対策専用 MTA を実際の MTA の前にフロントエンド MTA として置き，そこで Greylisting に独自の spam 判定条件を追加した方式を実装し実験するという手法を取る．独自の情報とは，接続時間間隔，接続ホストに対する DNS チェックとパラノイドチェックである．

また，MUA がメイルを受信する際に spam 判定基準を与えることによるユーザ側での spam 対策を組み合わせることで，より効果的な spam 対策を目指す．図 2 に

[*1] 京都の高級なお店などでよく言われる「一見さんお断り」からきている．一見さんとは「初めてのお客さん」「(お店の人が) 知らないお客さん」のことで，そのようなお客さんは受け入れないお店がある

本研究で作成するシステム構成を示す．

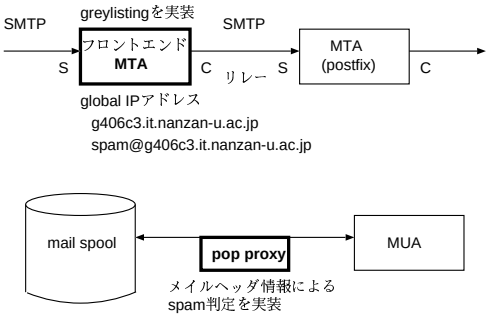

図 2　本研究でつくるシステムの構成

3　MTA における対策の実現

以下に配信方法による対策のシステム概要図 (図 3) を示す．antispam-smtpd で SMTP 接続をうけ antispam-client で postfix と通信する．antispam-resending は，何らかの理由で postfix との通信に失敗したメイルを再送するものである．よって，postfix という一般的な MTA ソフトウェアの前に設置することにより 2 節で述べたように簡単に spam 対策ができることを目指し設計した．

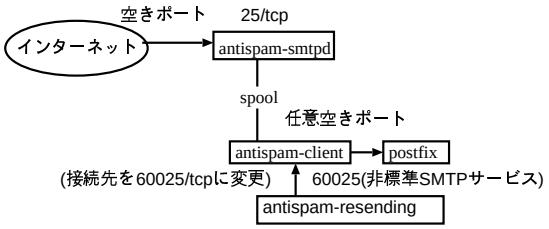

図 3　フロントエンド MTA のシステム概要

本研究では，接続ホストの状態を 4 つと状態遷移条件を定義した．各状態の遷移条件は OR である．以下に状態と遷移条件を記す．なお，始めて接続してきたホストは「一見さん」として状態 grey とする．

- black: spammer の可能性大 接続を途中で切る
  - black への遷移条件 (white, dark, grey から)
    6 時間以上再送がない
    再送時間が 6 分 30 秒より短い
    DNS が引けない
    宛先の数が 50 件以上
    管理者が blacklist に登録する
- dark: spammer の可能性あり 応答を遅らせる
  - dark への遷移条件 (grey から)
    再送時間が 6 分 30 秒以上 30 分以内
- grey: 通常のホスト 一時エラーを返す
  - grey への遷移条件 (black, dark, white から)
    whitelist，blacklist にない初めての接続
    再送時間が 30 分以上
- white: whitelist 通常の SMTP 応答をする
  - white への遷移条件 (black, dark, grey から)
    管理者が whitelist に登録する

プログラムを Perl で記述し，DB として Postgresql を利用した．あらかじめ，SMTP の接続記録を一時的に保持する DB テーブル (tempdata) と「一見さん」をお断りするための DB テーブル (greydata) と MTA のスプーリング機能としての DB テーブル (spool) と接続ホストごとの情報を保持する DB テーブル (hosts) を用意した．greydata，tempdata 両テーブルに関して一意な情報を決めるものとして接続先アドレス，送信元ドメイン，宛先アドレスを，spool テーブルには MTA がメイルを一意に決定するもとして定める Message-id をプライマリキーとして設定した．なお実験 MTA であるので，SMTP のコマンドも最低限必要な HELO(EHLO), MAIL，RCPT，DATA，RSET，QUIT のみを実装した．また，SMTP セッション内で接続クライアントに対し DNS 逆引きとパラノイドチェックを行っている．

次に SMTP サーバとして機能する antispam-smtpd の主な処理を SMTP セッションの流れに沿って以下に示す．なお始めに作成サーバの動きを，括弧でコマンド名とクライアントの流れを説明する．4 はフローチャート (図 4) を用いて説明する．

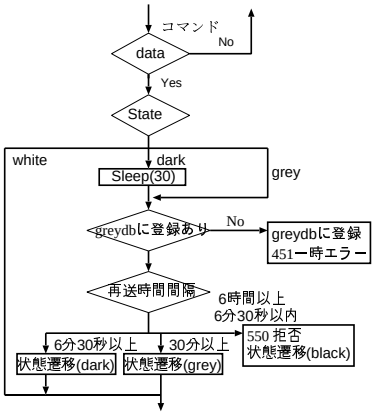

図 4　DATA コマンドのフローチャート

1. 接続 IP アドレスを接続時間，blacklist，whitelist にないか，また前に接続があれば状態を検索し DB に保存し応答する (HELO(EHLO): SMTP クライアントが SMTP サーバに接続し，ホスト名を名乗る).
2. 指定された送信元ドメインに関して DNS が引けるか確かめ結果に応じた応答を返し，DB に送信元アドレスと DNS の結果を保存する (Mail: 送信元アドレスを指定).
3. 宛先の数を確認し結果に応じた応答を返し，DB に宛先アドレスを保存する (Rcpt: 宛先アドレスを指定).
4. 図 4 で示した部分．まず接続ホストの状態を調べる．black の場合は永久エラーを返し，white の場

