

L-007

経路情報を用いたベイズスパムフィルタ作成に関する検討  
A study of Bayes SPAM filter using routing information

山崎 仁<sup>†</sup>                      白川 正知<sup>‡</sup>                      古川 泰男<sup>‡</sup>  
Jin Yamasaki,                      Masatomo Shirakawa,                      and Yasuo Furukawa

1. はじめに

現在、ネットワークの発展により、誰でも気軽にメールを送受信できるようになったが、それと同時に SPAM メールも増大し、トラフィックのかなりの部分を SPAM メールやウイルスが占めているといわれている。SPAM メールへの対策として様々な技術が開発されている。例えば、大手ソフトウェア会社では、メールシステム全体に ID を組み込み、望ましくないメール送信そのものを防止しようとしている [1]。また、エンドユーザのセキュリティ対策として、いくつかの製品も存在する [2][3]。しかし、ID 方式の場合、メールシステム全体の再構築が必要で、普及までに時間がかかると思われる。また、従来の製品は、特定のサイトやキーワードをフィルタリングするものが多く、SPAM メールの判定効率の点で劣る。

そこで、SPAM メールの判定効率の向上のために、ベイジアンフィルタが注目されている [4]。[4]では、99.5%以上の SPAM メールを判別することができる、記されている。本論文では、メールのヘッダ情報と経路情報を用いて、SPAM メールであるかをベイズ理論によって判定する手法を提案する。

2. 提案システム

本論文で提案するシステムの概要について説明する。提案システムは、2つの機能で構成される。すなわち、

1. メールヘッダ情報が偽造されている場合、SPAM メールと判断
2. メールヘッダが偽造されていなくても、ユーザが SPAM メールと判断した場合は、そのメールの経路情報からベイズ理論によって SPAM メールと判断

2.1 メールヘッダの偽造による判定

図1のように電子メールのヘッダ部分には、そのメールが途中経由してきた中継サーバを示す「Received:」ヘッダがある。基本的にメールサーバは追加することだけが許可されているので、中継地点が多いほど「Received:」ヘッダは増えていく。また、このヘッダは、下から上に付け加えられる。よって、上にいくほど受信者に近く、下にいくほど送信者に近い情報になる。このように、「Received:」ヘッダは迷惑メールを追跡する上で、ある程度送信者を限定できる。しかし、図2に示すように、

SPAM メールの場合メールヘッダは偽造されていることも多い。

```
Return-Path: <xxxx@a-xx.com>
Delivered-To: saxxxxx@xxxxx.freemail.ne.jp
Received: (qmail 5901 invoked from network); 2 Jul 2004 18:33:24 +0900
Received: from unknown (HELO a-xx.a-xx.com) (210.xxx.xxx.2)
by xxxx.freemail.ne.jp with SMTP; 2 Jul 2004 18:33:24 +0900
Received: from a-xx.com (winkxxxxx.winkxxx.ne.jp
[202.xxx.xx.242])
by a-xx.a-xx.com (8.10.0/3.6W) with SMTP id
i629ap801225
for <saxxxxx@xxxxx.freemail.ne.jp>; Fri, 2 Jul 2004
18:36:51 +0900
Date: Fri, 2 Jul 2004 18:36:51 +0900
Message-Id: <200407020936.i629ap801225@a-xx.a-xx.com>
To: saxxxxx@xxxxx.freemail.ne.jp
Subject: XXXXXXXX
From: "xxxx@a-xx.com" <xxxx@a-xx.com>
Mime-Version: 1.0
Content-Type: text/plain; Charset=ISO-2022-JP
Content-Transfer-Encoding: 7bit
```

図1 メールヘッダの例

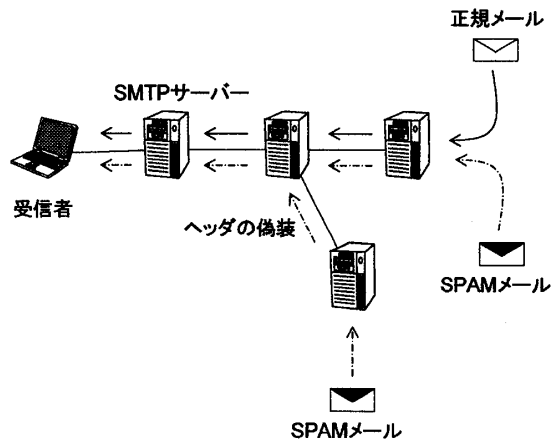


図2 メール経路情報

そこで、「Received:」ヘッダに書いてある内容が正しいか、受信者に近いヘッダ情報から順に DNS サーバに問い合わせ、「Received:」ヘッダの情報が正しいか、つまり中継サーバが正しく管理されているかを調べる。仮に、偽造されている場合は、SPAM メールと判断できる。この手法は、従来のメールサーバにセキュリティ機能として、実装されている。しかしながら、ホストの判定のみに使用されている。本提案手法では、一つのメールに対する中継メールサーバを一組とした情報として取り扱う。す

