4W-1

送信者情報を利用した迷惑メール対策

野原 史朗　　志田 晃一郎　　横山 孝典
武蔵工業大学

1　研究背景

近年ではインターネット環境の整備などにより，メールがネットワーク上における主要なコミュニケーション手段の１つになっている．メールは送信コストが安く，インターネット環境さえあればいつでも瞬時に相手に届けることができるので，ビジネスなどにも用いられる反面，迷惑メールが急増しており，対策技術が求められている．

2　従来研究

2.1　アドレス指定拒否

アドレス指定拒否とは，迷惑メールを受信した際に，そのメールの送信者アドレスを受信拒否リストに加えることで，以降その送信者アドレスからのメール受信を行わない手法である．しかしこの手法では，送信者アドレスを偽装されると効果がなくなってしまう．

2.2　送信者アドレス偽装対策技術 [1]

前述の送信者アドレス偽装を行わせないような送信システムも存在する．SMTP AUTH という送信時認証技術を拡張したもので，従来用いられていたアカウント名とパスワードに，送信者アドレスも認証項目に追加してある．

つまりこのシステムでは，メールサーバに登録されているユーザであり，かつ，送信者アドレスを偽装していないユーザのみが，メールを送信できるようになっている．

3　問題点

近年では，渡りという手法が特に問題となっている [2]．この手法の概要を図１に示す．この手法ではまず，迷惑メール送信者は，登録してあるメールサーバ（A 社）から大量にメールを送信する．その際，送信者アドレスを偽装せずに送信するため，アドレス指定拒否による対策が考えられる．しかし，数日後にはそのメールサーバの利用をやめ，別のメールサーバ（B 社）に登録し，新しい送信者アドレスで再度大量送信を行うので，アドレス指定拒否も効果がない．

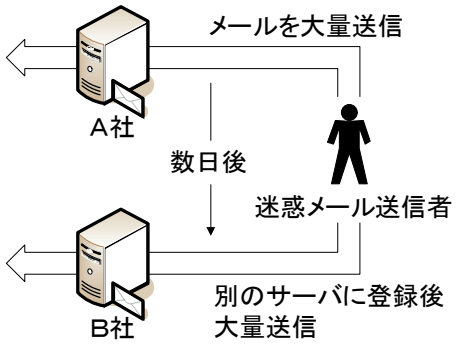

図 1．渡り

4　研究目的

送信者アドレス以外の送信者情報をメールに付与することで，迷惑メールを振り分けることを目的とする．

5　提案手法

提案するシステムの概要を図２に示す．提案手法ではメール送信の際に送信側サーバにおいて，送信者のアカウント情報をメールヘッダに追加する．そしてその情報を用いてフィルタリングを行う．

メールヘッダに追加する送信者アカウント情報は，以下のものを用いる．

- アカウント取得からの日数
  渡りを行っている送信者は，日数が浅い傾向にある．
- 一日あたりの送信数
  未承諾広告などの送信者は，送信数が多い傾向にある．
- 送信エラー率
  総当りで作成したアドレスに対して送信している場合，送信エラーが多い傾向にある．

これらの情報をデータベースに格納しておき，適宜読み出す．

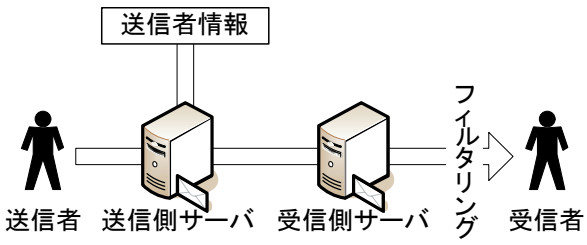

図 2．提案システムの概要

5.1　送信側サーバ

メールサーバは，データベースアクセスや，ヘッダ追加などを行えるように機能を拡張する．メールサーバの動作概要は図３のようになる．

The Countermeasures Against Spam Mail Using Sender Information
Shiro Nohara, Koichiro Shida and Takanori Yokoyama,
Musashi Institute of Technology

1. メール送信要求を受ける.
2. 送信者アカウントの情報をデータベースから読み出し，メールヘッダを追加する.
3. データベース上のアカウント情報の更新を行う.
4. メールを送信する.

