電子メイルヘッダの調査による spam メイル判定の提案

2008MI007 青山 尚樹
指導教員　後藤 邦夫

1　はじめに

近年，インターネットサービスの普及に伴い，その安全性や信頼性が問われている [5][4]．特に spam(スパム) メイルと呼ばれる迷惑メイルについては，個人情報の流出などがメディアで多く取り上げらている．

本研究では，昨年度の研究 [3] で完成しなかった spam 判定プログラムを完成させ，さらに品質の向上を目指す．また，送信方法が不正なメイルを検出する．

spam メイルかの判定をするために，blacklist との照合，SPF レコード，MX レコードによる判定，Domain Name System(以下，DNS) による逆引き正引き判定の 4 種類のルーチンを実行する．本研究では spam メイルを判別するために，リアルタイムキャプチャではなく，溜め込んだメイルのヘッダを利用する．

リアルタイムキャプチャはリアルタイムで情報が得られることが挙げられる一方，ヘッダにある情報の読み取りが困難であることや，運用が困難で実験ができない．それに対しメイルヘッダは，判定済みの蓄積データがあり，Date，subject などその他の情報も使い易いメリットがあるため，本研究ではメイルヘッダを使用する．

2　システムの概要

この節では，本研究のシステムの概要の基本的な考え方について述べる．

2.1　概要

メイルヘッダの内容の一例を図 1 で示す．ヘッダから読み取れる基本的な情報を抜粋した．

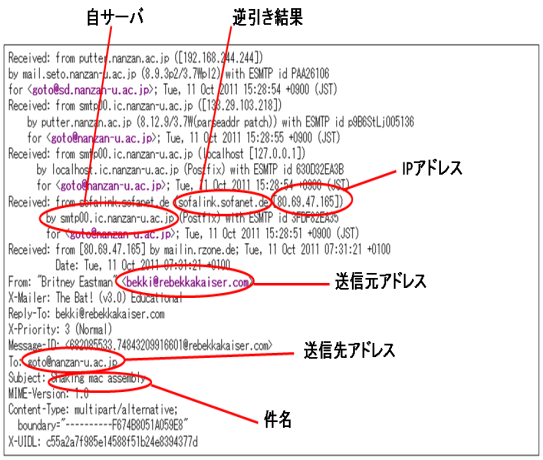

図 1　メイルヘッダ情報

2.2　spam 判定方法

spam 判定プログラムを作成し，その実行結果をもとに総合判定を考察する．本研究では，単純加算算術による判定と加重算術平均による判定の 2 種類を出力し各判定結果を比較する．実験のために溜め込んだメイル適切なポイント付けと重みつけを，実行結果の統計から決定した．From 行がない，アドレスが存在しない，SPF レコード調査で fail または softfail の場合は送信方法が不正と考え spam メイルと判断する．また，SPF レコード調査で pass の場合，正しい送信方法であると考える．

判定方法の流れを図 2 に示す．

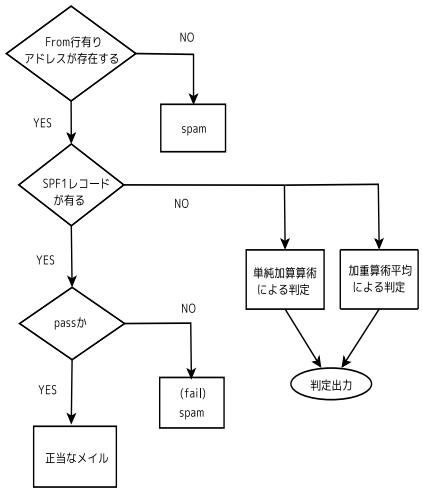

図 2　判定方法のフローチャート

4 種類の判定ルーチンについて説明する．

- blacklist 照合による判定 [1][2]
  blacklist は spam に関係するアドレスの一覧を公表した DB である．照合方法は，メイルヘッダから読み込んだ IP アドレスを逆順にし，それを各 blacklist に登録されているか調べる．例えば spamhaus の場合，IP アドレスが A.B.C.D の場合，D.C.B.A.zen.spamhaus.org の A レコードを検索し，Address(127.0.0.x) が得られれば，blacklist に登録されている．RBL.JP の照合方法も同様に，D.C.B.A.all.rbl.jp を調べる．Address が返ってこない場合は登録されていないことになる．

