社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ＤＰＳ－117　(25)
2004－CSEC－　24　(25)
2004／3／5

遅延評価による分散協調型spamフィルタの検出率向上

漣　一平[1]　山井成良[2]　岡山聖彦[1]　宮下卓也[2]　丸山　伸[3]　中村素典[4]

[1]岡山大学 工学部
[2]岡山大学 総合情報処理センター
[3]京都大学 大学院情報学研究科
[4]京都大学 学術情報メディアセンター

概　要

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，MUA(Mail User Agent)による電子メール取得時にまでフィルタリング適用時期を遅延させる方式を提案する．これにより他の利用者によるspamメールの登録が増加し，検出率の向上が期待できる．またPOPプロクシによる提案方式の実装を行い，十分に小さいオーバヘッドでかなりの検出率改善効果が期待できることも示す．

An Accuracy Improvement Method of Distributed Cooperative Spam Filter Using Delayed Evaluation Technique

Ippei Sazanami[1], Nariyoshi Yamai[2], Kiyohiko Okayama[1], Takuya Miyashita[2], Shin Maruyama[3], and Motonori Nakamura[4]

[1]Faculty of Engineering, Okayama University
[2]Computer Center, Okayama University
[3]Graduate School of Informatics, Kyoto University
[4]Academic Cener for Computing and Media Studies, Kyoto University

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 cooperative 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 cooperative filters, we propose a delayed evaluation technique such that filtering is performed when a Mail User Agent (MUA) has access to the mail server. This method can generate an additional period for registration of new spam mails received by other users, therefore we expect it improves the accuracy. We have also implemented the proposed method as a POP proxy, and shows that the accuracy may be improved in small overheads.

1 はじめに

電子メールはWWWと並んでインターネットにおいて最も普及しているサービスの1つであり，社会的な活動を支える通信手段としてもはや必要不可欠な存在となっている．一方，電子メールはセキュリティ上最も問題の多いサービスの1つである．特に，広告等を目的に不特定多数の利用者に一方的に送信されるspamメール (UBE(Unsolicited Bulk E-mail) あるいはUCE(Unsolicited Commercial E-mail) とも呼ばれる) の蔓延は大きな社会問題にまでなっており，その対策は重要である．spamメールによる被害には，(1) 一般の利用者は，受信した大量の電子メールの中から少数の非spamメールを選別するために時間を浪費し，場合によっては非spamメールを誤って削除する危険性がある，(2) 不必要なメールの受信により計算機資源，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メールの中継に自組織のメールサーバが用いられることにより，当該spamメールの発信に関与していると疑われる，(4)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り，当該spamメールの発信に関与していると疑われ，また宛先不明を通知するエラーメールの大量発生によりメールサーバが過負荷になる，などがある．以下では，これらのうち一般の利用者にとって最も影響が大きい(1) について議論する．

上記 (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ている．これは，利用者が電子メールを読む前に内容を検査してspamメールかどうかを判断し，spamメールでないと判断したものだけを利用者に提供する方法である．この方法で用いられる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が挙げられる [1]．このうち，ルールベースフィルタ及びベイジアンフィルタは，spamメールによく現れる特徴を予めルールとして記述したり，学習させたりしておき，その特徴に基づいて受信した電子メールがspamメールかどうかを判定する技法である．これらの技法はspamメールを正しく判断する検出率は比較的高いが，非spamメールを誤ってspamメールと判断 (false positive) する誤検出率も無視できず，実用上大きな問題となっている．一方，分散協調型フィルタは各利用者がspamメールと認識した電子メールを利用者間で共有するデータベースに登録するようにしておき，データベースへの登録の有無により電子メールがspamメールかどうかを判定する技法である．この技法は検出率は比較的低いが，誤検出率は事実上無視できるほど小さい点で注目されている．

本稿ではこれらの技法のうち分散協調型フィルタを対象とし，その検出率を向上させる手法を提案する．具体的には，従来ではMDA(Mail Delivery Agent) による電子メール受信時に行っていたデータベースとの照合を，MUA(Mail User Agent) による電子メール取得時にまで遅らせるようにする．これにより，他の利用者による同一spamメールのデータベース登録に時間的な余裕が生じ，特に宛先に複数のアドレスが指定されているspamメールなど，従来はほぼ同時にMDAが受信したためにデータベースへの登録が間に合わなかったspamメールについても検出できる効果が期待される．

2 従来のspamフィルタと問題点

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フィルタで用いられている技法を示し，その問題点を明らかにする．

2.1 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要性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[2]が挙げられる．

