2005－ＤＳＭ－37（15）
2005／5／13

宛先不明メールを利用した分散協調型spamフィルタの認識率向上

漣　一平[†]　山井　成良[†]　岡山　聖彦[†]

宮下　卓也[††]　丸山　伸[†††]　中村　素典[†††]

† 岡山大学　〒700-8530　岡山市津島中一丁目１番１号
†† 津山工業高等専門学校　〒708-8509　岡山県津山市沼 624-16
††† 京都大学　〒606-8501　京都市左京区吉田本町
E-mail: †sazanami@net.cne.okayama-u.ac.jp

あらまし　電子メールにおいて大きな問題となっている spam メールへの対策方法として，spam フィルタがよく用いられている．この方法で用いられる代表的な技法のうち，分散協調型フィルタは誤検出率が低いという利点を持つ反面，検出率も低いという欠点がある．本稿ではこの欠点を改善するため，宛先不明メールを潜在的な spam メールとみなし，分散協調型 spam フィルタの spam メール認識率を向上させる方式を提案する．これにより，従来の分散協調型 spam フィルタでの利用者による spam メール登録のみの場合に比べて，認識率の向上が見込める．また，メールゲートウェイにおける提案方式の実装を行い，提案方式の利用により分散協調型 spam フィルタの認識率向上が期待できることを示す．

キーワード　spam メール，分散協調型フィルタ，メールゲートウェイ

An Accuracy Improvement Method of Distributed Collaborative Spam Filter Using Invalid Recipient Mail

Ippei SAZANAMI[†], Nariyoshi YAMAI[†], Kiyohiko OKAYAMA[†],

Takuya MIYASHITA[††], Shin MARUYAMA[†††], and Motonori NAKAMURA[†††]

† Okayama University　1-1, Tsushima-Naka, 1-Chome, Okayama, 700-8530 Japan
†† Tsuyama National College of Technology　Numa 624-16, Tsuyama-City, Okayama, 708-8509 Japan
††† Kyoto University　Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan
E-mail: †sazanami@net.cne.okayama-u.ac.jp

Abstract　Spam filters are commonly used for a kind of protection measures of spam mail, which is one of the most serious problems on e-mail environment. As a kind of filtering methods, distributed collaborative filter is remarkable since its false positive rate is very small. However, this method has a significant drawback that its accuracy is considerably low. In this paper, in order to improve the accuracy of distributed collaborative filters, we propose a new method using mails sent to non-existent addresses that may potentially be spam mails We have also implemented the proposed method on a mail gateway, and shown that the accuracy may be improved.

Key words　spam mail, distributed collaborative filter, mail gateway

1. はじめに

spam メールとは，広告等を目的とした不特定多数の利用者に対して一方的に送信される電子メールのことであり，迷惑メールとも呼ばれる．メールによる広告は他の広告手段に比べてほとんどコストがかからず，また，インターネット利用者の増加に伴いその広告効果が高まったため，ここ数年で爆発的に増加している．米国のスパム対策企業 Brightmail 社 (現 Symantec 社) による 2004 年 3 月の調査では，全世界のメールの内 63%が spam メールであったとされている [1]．spam メールの蔓延は世界的な社会問題にまでなっており，各国で法律による規制も行われている．日本においても 2002 年 7 月に迷惑メール規制法が施行されたが，この法律だけでは spam メールの増加に歯止めをかけることができず，手口の巧妙化，悪質化も進んでいる．また，最近は単なる広告だけにとどまらず，利用者をだまし，個人情報等の搾取や詐欺を試みるフィッシングメール，架空料金請求メール等の極めて悪質な spam メールや，ウィルス，スパイウェアに感染したインターネット常時接続コンピュータ (ゾンビ PC) からの spam メール発信の増加も問題となっており，spam メール対策の重要性が一層高まっている．

spam メールによって一般の利用者が受ける被害は，受信した大量のメールから spam メールと非 spam メールを選別するために時間と労力を浪費することや，その煩雑性に伴い非 spam メールを見落とす危険性があることである．