- DNS 逆引き正引きによる判定
  送信元の IP アドレスとドメイン名をチェックする．PTR レコードと A レコードを取得し，その

ホスト名の一覧を印字し，正引き，逆引きをして照合する．逆引きでは，逆引きした結果をさらに正引きし元の IP アドレスと合致するか調べる．DNS の機能はそれに加えて，メイル配送 (MX レコード)，SPF 調査 (TXT レコード) など他のルーチンでも利用する．

- DNS(SPF レコードによる判定)
  SPF レコードは電子メイルにおける送信ドメイン認証の仕組みである．From 行のドメインと，Received 行の相手の IP アドレスから SPF 調査ができる．
- DNS(MX レコードによる判定)
  From 行アドレスドメインの MX レコードの有無を調べ，判定の参考程度に考える．

3 システムの実現

この節では spam メイル対策として実行している判別処理の仕組みについて説明する．

3.1 システムの構成

本研究では，OS に Ubuntu10.04LTS を使用する．プログラムは，文字列の抜き出しが容易な Perl スクリプトを用いる．

3.2 実行の手順

以下のことを実行する．

1. メイルの読み込み，ヘッダ抽出
2. spam メイル判定
   - RBL.JP，spamhaus で提供されている blacklist と照合
   - 提供されている DNS モジュールの利用
     Net::DNS モジュールを使いドメインの DNS の MX，A レコード，PTR レコードを調べて各ルーチンをチェック
   - 提供されている SPF モジュールの利用
     Mail::SPF モジュールを使い送信元ドメインを認証
3. spam メイル判定の結果から判別

4 実験と評価

あらかじめ用意した spam メイル 500 通と，spam でない通常のメイル 500 通を spam 判定プログラムで実行し，統計をとった．実験結果を表 1，表 2 に示す．

表 1 の結果により，通常メイルの多くは SPF レコードが存在し調査結果が pass であることがわかった．

表 2 の結果から各判定ルーチンの重要度を考察した．単純加算算術による判定と加重算術平均による判定の 2 種類の総合判定をした．その結果，2 種類の総合判定結果が近似したことから，各ルーチンに適切なポイント付けと，重みをつけることができたと考える．加重算術平均は，個々のデータの重みが違うときに使われるため，本研究で有効であると考え使用した．各総合判定を出力し比較することにより，より正確な spam 判別が可能になったと考える．

表 1　spam でない通常のメイル 500 通の集計

SPF 調査で pass	339 通 (70 %)
SPF 無しで正しい送信方法	120 通 (20 %)
SPF 調査で softfail，その他	18 通

表 2　spam メイル 500 通の集計

SPF 調査で fail または softfail	約 100 通
From 行なし，アドレスが存在しない	50 通
SPF 調査で pass	22 通
spamhaus 登録済	39 通
RBL.JP 登録済	8 通
逆引き失敗	12 通
正引き失敗	54 通

5 おわりに

spam 判定プログラムを用いることによって spam メイル判定が容易になり，より正確になると考える．そしてメイル使用の手助けになるのではないかと考える．さらに，今後の研究課題として以下のことが挙げられる．

- インターネットを用いた実験
- spam メイル対策プログラムの品質の向上，評価，改善

上記の研究課題を完成させることにより，spam メイルの対策技術が進歩すると考えられる．

参考文献

[1] RBL.JP プロジェクト：RBL.JP (accessed December 2011). http://www.rbl.jp.

[2] spamhaus: The Spamhaus Project (accessed June 2011). http://www.spamhaus.org.

[3] 加藤雅斗，松本征也，南部勝巳：ゲートキーパーへの迷惑メイル対策機能の追加，卒業論文，南山大学数理情報学部 情報通信学科 (2010).

[4] 警察庁：わが国におけるインターネット治安情勢の分析について (平成２０年度第 1/四半期) (accessed June 2011). http://www.npa.go.jp/cyberpolice/detect/pdf/080723.pdf.

[5] 警察庁:インターネットの観測結果等 (accessed June 2011). http://www.npa.go.jp/cyberpolice/detect/pdf/20110428.pdf.