この技法では経験的に判明したspamメールの特徴をルール化するため，典型的なspamメールの検出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検出する危険性が無視できない点も問題である．

2.2 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 メールを検出する技法である [3]．すなわち，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[4] 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 メールを誤検出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 メールの検出率が低下して誤検出率が上昇してしまう現象が報告されている [5]．

2.3 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．すなわち，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)[6]，Razor[7]，Pyzor[8] 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．MDA 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 メールと見做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，通常メールと見做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，MUA 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．

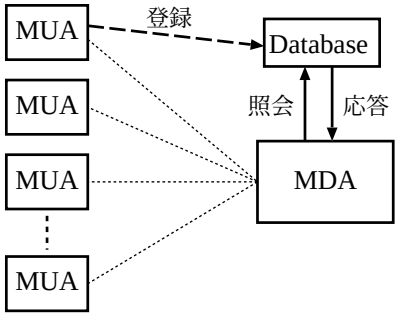

図 1: 分散協調型フィルタの処理過程

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，また検査時までに他の利用者がデータベースに登録しないと spam メールを検出できないため，検出率は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る可能性を持ち，誤検出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 分散協調型フィルタの改良

前節で述べたように，分散協調型フィルタは誤検出率が事実上無視できるという，他の技法にはない特徴を持つ．そこで，本研究では分散協調型フィルタをもとに，その問題点である検出率を改善する技法を提案する．

3.1 提案方式の概要

3.1.1 遅延評価の導入

従来の分散協調型フィルタでは，殆どの場合 MDA にフィルタ機能が導入されており，電子メール受信

時にデータベースとの照合が行われていた．したがって，特に 1 通につき多数の宛先アドレスが指定されているものや発信時の宛先アドレスがメーリングリストに対応するものなど，多数の利用者がほぼ同時に受信する spam メールについては，データベースに登録されないうちに照合が行われるため，これを spam メールと判定することができず，これが分散協調型フィルタにおける検出率低下の要因の 1 つになっていると考えられる．

そこで，本稿ではデータベースとの照合を MDA におけるメール受信時ではなく，MUA による電子メール取得時まで遅延させる方式を提案する．これにより，上記のようにたとえほぼ同時に受信した spam メールであっても MUA による電子メール取得時刻が利用者毎に異なれば，データベースに登録する時間的余裕が生じるため，2 番目以降に電子メールを取得する利用者がこの spam メールを検出することが可能になり，検出率の改善効果が期待できる．

なお，MDA においてメールボックスの件数やサイズによる受信制限が行われている環境では，遅延評価の導入によりメールボックスが満杯になる危険性が生じるという問題がある．これに対しては，メール受信時におけるフィルタリングと併用することにより解決可能である．

3.1.2 MDA 側へのフィルタ機能の導入

データベースとの照合を MUA による電子メール受信時まで遅延させる方法として，MUA に分散協調型フィルタを導入する方法が考えられる．しかし，この方法では最終的には破棄される spam メールまでも一旦 MUA まで転送する必要があるため，無駄な通信が生じることになる．これはダイアルアップ回線や携帯電話を用いて電子メールを受信する場合など，通信回線が低速であったり通信時間や通信量に応じて課金されたりした場合に通信費用や時間を浪費することにつながる．

そこで，提案方式では MUA による電子メール取得時に MDA 側でフィルタリングを行い，通常メールと判定された電子メールのみを MUA に転送する技法を採用する．この技法により，MUA では既存のソフトウェアをそのまま用いることができ，提案方式の適用環境に制限が課せられなくなることも期待できる．

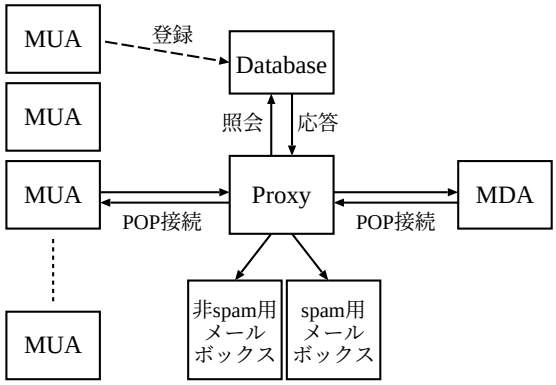

図 2: 試作システムの構成

3.2 提案方式の設計・実装

MUA による電子メール取得手段としては，POP (Post Office Protocol)[9]，IMAP (Internet Mail Access Protocol)[10] など，いくつかの方法が使われているが，これらのうち POP は殆ど全ての MUA でサポートされており，最も普及していると思われる．そこで，我々は上記の提案方式に基づき，POP アクセス時に spam メールをフィルタリングする機能を設計し，試作した．

