チャレンジ–レスポンスとベイジアンフィルタリングを併用した迷惑メール対策の提案

岩永 学[†] 田端利宏[††] 櫻井幸一[††]

迷惑メール対策のアプローチの 1 つとして，ホワイトリストに登録されている送信者からのメールのみを受信者に表示し，ホワイトリストに含まれていない送信者に対してはチャレンジ–レスポンスによってホワイトリストへの登録を求めるという方式がある．エラーメールは受け取る必要のある正当な電子メールであるにもかかわらず，エラーメールの送信者である MTA（Mail Transfer Agent）はこの登録を行うことができず，このままではエラーメールを表示することができない．したがってエラーメールはチャレンジ–レスポンスの例外として受け入れる必要がある．その一方で，エラーメールの形式をとる迷惑メールも存在し，このような電子メールを受け取ってしまうという問題がある．本論文ではエラーメールの形式をとる迷惑メールへの対策として，チャレンジ–レスポンスとベイジアンフィルタリングを併用する方式を提案し，この方式による迷惑メール対策の効果について評価を行う．

Proposal of Anti-spam Scheme Combining Challenge-response and Bayesian Filtering

Manabu Iwanaga,[†] Toshihiro Tabata[††] and Kouichi Sakurai[††]

Some anti-spam schemes are based on challenge-response, a principle that a recipient reads only messages from senders who are registered by the recipient. In these schemes, request for setup is sent to senders who are not registered. Since bounce messages are legitimate but MTA cannot reply to request, we should have some exception to receive for them. However, spammers can abuse this exception to send spam to users, disguising their spam with bounce messages. In this paper, we propose an improved scheme, combining challenge-response and Bayesian filtering, then perform some tests on the effect of our scheme to avoid those spam.

1. はじめに

近年の電子メールの普及にともない，送信に要する費用の少なさから迷惑メールが急増している．迷惑メール送信者による送信の手口は巧妙化しており，送信者アドレスの詐称をはじめとした電子メールのヘッダなどの偽造も行われている．このため，ヘッダの内容に対する単純な検査だけでは，迷惑メールを十分に検出することは難しい．

様々な迷惑メール対策方式が提案され，使用されている．迷惑メール対策のアプローチの 1 つとして，ホワイトリストがある．これは，受信者の持つリストに登録されている送信者からの電子メールのみを受信者に表示するというものである．ホワイトリストには，正当な目的を持った新たな送信者のために，チャレンジ–レスポンスと呼ばれる自動登録システムが一般に併用される．

ホワイトリストを使用する場合，そのままではエラーメール（バウンス）を受信することができない．エラーメールは MTA により作成されるが，一般に MTA はチャレンジ–レスポンスに対応していないからである．送信した電子メールが正常に配送されなかったことを知るためには，エラーメールをチャレンジ–レスポンスの例外として受け入れる必要がある．しかしエラーメールを無条件に受け入れると，迷惑メールをエラーメールに見せかけることにより，迷惑メール送信者は，受信者の迷惑メール対策を回避することができる．エラーメールの形をとった迷惑メールの問題としては，論文 1) などで指摘されている，送信元メールアドレスを詐称した迷惑メールによるエラーメール

† 九州大学大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University
†† 九州大学大学院システム情報科学研究院
Faculty of Information Science and Electrical Engineering, Kyushu University

の問題がある．また，Mydoom ウィルス[2] のように，自身の複製を拡散する際にエラーメールを装った電子メールを作成するウィルスも現れており，エラーメールを装った迷惑メールに受信者を注目させ，その内容を読ませようとする迷惑メールについても対策を行う必要がある．

本論文では，迷惑メール送信者が迷惑メールをエラーメールに見せかける方法を述べ，チャレンジ–レスポンスとベイジアンフィルタリングを併用する迷惑メール対策方式を提案する．また実験により，迷惑メールや正当な電子メールに対する提案方式の精度について評価した結果を報告する．