spam フィルタの代表的な技法としては，ルールベースフィルタ，ベイジアンフィルタ，分散協調型フィルタなどが挙げられる．ルールベース，ベイジアンフィルタは spam メールによく現れる特徴をルールとして記述したり，学習させたりしておき，それを元に spam メールの判別を行うものである．これらの技法は受信した spam メール中のどの程度の割合を spam メールと認識できるかを示す認識率は高いが，非 spam メールを誤って spam メールであると認識してしまう割合を示す誤認率も無視できず実用上の大きな問題になっている．またフィルタルールや学習設定に気を配らなければ，日々巧妙化する新種の spam メールに対応することができない．一方，分散協調型フィルタは各利用者が spam メールと認識したものを利用者間で共有するデータベースに登録し，そのデータベースへの登録の有無により spam メールかどうか判別する技法である．この技法では，一定数以上の利用者が同一の spam メールを受け取ってそれをデータベースに登録した後に，フィルタリングされるため，認識率は比較的低いが，誤認率が事実上無視できる程低いという特徴がある．そのため，認識率を向上させることができれば，非常に優れた spam フィルタとなり，これまで誤認率が問題で spam フィルタを導入できなかった環境等への適用も考えられる．

本稿では分散協調型フィルタを対象として，その認識率を向上させる手法を提案する．具体的には，従来の手法では spam メールの可能性が高いにも関わらず，spam メール認識の判断材料として利用されていなかった，実在しないメールアドレスへ送信される宛先不明メールを潜在的な spam メールとみなし，MTA(Mail Transfer Agent) がメール転送の処理を行なう段階でデータベースに登録する．これにより，従来の利用者からの登録のみの場合より，spam メールの認識率の向上が期待できる．また，分散協調型フィルタのデータベースへの登録の後，SMTP(Simple Mail Transfer Protocol) [2] セッションをエラーで終了させ，宛先不明メールを受信しないようにする．これにより，宛先不明であったことを spam 送信者に通知するエラーメールが発生しないため，自ドメインが発信者アドレス詐称による DoS(Denial of Service) 攻撃や spam メール配送に利用される危険性が低下する．さらに，メールゲートウェイで，一般的に宛先不明であることが判明する前に行われているウィルスチェック等の処理を，宛先不明メールに対して行わないようにすることでメールゲートウェイの負荷の低減も期待できる．

以降，まず第 2 章において，一般的な spam フィルタ技法の特徴と問題点について述べる．次に第 3 章で分散協調型フィルタの認識率改善方式を提案し，第 4 章では提案方式の実装と有効性について述べる．第 5 章では結論と今後の課題について述べる．

2. 従来の spam フィルタの問題点

本章では，spam メールによる被害について述べ，一般的な spam フィルタの概要と問題点について述べる．

2.1 spam メールによる被害

多くの spam メールにみられる特徴は，

- 大量に送信される
- 実在しないメールアドレスにも送信される
- ほとんどが送信者アドレスを詐称している
- 一度送信が成功すると繰り返し送信される

などが挙げられる．一般的な spam メールによる被害は，

(1) 利用者が大量のメール中から非 spam メールを選別するために時間と労力を浪費する．また，その煩雑性により誤って非 spam メールを削除したり，見落とす危険性がある．

(2) spam メール受信のために，計算機資源，ネットワーク資源，通信費用，通信時間などを浪費する．

(3) spam メールの発信者アドレスを自組織のものに詐称されることによって，spam メールの発信に関与していると疑われる．また，その詐称アドレス宛に宛先不明を通知するエラーメールを大量に受信することにより自組織のメールサーバが過負荷になる．

などが挙げられる．このうち，一般の利用者にとって最も影響が大きい (1) の被害への対策方法として，spam フィルタがよく用いられる．これは，利用者がメールを読む前に内容を検査して spam メールであるかどうかを判断し，spam メールと非 spam メールを振り分ける方法である．次節では，従来の spam フィルタの代表的な技法とその問題点について述べる．