試作システムでは，分散協調型フィルタは POP サーバとしてよく使われている qpopper[11] を改造し，POP プロクシとして実装した．これは稼動中の任意の POP サーバと組み合わせて利用することを可能にするためである．また，試作システムは MDA の動作には影響を与えないため，MDA におけるフィルタリングと併用することも可能である．なお，試作システムでは基本となるデータベースシステムに DCC を用いた．試作システムの構成を図 2 に示す．

本システムは，以下のように動作する．

(1) MUA は POP プロクシとの間で利用者認証を行う．このとき，もし USER 及び PASS コマンドを用いられた場合には，これを POP サーバに中継するように動作し，同時に POP プロクシ・POP サーバ間でも利用者認証を行う．もし APOP を用いられた場合には，POP プロクシは POP サーバと共有したパスワードファイルを用いて認証を行い，これに成功した場合に限り POP プロクシ・POP サーバ間でも利用者認証を行う．ただし，いずれの場合も認証成功後は直ちには結果を MUA に返

さない．

(2) POP プロクシは POP サーバとの間で LIST コマンド (一覧表示用)，RETR コマンド (メッセージ取得用)，DELE コマンド (メッセージ削除用) を用いて POP サーバに格納された電子メールを取得する．

(3) POP プロクシは DCC を用いて取得した各電子メールがデータベースに登録されているかどうかを確認する．もし，データベースに登録されている電子メールがあれば，これを spam 用メールボックスに格納し，データベースに登録されていない電子メールだけを非 spam 用メールボックスに格納する．

(4) POP プロクシは非 spam 用メールボックスを開いて電子メール受信件数を確認し，この件数とともに MUA に認証成功を通知する．これ以降は，通常の POP サーバとして動作する．

なお，本システムにおける通信過程を図 3 に示す．

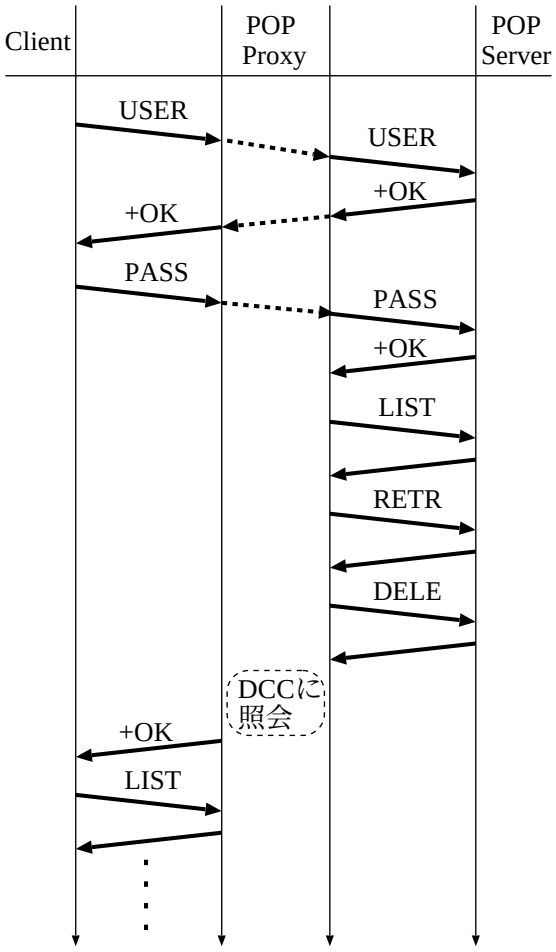

図 3: 試作システムの通信過程

4 性能評価

4.1 試作システムの spam メール検出率

試作システムの spam メール検出率は，提案方式の有効性を評価する上で最も重要な指標である．しかし，この値は受信する電子メールの種類，数やタイミング，試作システムの利用者数，各利用者が電子メールを取得するタイミングなど多くの要因に依存するため，現時点で正確に測定することは困難である．そこで，試作システムを運用する代わりに著者の一人がこれまでに受信した spam メールを分析し，期待される検出率を見積もることにした．