2. 用　　語

エラーメール　電子メールの配送に問題が生じた際に，送信者へ通知するための電子メール．受信者メールアドレスの不存在，受信者ドメインの不存在，メールサーバへの接続不能などの理由で，電子メールが正常に配送されなかった場合に，MTA によって作成される．エラーメールには配送状況の通知のほか，多くの場合正常に配送されなかった電子メールの本文が添付される．バウンスメールとも呼ばれる．

通常の迷惑メール　受信者の承諾のないまま大量の受信者に対して送信される電子メールのうち，エラーメールの形式をとらないもの．

迷惑エラーメール　通常の迷惑メールの配送が失敗した際に，送信者アドレスとして詐称された利用者に送り返されるエラーメール，および，迷惑メール送信者によって偽造されたエラーメール．

迷惑メール　通常の迷惑メールおよび迷惑エラーメール．

正当なエラーメール　迷惑エラーメールでないエラーメール．

正当な通常の電子メール　迷惑メールでない電子メールのうち，エラーメールでないもの．

正当な電子メール　正当な通常の電子メールおよび正当なエラーメール．

誤検出　正当な電子メールが誤って迷惑メールと判定され，受信者に表示されないこと．

見逃し　迷惑メールが誤って正当な電子メールと判定され，受信者に表示されること．

非応答　登録要求の電子メールが送信され，なおかつその送信から一定期間のうちに送信者が応答しないこと．登録を要求する電子メールが送信者に配送されなかった，登録を要求する電子メールを送信者が無視したなどの理由によって起こる．

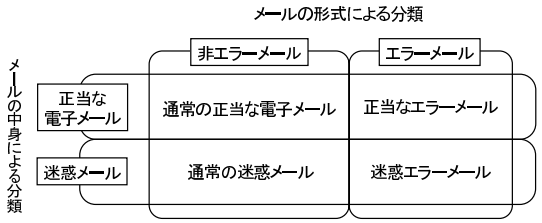

図 1　電子メールの分類の概念
Fig. 1　Notion of classification of email.

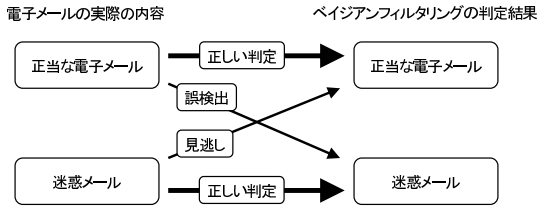

図 2　誤検出と見逃しの概念
Fig. 2　Notion of false-positive and false-negative.

非応答率　正当な通常の電子メールに占める，送信者が応答しないことにより誤って迷惑メールと見なされる電子メールの割合．

元の電子メール　エラーメールや，チャレンジ–レスポンスにおける登録要求の，発生原因となった電子メール．

迷惑メール確率　ある電子メール，もしくはある単語を含む電子メールが，統計的にみて迷惑メールである確率．

エラーメール，通常のメール，迷惑メール，正当な電子メールなどの概念は**図 1** のように図示することができる．また，誤検出と見逃しの関係については**図 2** のように表すことができる．

3. 既存の迷惑メール防止方式

3.1 メール内容による方式

迷惑メールに対する単純な対策としては，本文やヘッダに含まれる特徴を基にしたフィルタリングがよく用いられている．従来，メール内容によるフィルタリングは Spamassassin[3] のようにルールベースのものが中心だったが，近年，Graham の論文[4),5)] をはじめ，ベイジアンフィルタリングとよばれる統計的手法がよく用いられるようになってきた．ベイジアンフィルタリングでは，電子メール中に現れる単語（token）の正当な電子メールと迷惑メールのそれぞれにおける出現確率を求め，この確率をもとに判定対象の電子メールの迷惑メール確率を計算し，この確率が一定の

閾値を超えたものを迷惑メールと判定する．ベイジアンフィルタリングの特徴として，判定した電子メールに含まれている単語を学習し出現確率のデータを更新することにより，以後の電子メールの判定精度を改善する点がある．