2.2 従来の spam フィルタとその問題点

前節でも述べたように，利用者による非 spam メール選別の負担を軽減する方法として spam フィルタがよく用いられている．本節では，従来の spam フィルタ技法を示し，その問題点を明らかにする．

a) ルールベースフィルタ

ルールベースフィルタは，たとえばヘッダに偽造の痕跡がある，HTML のみで記述されている，特定のキーワードを含むなど，spam メールと通常メールを識別できる特徴及びその重みが予めルール及びスコアとして記述されており，受信メールに合致したルールの合計スコアが一定値以上であると，これを spam メールと判定する技法である．この技法に基づく代表的な spam フィルタとしては SpamAssassin [3] が挙げられる．

この技法では経験的に得られた spam メールの特徴をルール化するため，典型的な spam メールの認識率は非常に高い反面，新たな手口の spam メールについては新たなルールを作成する必要があるため柔軟性に欠けるといえる．また，spam メールと同じ特徴を持つ非 spam メールを受信した場合にはこれを誤認する危険性が無視できない点も問題である．

b) ベイジアンフィルタ

ベイジアンフィルタは，単語や 3 連文字などの出現頻度を基に spam メールを認識する技法である [4]．具体的には，ベイジアンフィルタでは過去に受信した spam メール及び非 spam メールの単語等の出現頻度を分析してベイズの定理により各単語等に対するスコアを算出しておき，受信メールに出現した単語等の合計スコアが一定値以上であると，これを spam メールと判定する．この技法に基づく代表的な spam フィルタとしては bogofilter [5] が挙げられる．

この技法では，利用者の判定に基づいて単語等のスコアを学習させることができるため，利用者に応じた調整が可能であり，また新たな手口の spam メールについても学習によりある程度適応することが可能である．一方，この学習には利用者の介在が不可欠であり，また一般にスコアの再計算にはかなりの時間を要する点が問題である．また，ルールベースフィルタと同様に，高スコアの単語等が偶然多数含まれる非 spam メールを誤認する危険性も無視できない．特に，最近では新たな手口として無作為に選択された多数の単語等 (word salad と呼ばれる) を含む spam メールが見受けられるようになっているが，これを学習させたためにスコアが大きく乱され，spam メールの認識率が低下して誤認率が上昇してしまう現象が報告されている [6]．

c) 分散協調型フィルタ

分散協調型フィルタは，上記の 2 つの技法とは異なり，同一内容の電子メールが多数の利用者に送信されるという spam メールの性質を利用した技法である．分散協調型フィルタでは利用者間で共有する spam メールのデータベースを導入し，このデータベースへの登録の有無により電子メールが spam メールであるかどうかを判定する．その際，高速化を図るため，spam メールの本文全体ではなく一種のチェックサム (本文中の部分的な改変に対応するために工夫をしているものが多い) を計算してこれをデータベースへの登録・照合に用いる．代表的な分散協調型フィルタとしては DCC(Distributed Checksum Clearinghouse) [7]，Vipul's Razor [8]，Pyzor [9] などがある．

分散協調型フィルタにおける典型的な処理過程を図 1 に示す．MTA からメールを受け取って利用者のメールボックスへの振り分けなどを行う MDA(Mail Delivery Agent) は電子メールを受信するとまずチェックサムを計算し，同じチェックサムが登録されているかどうかをデータベースと照合する．その結果，同一チェックサムの登録が確認されれば，これを spam メールとみなして破棄したり通常メールとは別に格納したりする．同一チェックサムの登録が確認されない場合には，通常メールと見なされてメールボックスに格納されるが，後に利用者が電子メールを読んだ際に spam メールであると判断すれば，利用者がメールを送受信するプログラムである MUA(Mail User Agent) の持つ転送機能などを利用してデータベースへの登録を行う．

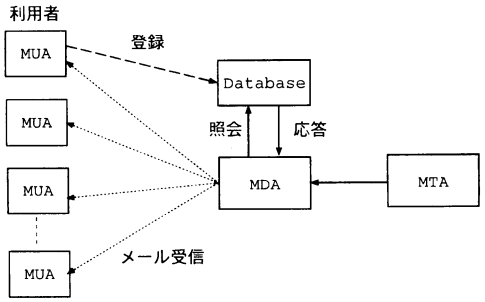

図 1 分散協調型フィルタの処理過程

この技法はベイジアンフィルタと同様に利用者の介在が不可欠であり，また spam メール判定時までに他の利用者がデータベースに登録していないと spam メールを認識できないため，認識率が比較的低いという欠点がある．一方，データベースへの登録は利用者の判断に基づくため，他のプログラムに基づく技法では見逃してしまう spam メールにも適応でき，誤認率は事実上無視できるほど小さい点が優れている．また，データベースへの登録に要する時間はベイジアンフィルタでの学習の場合と比較して短時間で済むという点も本技法の利点である．