受信した spam メールは 3 つの個人用アドレス宛といくつかのメーリングリスト用アドレス宛に 2003 年 11 月 10 日から 2004 年 1 月 8 日までの間に送られたもので，SpamAssassin 2.60 により検出されたものと，受信者が見て判定したものからなる．このうち，DCC により同一 spam メールが受信した spam メール集合中に存在すると判定されたもの，宛先として複数のアドレスやメーリングリスト用アドレスが指定されているなど，他の利用者によるデータベースへの登録が期待できるものを分散協調型フィルタで検出可能であると見做して分類した．その結果を表 1 に示す．この表から，3 つの個人用アドレスに届いた spam メールしか収集していない状況にもかかわらず，全体では約 82%，SpamAssassin(表中の SA) では検出できなかった電子メールについても約 61%を正しく spam メールと判定できる可能性があり，ある程度の有効性が期待できる．また，SpamAssassin との併用により，誤検出率は同一のままで spam メール検出率を約 95%まで改善する可能性も期待できる．更に，これらの数値は分散協調型フィルタの利用者の増加に連動することが期待できるため，提案方式が普及すると更に増加することが期待される．

4.2 POP プロクシのオーバヘッド

提案方式では POP プロクシへのアクセス時に初めてフィルタリングするため，POP プロクシのオーバヘッドが利用者から見た応答時間に大きく影響を与える．そこで，我々は POP プロクシを用いた場

表 1: spam メールの分類

		提案方式		
		検出可能	検出不能	合計
S	検出可能	2,803	475	3,278
A	検出不能	314	205	519
合計		3,117	680	3,797

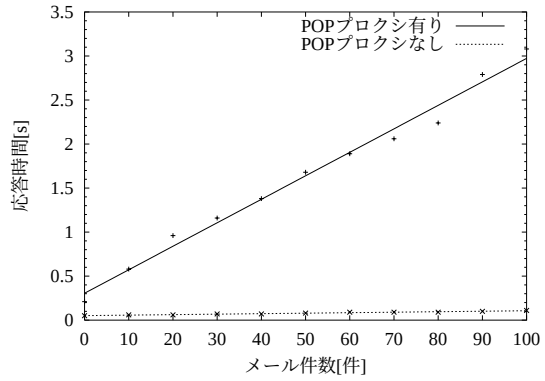

図 4: POP プロクシの有無による応答時間の違い

合と用いない場合の応答時間を比較して POP プロクシのオーバヘッドを評価するための実験を行った．

本実験では，POP サーバに 0〜100 通の電子メールを格納しておき，POP プロクシを用いた場合と用いていない場合において利用者認証用の PASS コマンドを入力してから応答が返されるまでの時間をそれぞれの場合について 100 回計測し，その平均値を求めた．その際，spam メールと非 spam メールの割合は 7:3 とし，電子メール 1 通あたりの平均サイズは約 4.3kB とした．その結果を図 4 に示す．

この結果より，電子メール 1 件当たりのオーバヘッドは高々0.03 秒程度であることがわかる．これは実用上十分に小さい値であるといえる．

5 まとめ

本稿では，誤検出率が事実上無視できるという特徴を持つ分散協調型 spam フィルタを対象とし，その問題点であった spam メール検出率を向上させるため，同フィルタにおけるデータベースとの照合を MUA による電子メール取得時まで遅延させる方式を提案した．また，POP プロクシを用いて提案方式を実装し，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．

今後の課題としては実運用を通じての提案方式の性能評価が挙げられる．この場合，メール受信時にも同一の分散協調型フィルタを適用しておき，その後 MUA によるメール取得時に POP プロクシにおいて検出できた spam メールの件数を測定すれば遅延評価の効果を正しく検証できると思われる．

また，spam メールの発信者アドレスとして利用者自身のアドレスに詐称された場合など，ほぼ同一内容のエラーメールが短時間に多数届く状況に対しても，提案方式ではたとえば一度に MUA が取得するメッセージ数を 10 通程度に制限するなどの方法で対応することが可能である．今後はこのような状況に対する有効性についても検証したい．

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．

参考文献

[1] David Mertz: “Six approaches to eliminating unwanted e-mail”, http://www.ibm.com/developerworks/linux/library/l-spamf.html, September 2002.

[2] “SpamAssassin: Welcome to SpamAssassin”, http://www.spamassassin.org/index.html.

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

[4] Eric S. Raymond: “Bogofilter Home Page”, http://bogofilter.sourceforge.net.

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

[6] Rhyolite Software: “Distributed Checksum Clearinghouse”, http://www.rhyolite.com/anti-spam/dcc/.

[7] Vipul Ved Prakash: “Vipul’s Razor: home”, http://razor.sourceforge.net/.

[8] “Pyzor”, http://pyzor.sourceforge.net/.

[9] J. Myers, M. Rose: “Post Office Protocol - Version 3”, RFC1939, May 1996.

[10] M. Crispin: “INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1”, RFC 3501, March 2003.

[11] QUALCOMM Incorporated: “Qpopper Information ”, http://www.eudora.com/qpopper/.