3.2 チャレンジ–レスポンス

Gabberら[6]やHall[7]，Jakobssonら[8]，Mailblocks[9]の方式などにおいては，チャレンジ–レスポンスとよばれる手法が使用されている．この手法では，ホワイトリストに含まれている送信者からのメールのみが利用者に表示され，それ以外の送信者からの電子メールに対しては自動的に登録作業の実行を要求する電子メール（チャレンジ）が返信される．登録要求を受け取った送信者が登録作業を行うと（レスポンス），この送信者は自動的にホワイトリストへ追加される．しかし，迷惑メールのように機械的かつ大量の受信者に対して送信されている電子メールの場合，送信者が登録作業の実行を要求する電子メールを1通ずつ読んで登録作業を行うことは，ほとんど不可能である．登録作業が行われない場合，保留されていた電子メールは廃棄される．

登録作業の例としては，電子メールの本文中に記述されたWebページへアクセスし，ブラウザに表示される画像の中の文字をテキストボックスに入力して送信するという手法がある．この手法では，コンピュータによる文字認識が難しく，なおかつ人間には容易に読み取れるような文字を表示することで，人間による手作業の登録作業とコンピュータによって自動的に行われる作業とを識別しようとしている．

また，数学的に証明可能な計算量的コストを送信者の計算機に支払わせるという手段[10]を用いた方式[6)~8)]も存在する．計算量的コストを用いたこれらの手法は，迷惑メール送信者は大量送信を行うことでコストを低減するという点に着目し，電子メールの送信者に，送信する電子メールの数に比例した計算量的コストを負担させることで大量送信を防ぎ，迷惑メール送信をコストの面から非効率なものにしようとしている．

チャレンジ–レスポンスの1つであるJakobssonらの方式[8]は，送受信者間の共通鍵と送信するメッセージから生成されるMAC（Message Authentication Code）を電子メールのヘッダに付加し，受信側でこのMACを検査することで，チャレンジ–レスポンスを電子メールの盗聴・改竄に対して安全なものにしている．この方式ではチャレンジ–レスポンスや送受信時のMACの作成・検査のために，送受信者の計算機上にそれぞれメールプロキシとよばれるソフトウェアを導入し，メールソフトの送受信時にメールサーバとの間に割り込む形で動作する．メールプロキシは自らの動作する計算機以外からの接続を受け付けない．送信者は登録作業を行う際メールプロキシを用いて，計算量的コストを消費したことを証明する情報 $cookie$ と，自らが作成した公開鍵 y_S の対を受信者に送信する．受信者は $(cookie, y_S)$ を受信すると，共通鍵 K_{RS} を作成し，y_S で暗号化したうえで送信者に送る．以後この送信者が電子メールを送信する際には，送信側のメールプロキシが K_{RS} を用いて送信するメッセージに対するMACを計算し，電子メールのヘッダ部分にその値を付加する．受信側のメールプロキシはメッセージと K_{RS} から再度MACを計算し，MACの値が一致すれば，この電子メールを正当な送信者からのものとして利用者に表示する．MACはメッセージ内容から計算されるため，電子メールの通信路上での改竄は受信者のメールプロキシによって検出される．

メールプロキシが送信に関与する電子メールは，利用者が送信しようとする電子メール，および未知の送信者からの電子メールに対する登録要求であり，外部の攻撃者が利用者のメールプロキシを利用して第三者に任意の電子メールを送信することはできない．また，登録要求の中に元の電子メールの内容を含むと，詐称された電子メールアドレスに対して登録要求を送った際に登録要求自体が迷惑メールとなってしまうことから，登録要求は元の電子メールの内容を含んではならないといえる．

Jakobssonらの方式において，受信者側のシステムは以下のような順序で受信した電子メールの扱いを決定する．

(1) 正当なMACが付属しているならば，受信者に表示する．
(2) 登録要求の電子メールや，それに対する登録作業の電子メールならば，自動的に登録を行う．
(3) どちらでもなければ，送信者に登録要求のメッセージを返信し，受信者には表示しない．この方式では通信路上での電子メールの改変を防ぐため，登録作業が行われる前に送信された電子メールは登録作業終了後に再送される必要がある．