アカウント情報の更新に関しては，一日あたりの送信数は毎回更新される.アカウント取得からの日数は,データベースにアカウント取得日を格納しておき，拡張部の中で取得からの日数を計算するため，アカウント登録時に書き込まれ，以降は更新されない．送信エラー率は，累積送信数と送信エラー数をデータベースに格納しておき，拡張部内でエラー率を計算する．送信エラー数の更新は，エラーメールを受信した際に行う.

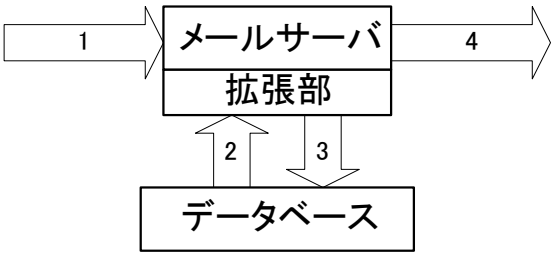

図 3．送信側サーバの動作概要

5.2 受信側

受信側では，受信したメールを，正常なメールと迷惑メールとに振り分ける必要がある．そこで今回は，振り分けフィルタとして広く用いられているベイジアンフィルタを使用する．ベイジアンフィルタはデータが入力されると，過去に振り分けられたデータを参照し，入力されたデータがどちらに分類されるかを判断する．判断したデータは過去の振り分けデータとして蓄積され，判断に誤りがあった場合はユーザが修正を施せるようになっている.

6 実装

今回提案したシステムでは，送信者アカウント情報付加機能と，データベースアクセス機能を送信側サーバに持たせる必要がある.

送信者アカウント情報付加機能に関しては，sendmail[3] というメールサーバにおいて提供されている API の中に，送信者アドレス取得関数や，ヘッダ操作関数などがある．そこで，その関数を用いて取得した送信者アドレスを元に，データベースから取得した送信者アカウント情報をヘッダに追加する.

データベースアクセスに関しては MySQL というデータベースにおいて提供されている C 言語用 API を用いる.

7 評価

今回提案した,メールに送信者情報を付加した場合と,従来の場合とで，受信側での振り分け精度にどの程度の違いが出るかを評価する.

ISP などのメールサーバーに本機能を組み込み，実際に得た送信者情報を元に評価を行う事が望ましいが，今回は送信者情報をランダムに決定し，評価を行った．正常な送信者と，渡りを行う迷惑メール送信者には，送信者情報に違いがあると考えられる．今回設定した送信者情報の範囲を表 1 に示す．これらの範囲の元でランダムに生成した送信者情報を利用して，正常なメールを 400 通，迷惑メールを 3600 通送信し，ベイジアンフィルタに学習を行わせた..

この学習データを用いて，別途送信した正常なメール 500 通，迷惑メール 500 通を振り分けた．また，従来手法との比較用に，送信者情報を付加しないメールに関しても同じように学習と振り分けを行った．振り分け結果を表 2 に示す．従来手法では 84% の振り分け精度だったが，提案手法では 98% となり，振り分け精度の向上が確認できた.

表 1．送信者情報

	正常な送信者	迷惑メール送信者
アカウント取得からの日数	3600日以内	5日以内
一日あたりの送信数	100通以内	10000通以内
送信エラー率	10%以内	90%以内

表 2．振り分け結果

メール種別	Clean		Spam		精度
判定結果	Clean	Spam	Clean	Spam	
従来手法	375	125	35	465	0.84
提案手法	500	0	25	475	0.98

8 おわりに

本研究では，送信者情報をメールに付加し，受信側での振り分けに利用する手法を提案し，従来手法との精度の比較を行った.

今後は，学習数を変化させた場合，従来手法と提案手法とでどのように精度が変化するかといった調査や，どの送信者情報がどの程度振り分け結果に影響したかの調査を行う予定である.

参考文献

[1] 松原義継,”偽の送信者メールアドレスを持つメールの配送を防止するフィルタ”，情報処理学会論文誌 Vol.47 (2006), No. 4, pp. 992-999.

[2] IronPort Systems
http://www.ironport.com/

[3] SENDMAIL
http://www.sendmail.com/jp/