3. 分散協調型 spam フィルタの認識率向上

前節で述べたように，分散協調型フィルタは誤認率が無視できるほど低いという特徴を持っているため，優れた spam フィルタとして期待されている．そこで，本稿では分散協調型フィルタの欠点である比較的低い認識率を向上させる技法を提案する．

3.1 提案方式の概要

前述したように，従来の分散協調型フィルタでは，spam メール認識のための判断材料は，利用者によって spam メールと判断および登録されたものだけである．これは，信頼性の高い spam メール認識のために一定数以上の利用者からの登録があったメールを spam メールとみなしているからである．spam メール認識率の向上のためにおとり用のメールアドレスを設けて，そのアドレスに送られてきたものを spam メールとみなして判断材料に利用するという手法もあるが，この場合，少数のおとり用メールアドレスの利用では，収集できる spam メールの数が少なく，認識率向上の効果が低いと思われる．

そこで本稿では，実在しないメールアドレスに送られる宛先不明メールを分散協調型フィルタの spam メールの判断材料に利用することを提案する．なぜならば，無作為に選ばれたメールアドレスに対して大量に送信されるという多くの spam メールにみられる特徴から，宛先不明メールは spam メールである可能性が高いからである．これにより，従来の技法では MTA で User Unknown エラーとなって，利用者によって認識されることが無かった spam メールをデータベースに登録することができ，分散協調型フィルタの認識率向上を見込める．ここで提案方式では宛先不明メールを spam メール認識のための判断材料に利用することにより，分散協調型フィルタの特徴である，利用者による spam メール認識が行われない．しかし，宛先不明メールの数は有効なメールアドレスに送られるものと比較して非常に多く，そのほとんどが spam メールであることから，spam メール認識の判断材料として妥当であると考えられる．

3.2 メールゲートウェイでの宛先メールアドレスの存在確認

本稿では，提案方式を適用する環境として，岡山大学をはじめ一般的な組織で多く利用されていると思われる，メールゲートウェイと階層化されていない複数の末端メールサーバから構築されている環境を想定する．従来の技法では，外部から届けられた宛先不明メールをメールゲートウェイで一旦受信した後，末端のメールサーバへ転送される時点で宛先不明であることが判明する．そのため，宛先不明メールであってもメールゲートウェイでディスクに保存され，末端の受信メールサーバへ転送を行う前にウィルスチェック等を行う．受信者の存在しないメールに対してこれらの処理を行うことは無駄であり，ディスク容量を圧迫する，ウィルスチェックによる CPU 負荷が高くなる等，メールゲートウェイにおける spam メールによる被害が発生する．大量の宛先不明メールを処理することによって，メールゲートウェイが過負荷に陥ることが問題となっている．

この問題を解決するために，提案方式では，メールゲートウェイにおいて宛先メールアドレスが転送先の末端メールサーバに実在するかどうかの確認を SMTP セッションを用いて透過的に行う．また，SMTP セッションをメールゲートウェイにメールが届く毎に開始するのではなく，一度末端のメールサーバに宛先アドレスの存在を問い合わせた後，可能な限り接続を維持してコネクションを再利用する手法を提案する．これにより，高速に宛先メールアドレスの存在有無の確認が可能である．

上記手法によって，宛先不明であると判明したメールに関して，分散協調型フィルタの spam メールの判断材料に利用するために，宛先不明が判明した時点ではそれを送信者には通知せず，メール本文を受信し，分散協調型フィルタのデータベースへの登録を行う．提案方式では，宛先不明メールに対して，本文の受信が正常に完了したという応答をせず，エラーにより失敗したと示すことにより宛先不明メールの送信を成功させないようにする．