合は，通常の本文入力応答を返す．その後，テーブル greydata に現接続の IP アドレス，送信元アドレス，宛先アドレスがすべて一致するものがあるか検索する．一致するものがなく状態が grey，dark の場合は一時エラーを返す．一致するものがある場合，前接続時間と現接続時間間隔により状態遷移，状態に応じた応答をする (DATA: 本文の入力)．

5. 真の MTA である Postfix に受け取ったメッセージを送信する．

4 メイルヘッダ分析による対策の実現

電子メイルには配信経路などの情報が書かれたヘッダ部分がある．本研究は，この部分を分析する POP proxy を実装することによりユーザやソフトウェアに spam 判定基準を与えることを目的とした．

4.1 POP proxy

POP proxy とは，メイルヘッダに書かれている情報のうち，信用できる自ドメインのメイルサーバが他から受信したときに MTA でつけた Received フィールドと宛先，送信元を抜き出して分析し，X で始まる実験ヘッダ行を追加する．追加した X で始まるヘッダ行の例を図 5 に示す．

```
X-HeaderCheck: F:yahoo.com MX:mx4.mail.yahoo.com(4)
  HELO:m6o.mlb2r9.net FQDN:NOFQDN IP:207.204.245.7
  NUMTO:1 T:iq.nanzan-u.ac.jp NUMCC:1 HOST:NOFQDN
  by:acdc.iq.nanzan-u.ac.jp  W:0
X-HeaderAnalysis: FROM reachability 2 (MX),
  senderhostmatch 0/2,TO senderhost-match 4/4
X-Spam-Prob: 1.00 (ML or Alias: 1 - (reach)1 * (from)0)
```

図 5　X-header の例

1. X-HeaderCheck: Recieved フィールド内の情報を分析し以下のように表示する．
   - F: A レコード MX: MX レコード (MX の数)
   - HELO: HELO を実行したホスト名
   - FQDN: 接続ホストの FQDN
   - IP: 接続ホストの IP アドレス
   - NUM TO: 中継回数
   - T: To で指定したドメイン
   - NUMCC: Cc の数
   - HOST: 中継ホストのドメイン
   - by: 受信 MTA のドメイン
   - W: whitelist に該当すれば 1，しなければ 0
2. X-HeaderAnalysis: ヘッダ情報に偽りがないか分析し以下のように表示する
   - FROM reachability: DNS 調査結果を表示．送信者が信頼できるか確認する．メイルが届く可能性があれば信頼できるホスト (MX，A レコードが引ける)，なければ信頼できないホストとする (DNS 情報がひけない)．
   - senderhostmatch，to senderhost-match: ヘッダに偽りがないか表示．
     送信元ドメインと送信 MTA のドメイン，宛先ドメインと受信 MTA のドメインを「.」で区切り後ろから同じ単語か調べる．
3. X-Spam-Prob: 1，2 の結果を定義した実験式に代入し spam 確率を求め表示させる．

4.2 spam 確率実験式

前節で示した X-Spam-Prob について説明する．Received フィールドを分析した結果より 4 つの式を提案しそれぞれに関して ROC 曲線を描き (図 6) 判定式の性能を比較する．以下に式を示す．

$$SpamProb = 1 - (reach \times from_{match}) \quad (1)$$

$$SpamProb = 1 - (reach \times from_{match} + to_p) \quad (2)$$

$$SpamProb = 1 - (reach \times to_{match}) \quad (3)$$

$$SpamProb = 1 - (reach \times to_{match} + f_p) \quad (4)$$

- $reach$: DNS 問い合わせの結果が MX なら 1 を，A なら 0.75 を，none ならば 0 を代入
- $from_{match}$: From のドメインと発信ホストのドメインの合致率を代入
- to_{match}: To のドメインと受信ホストのドメインの合致率を代入
- to_p: to_{match} の値により与えられる重みを代入
- f_p: $from_{match}$ の値により与えられる重みを代入

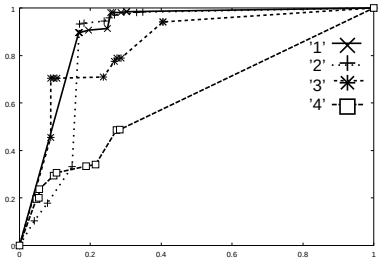

図 6　ROC 曲線 (横軸: 誤判定率 縦軸: 判定率)