<sup>†</sup> 豊橋技術科学大学大学院工学研究科, Graduate School of Engineering, Toyohashi University of Technology  
<sup>‡</sup> 豊橋技術科学大学未来技術流動研究センター, Research Center for Future Technology, Toyohashi University of Technology

なわち、メールの中継経路から SPAM メールかどうかを判定する。これにより、送信サーバが動的 IP アドレス割り当てによって変化しても、柔軟にかつより詳細に対応できる。

### 2.2 ベイズ理論による判定

ヘッダ情報が偽造されていない場合でも、SPAM メールの場合がある。そこで、経路情報を用いてベイズ理論により SPAM メールと判断できるようにする。図3のように、「Received:」ヘッダ情報の中継サーバ IP アドレスを用いてあらかじめ木を作成する。各ノードは、IP 情報とそこを経由したメールが SPAM メールか SPAM メールでないかの情報を保持する。

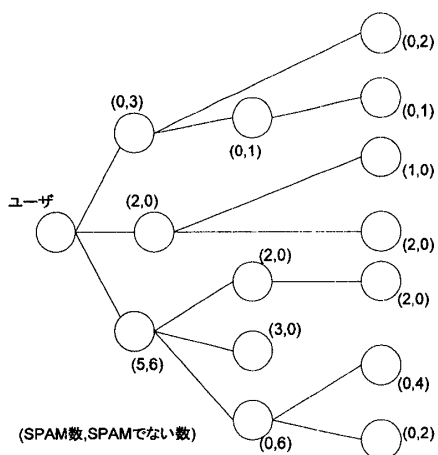


図3 木構造

SPAM メールかどうかを判定したい場合は、経由するノードが SPAM メールを中継した回数としない回数の情報を使用する。すなわち、Paul Graham 方式を応用すると、あるノード  $n$  が SPAM メールを中継した回数を  $sCount$ 、そうでない回数を  $gCount$ 、SPAM メールの総数を  $sTCount$ 、そうでないメールの総数を  $gTCount$  とすると、各ノードでの SPAM 確率  $pg_n$  は、

$$pg_n = \frac{\frac{sCount}{sTCount}}{\frac{2 \times gCount}{gTCount} + \frac{sCount}{sTCount}} \quad (1)$$

ここで、あるノードを経由したメールが全て SPAM メールだった場合は、 $pg_n = 0.99$  に、逆に全て SPAM メールでなかった場合は、 $pg_n = 0.01$  とする。また、木に存在しなかった場合は、 $pg_n = 0.4$  とする。

各ノードでの SPAM 確率を求めた後、それらの結合確率を求めることで、全体の SPAM 確率とする。ノードが 3 つの場合、結合確率  $pg$  は、

$$pg = \frac{pg_1 pg_2 pg_3}{pg_1 pg_2 pg_3 + (1 - pg_1)(1 - pg_2)(1 - pg_3)} \quad (2)$$

となる。この値が 0.9 以上の場合、SPAM メールと判断する。

### 3. まとめ

本論文では、基本技術の検証のために、プログラムを作成した。サンプルメール 2000 通 (SPAM メール: 1500 通) を使用し、ベイズ理論に用いる木を構成した。サンプル SPAM メールは、実際に受信したものを使用した。プログラム言語は C++ 言語を使用し、WindosXP(R) 上で構成した。

200 通のうち SPAM メール 100 通、ヘッダ偽装 SPAM メール 50 通のテストメールを、作成したプログラムに対して実行した。その結果、SPAM メール 100 通のうち、60 通のメールを SPAM メールとして認識した。また、SPAM でないメールを SPAM として認識することはなかった。

今後の研究において、さらなる検証をしつつメールシステムに組み込み、実装する予定である。

### 参考文献

- [1] <http://www.microsoft.com/mscorp/twc/privacy/spam.mspcx>
- [2] <http://www.qurb.com/>
- [3] <http://www.vectant.co.jp/>
- [4] <http://www.paulgraham.com/spam.html>