この技法により，宛先不明メールの内容を分散協調型フィルタにおいて spam メール認識の判断材料に利用することができ認識率の向上が期待できるほか，宛先不明を送信者 (ほとんどの spam メールで詐称されている) に通知するエラー通知メール (バウンスメール) の発生を防ぐことができるため，エラーメールを利用した DoS 攻撃や spam 配送を抑制できる．

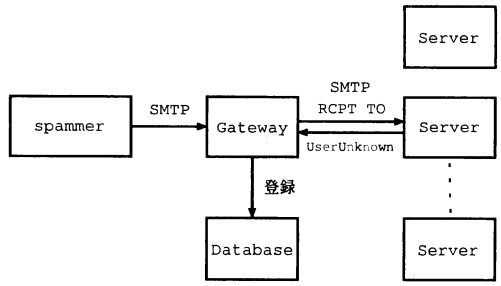

図 2 試作システムの構成

また，多くの spam メール送信者は，一度送信が成功すると繰り返し送信を試みるため，送信を成功させないことにより，spam メールが繰り返し送信される可能性が低下する．

4. 試作システムの実装と性能評価

4.1 提案方式の設計・実装

提案方式を実現するためには，末端のメールサーバに対して宛先メールアドレスの存在を確認する処理と，宛先不明メールを分散協調型フィルタのデータベースへ登録するという処理をメールゲートウェイにおいて行う必要がある．

まず，末端メールサーバへの宛先メールアドレスの存在確認は，前述したように宛先不明メールであった場合，分散協調型フィルタのデータベースに登録する必要があるため，メールゲートウェイの SMTP セッションで，メール本文を受信する前に完了しなければならない処理である．これは SMTP プロクシとしての実装も考えられるが，今回は，最も普及している MTA であると思われる Sendmail [10] の機能拡張用 API である Milter [11] [12] を利用することにした．Milter を利用することにより，MTA での SMTP セッション内に任意の処理を追加することが可能となる．また，宛先アドレスの存在確認処理を高速に行うために，末端のメールサーバへの SMTP セッションを可能な限り再利用するようにした．

次に，宛先不明メールを分散協調型フィルタのデータベースに登録する処理については，前述の Milter を利用し，ゲートウェイでの SMTP セッション完了前に処理を行うようにした．対象とする分散協調型フィルタには DCC を用いた．宛先メールアドレスが複数指定されているメールに関しては，一定割合が宛先不明の場合のみ spam メールとみなし，登録するようにした．

上記の 2 つの処理をゲートウェイにて行うプログラムを perl スクリプトとして実装した．試作システムの構成を図 2，通信過程を図 3 に示し，実装部分の具体的な動作を以下に示す．

(1) メールゲートウェイにメール送信要求が到着し，SMTP セッションが開始される．

(2) メール送信ホストから RCPT TO コマンドにて宛先メールアドレスを受け取った後，そのアドレスから転送先の末

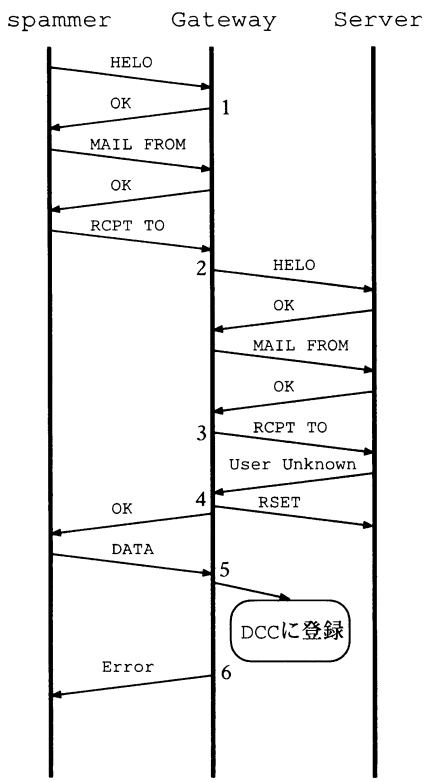

図 3　試作システムの通信過程

端メールサーバを特定し，末端メールサーバに接続して SMTP セッションを開始する．ただし，既に接続されている場合はその接続を再利用する．この場合 HELO コマンドを送る必要は無い．