4. エラーメールの形式をとった迷惑メール

4.1 エラーメール

電子メールの配送において，送信者からSMTP接続を受けて電子メールの配送を依頼されるMTAと，受信者のメールボックスを管理するMTAは異なるド

メインに属することが多く，送信者が受信者のメールボックスを管理する MTA に直接 SMTP 接続を行い電子メールを送信することは少ない．このため，多くの場合において，送信者は自らの属する組織の MTA にメールの配送を依頼し，この MTA から受信者のメールボックスを管理する MTA へ電子メールを配送することになる．

電子メールの配送は正しく受信される保証のないまま行われるので，宛先のアカウントが存在しない，受け取りが拒否された，MTA へ接続できないなどの理由で，電子メールの配送が失敗または遅延する可能性がある．インターネットにおける SMTP を用いた電子メールの配送について規定した RFC2821 [11] では，MTA は電子メールの配送に問題が生じた場合，送信者にエラーメールを送信しなければならないと定めている．エラーメールには，問題の発生と原因を伝える文章，エラー発生時の通信ログ，配送に問題が生じた電子メールの内容などが含まれるが，MTA として使用されているソフトウェアや設定などによってその書式は異なる．

4.2 エラーメールの形式をとった迷惑メール

エラーメールは電子メールの配送における問題を送信者に伝えるための重要な電子メールであり，通常の正当な電子メールと同様に，迷惑メール対策によって安易に失われてはならない．しかし，迷惑メールがエラーメールの形式をとって送られてくることがある．送信者メールアドレスの詐称による迷惑エラーメールの概念を**図 3** に示す．

迷惑メールがエラーメールの形式をとって配送される原因として，次のような可能性が考えられる．

(1) **間接送信**：迷惑メールの送信者が送信者メールアドレスの詐称を行い，その迷惑メールの配送が失敗することによって，詐称されたメールアドレスに送信先の MTA からエラーメールが届く．迷惑メール送信者の意図により，以下の 2 通りに分けて考えることもできるが，本研究では区別せず考える．

- 迷惑メール送信者が，自分の身元を隠す，送信した迷惑メールを読まれやすくするなどの理由から詐称を行い，配送の失敗により結果的に関係ない利用者にエラーメールが届く．
- 迷惑メール送信者が，迷惑メールをエラーメールとして利用者へ送ることを目的として，存在しないメールアドレスに向けて迷惑メールを送る．

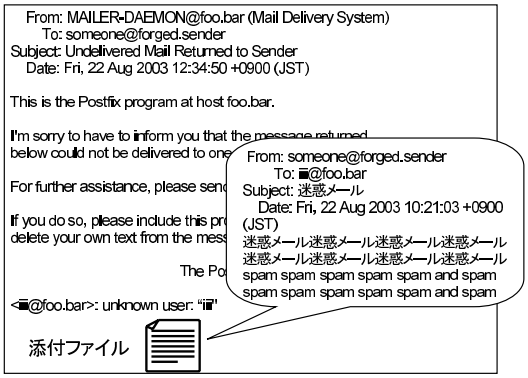

図 3 送信者メールアドレスの詐称による迷惑エラーメールの概念
Fig. 3 Notion of disguised spam with forged sender address.

(2) **直接送信**：迷惑メール送信者がエラーメール自体を偽造し，そのエラーメールの一部として迷惑メールを利用者に送信する．

間接送信による迷惑エラーメールは，エラーメールの作成自体は正規の MTA によって行われる．また，直接送信によるエラーメールについても，受信した MTA によって追加される received ヘッダ以外のヘッダや本文は迷惑メール送信者によって自由に作成できる．したがって，ヘッダの形式を検査するだけでは正当なエラーメールと迷惑エラーメールを区別することはできない．また，各利用者において自らが送信したメールを蓄積し，受信したエラーメールと照合することによって，利用者が送信したことのない電子メールに対するエラーメールを排除することが考えられる．しかし，迷惑メール送信者が盗聴を行うことができるならば，利用者の送信する電子メールを盗聴し，その電子メールに対するエラーメールを偽造することができる．このエラーメールは利用者が送信した電子メールに対するエラーメールに見えるため，迷惑メール送信者は送信したメールとエラーメールとの照合を回避できる．