ROC(Receiver Operation Characteristic) カーブとは，横軸に False-Positive Rate(誤判定率)，縦軸に True-Positive Rate(判定率) をとり，spam 判定基準つまり spam とみなす (1)〜(4) 式のしきい値を変化させたときの結果より描くことができる曲線である．この曲線は，式の有効性の比較を図で行うものである．本研究は，判定率が高く，かつ誤判定率が低いことが理想である．また推定する稼働状況は誤判定率 0.1〜0.2 とした．判定基準としては，曲線下面積の比較があげられる，よって (3) 式，(4) 式は除外される．また稼働状況を考慮にいれると (2) 式が (1) 式に比べ誤判定率 0.1 付近で判定率が低くなる．結果 4 つの式の中で最も効果的であるとして (1) 式を採用した．しかし，他の判定方法では他の式がよい可能性もある．

5 実験結果と評価

以下に本研究で提案した 2 つの手法の実験による評価を述べる．

5.1 MTA における対策の評価

2004 年 9 月 28 日から 2005 年 1 月 5 日までの 100 日間のデータを分析した結果を，表 1 から表 2 に示す．

表 1 に示すように，この間の全セッション数 36575 のうち，SMTP で blacklist ホストに永久拒否エラーを返した数が 91.2%，一時エラーを返した数が 2.3%，メイルを受け取った数が 6.3% である．以上より，91% の SMTP セッションを spam として拒否に成功したと言える．

表 1　SMTP セッションログ内訳

result	session	rate
永久拒否をした回数	33337	91.2%
一時エラーをした回数	858	2.3%
メイル受信数	2299	6.3%
その他	81	0.2%
総接続数	36575	100%

表 2 に示すように，blacklist に登録された全 222 のホストのうち，DNS 未登録のため状態が black に遷移した数が 3.6%，再送時間間隔が 6 分 30 秒以内だったため black に遷移した数が 32.9%，一時エラーを返した後再送がなかったため black に遷移した数が 63.5% である．black に遷移する条件として，宛先の数が 50 以上あった時も状態を black に遷移させるよう定めたが，結果として宛先を 50 以上指定してきたセッションはなかった．以上より，再送時間間隔が短いホスト，再送がないホストからの SMTP 接続を拒否することにより spam メイル受信拒否に大きな効果があることが示せた．

表 2　black に遷移した原因

原因	num	rate
DNS 未登録	8	3.6%
宛先の数	0	0%
再送時間間隔	73	32.9%
再送がない	141	63.5%
All	222	100%

5.2 メイルヘッダ分析による対策の評価

表 3，表 4 に示すのは前節で最も効果的であると示した式 (1) を採用した POP proxy に非 spam メイル 34000 件，spam メイル 13000 件を通過させた結果である．これより True Positive，すなわち spam が spam であると判定したものが 90%， また False Negative，すなわち spam であるメイルを誤って spam でないと判定したものは，1% 以下と spam を適切に判定できたので，効果的であると言える．なお，表の isvw とは大学が設置しているウイルス対策ゲートウェイを通過したもので Received 行が正しくないので除外した分である．

表 3　非 spam(30000 件)

Prob	rate
1	16.8%
0.9	0%
0.8	0.20%
0.7	2.55%
0.6	5.34%
0.5	0.97%
0.4	1.02%
0.3	0.49%
0.2	1.57%
0.1	0.03%
0	69.76%
isvw	24.45%

表 4　spam (13000 件)

Prob	rate
1	89.68%
0.9	0%
0.8	0.02%
0.7	0.99%
0.6	0.68%
0.5	0.08%
0.4	0.01%
0.3	0.09%
0.2	0.22%
0.1	0%
0	1.61%
isvw	31.29%

6 おわりに

本研究でのフロントエンド MTA で行う対策では，spammer からの接続を 91% の割合で SMTP 接続の時点で拒否できた．これは，同じ条件で実験を行ったわけではないので単純比較はできないが，TIC における MTA による spam 対策の実践報告 [2] で行われた一見さんお断り方式を利用した spam 対策での接続拒否 86% と比べ向上している．このことより，本研究のように再送時間間隔や DNS 情報を追加することでより多くの spam 拒否が可能となった．加えて，Bayesian filtering と違いメイル受信の数を著しく減少できた．しかし，本方式は現在の spammer の特徴を前提に構成したものであり，spammer が本方式で利用した特徴を逃れる場合には対応できない．

メイルヘッダ分析では，各 MTA でのヘッダの書き込みが RFC の基準を守っていないものが多く誤判定の要因となった．しかし，結果として spam を spam と判定できた．課題として spam 判定実験式を再考し，論理的根拠を与える必要がある．

現在様々な対策が考えられているが全ての MTA が RFC の基準を守っていれば，対策の効果が向上すると本研究で実感した．

参考文献

[1] 前野年紀: boycott spam mail, http://spam.qmail.jp/tfail.html，2004.

[2] 鈴木常彦，後藤邦夫，山口榮作，石川雅彦: MTA による spam 対策の実践報告，情報処理学会研究報告，2001-DSM-34(11) pp.61-64，2004.

[3] Evan Harris: Greylisting: http://projects.puremagic.com/greylisting/, 2004.

[4] Paul Graham: A Plan for Spam, http://paulgraham.com/better.html/，2003.