（3） 末端メールサーバに対し，RCTP TO コマンドで宛先メールアドレスが有効かどうか問い合わせる．複数の宛先メールアドレスがある場合，宛先アドレス毎に同様の処理を行う．

（4） 末端メールサーバからの応答により，宛先不明メールであると判明して (550 User Unknown というエラー応答) もメール送信者に対しては正常 (250 ok) と返答する．また，末端メールサーバとの SMTP セッションは終了せず，MAIL FROM コマンド，RCPT TO コマンドで設定したアドレスを初期化し，セッションを再利用するために，RSET コマンドを送信する．

（5） DATA コマンドを受け付け，メール本文を受信し，DCC のデータベースへ登録を行う．(複数の宛先メールアドレスを含む場合はあらかじめ設定した割合以上で登録する)

（6） 宛先不明メールの場合，本文受信の後，メール送信者の DATA コマンドに対する応答として，エラーを示し，メールの送信を失敗させる．

試作システムの通信過程を図 3 に示す．

4.2 性能評価

提案方式の有効性を評価するために，試作システムの spam メール認識率を求めた．この値は受信するメールの数，宛先不明メールを受信するタイミング，システムの利用者数，各利用者がメールを受信するタイミングなど多くの要因に依存するため正確に測定することは困難である．また，実際に稼働しているシステムを用いた実験はリスクが大きいため，実際のシステムで運用する代わりにすでに受信した spam メールを分析し，期待される認識率を推定することにした．

2005 年 1 月 8 日から 2005 年 1 月 13 日までの間に，岡山大学ドメイン宛に送られた宛先不明メールを得るために，末端のメールサーバが送信元にエラー通知メールの配送を試み，(送信元詐称により送信者不明のため) それが失敗となったダブルバウンスメールを収集し，その中から元の宛先不明メールを抽出したもの (2 つのメールゲートウェイで収集した合計 111,644 通) と，同期間に 1 つの個人用アドレス宛に送られてきた spam メール (560 通) を利用して実験を行った．

個人用アドレスに宛に届いた spam メールは，SpamAssasin(代表的なルールベースフィルタ，以下 SA と略記する) によって spam と判定されたもの，および SA では認識できず利用者が見て spam メールと判別したものからなる．また，判定に利用された SA は特別なチューニングが行われていないものであるが，判定ルールには従来の宛先不明メールを利用しない DCC でのチェックが含まれており，従来の DCC で認識可能かどうかヘッダに出力されている．これを利用して，提案方式の有効性を検証した．

実験内容は，宛先不明メールと同じチェックサムを持つ spam メールが利用者が受信したメール中にあるか調べ，あればその受信時刻が宛先不明メールより後であるかどうか調べた．宛先不明メールより後に受信したものであれば，その spam メールは提案方式で認識可能である．まず，DCC で全てのメールのチェックサムを計算し，次に，各々のメールの一番新しい Received ヘッダから受信時刻を取得した．そして，全てのメールの中で，同じチェックサムをもつメールを時系列順に並べ，1 通以上の宛先不明メールの後に利用者の受信した spam メールがあるかどうか調べた．利用者が受信した spam メールについては，実際に MUA で取得した時刻は不明であるが，それに最も近いと思われる上記の時刻を利用した．提案方式で認識できる spam メールのうち，従来の DCC で認識されていないものがあれば，宛先不明メールを利用することによって，分散協調型フィルタの認識率が向上することを示している．

4.3 結果・考察

実験結果を表 1 に示す．この表から，個人用アドレス宛に送られた 560 通の spam メールのうち，宛先不明メールを利用しない従来方式の DCC で認識可能な spam メールは 241 通あることがわかる．提案方式で認識が期待できる spam メールは，210 通あり，これらのうち従来方式の DCC で認識不可で，提案方式で認識可能な spam メールは 107 通ある．この 107 通が宛先不明メールの利用により新たに認識されたものである．これにより従来方式に提案方式を加えると，全体からみた認識率が宛先不明メールを利用しない場合の 43%から 62%までの向上が期待される．よって提案方式により分散協調型フィルタの

表 1　従来方式の DCC と提案方式の比較

		提案方式		
		認識可能	認識不可	合計
従来	認識可能	103	138	241
	認識不可	107	212	319
合計		210	350	560

表 2　SA と提案方式の比較

		提案方式		
		認識可能	認識不可	合計
SA	認識可能	65	147	212
	認識不可	145	203	348
合計		210	350	560

認識率を向上させることができるといえる．また，提案方式と従来方式の両方で認識可能なものが 103 通あるが，提案方式により認識されるタイミングのほうが早いため，より多くの利用者がフィルタリングの効果を得られる．

続いて，SA による spam メール認識率と提案方式を比較した結果を表 2 に示す．この表から，提案方式で認識可能な 210 通のうち，SA で認識不能なものが 145 通あることがわかる．SA で認識不能な 348 通のうち，その約 42%を提案方式よって認識可能であることから，分散協調型フィルタ以外と組み合わせても有効性があるといえる．

5. むすび

本稿では，誤認率が無視できる程小さいという特徴をもつ分散協調型 spam フィルタを対象として，その欠点であった比較的低い spam メール認識率を向上させるために，宛先不明メールを同フィルタの spam メールの判断材料として利用する方式を提案した．また，提案方式を実装し，試作システムの性能評価を通してその有効性を評価した．

今後の課題は，今回の提案方式の性能評価において，利用者がメールを受信する以前に，同じチェックサムをもつ宛先不明メールが 1 通でも登録されていれば spam メールとみなすという前提を設定したが，この値をどの程度増加させると性能に影響が出るか検証することが挙げられる．

また，提案方式と同様に利用者に頼らずにデータベースに spam メールを登録する方式として，ハニーポットと呼ばれるおとりアドレスを用いる方式がある．この方式は，Web 上におとりアドレスを公開するなどして，メールアドレス探索ロボットにおとりアドレスを収集させ，このおとりアドレスに送られてきた spam メールを自動的にデータベースに登録する方式である．この方式は，おとりアドレスに送られてきたメールだけしかデータベースに登録できないため，宛先不明メールを全て登録することのできる提案方式よりも検出率が劣ると思われるが，今後の課題として，実際に両者の検出率を比較して提案方式の有効性を確かめることが挙げられる．

謝辞

本研究の一部は平成 15〜16 年度科学研究費補助金 (基盤研究 (C)(2)，課題番号 15500039) の補助を受けている．

文　献

[1] シーネットネットワークスジャパン株式会社:
"「迷惑メール送信事業者のビジネスモデルを破壊する」MS のスパム対策とは",
CNET Japan : ニュース,
http://japan.cnet.com/news/sec/story/0,2000050480,20065358,00.htm.

[2] J. Klensin:
"SIMPLE MAIL TRANSFER PROTOCOL",
RFC 2821 , April 2001.

[3] SpamAssassin dev team:
"SpamAssassin: Welcome to SpamAssassin",
http://www.spamassassin.org/index.html.

[4] Paul Graham:
"A Plan for Spam",
http://www.paulgraham.com/spam.html, August 2002.

[5] Eric S. Raymond:
"Bogofilter Home Page",
http://bogofilter.sourceforge.net/ .

[6] John Graham-Cumming:
"How to Beat a Bayesian Spam Filter", in 2004 Spam Conference (unpublished),
http://www.spamconference.org , Januray 2004.

[7] Rhyolite Software:
"Distributed Checksum Clearinghouse",
http://www.rhyolite.com/anti-spam/dcc/.

[8] Vipul Ved Prakash:
"Vipul's Razor: home",
http://razor.sourceforge.net/ .

[9] Frank J. Tobin:
"Pyzor",
http://pyzor.sourceforge.net/ .

[10] Sendmail, Inc.:
"Sendmail Home Page",
http://www.sendmail.org/.

[11] Milter:
"Milter Home Page",
http://www.milter.org/.

[12] Charles Ying:
"Sendmail::Milter Home Page",
http://sendmail-milter.sourceforge.net/.