迷惑メールを十分に高い確率で検出するためには，正当なエラーメールと迷惑エラーメールを何らかの方法を用いて区別する必要があるが，以上のように，チャレンジ–レスポンスのみを用いた迷惑メール対策では，迷惑エラーメールに対応できない．

5. 提 案 方 式

チャレンジ–レスポンスを用いた迷惑メール対策において，エラーメールに正しく対応するためには以下の条件が必要である．

- 迷惑メールと関係のない，正当なエラーメールを

受け入れること．
- エラーメールの形をした迷惑メールを排除すること．
- エラーメール以外の，通常の電子メールについては従来どおりチャレンジ–レスポンスを行うこと．

正当なエラーメールと迷惑エラーメールを区別するためには，電子メールの内容に基づくフィルタリングが必要である．そこで，ベイジアンフィルタリングを行うことが考えられる．すなわち，エラーメールのうち，配送に問題が生じた電子メールの内容が正当な電子メールのものであるか，あるいは迷惑メールのものであるかをベイジアンフィルタリングで判別する方法である．本論文で提案する方式では，Jakobsson らの方式[8] をもとに，エラーメールに対してはベイジアンフィルタリングを行い，正当なエラーメールと迷惑エラーメールのどちらであるかを判定する．

ベイジアンフィルタリングが高い精度で判別を行うためには十分多くの電子メールを学習する必要があるが，エラーメールは通常のメールに比べ一般に数が少ないため，単純にエラーメールのみを学習・判定対象とするのでは十分な精度が得られないことが考えられる．そこで，提案方式では
- 正当なエラーメールに含まれている元の電子メールと，通常の正当な電子メール
- 迷惑エラーメールに含まれている元の電子メールと，通常の迷惑メール

はそれぞれ同じ語彙によって構成されていると見なし，チャレンジ–レスポンスによって正当な電子メール，もしくは迷惑メールと判定された電子メールについて，チャレンジ–レスポンスによる判定結果をもとにベイジアンフィルタリングの学習対象とする．また，ベイジアンフィルタリングが最初に持つ学習データについても，エラーメールか否かに関係なく，任意の正当な電子メールや迷惑メールから構成する．このようにすることで，エラーメールに対するベイジアンフィルタリングは，全電子メールに対してベイジアンフィルタリングを行った場合と同じ量の学習データをもとに判定を行うことができる．

具体的には，受信者側のシステムは，受信した電子メールの扱いを以下のような手順で決定する（処理の流れを**図 4** に示す）．

(1) 正当な MAC が付属しているならば，ベイジアンフィルタリングに正当な電子メールとして学習させ，受信者に表示する．

(2) 登録要求の電子メールや，それに対する登録作業の電子メールならば，自動的に登録を行う．

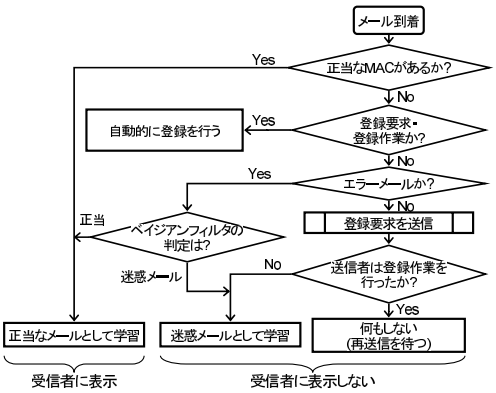

図 4 提案方式における電子メールに対する処理の流れ
Fig. 4 Processing flow for incoming message by our scheme.

(3) エラーメールの形式であるならば，ベイジアンフィルタリングで判定し，判定結果に従って学習させる．正当と判定されれば正当なエラーメールとして受信者に表示し，迷惑メールと判定されれば破棄する．

(4) どれにもあてはまらないならば，送信者に登録要求のメッセージを返信し，受信者には表示しない．その後一定期間内に登録が行われないならば迷惑メールとして学習させるが，登録が行われた場合は送信者によって電子メールの再送が行われるので何もしない．

本方式の利点は，登録された正当な送信者からのメールについては確実に受信者に表示するというチャレンジ–レスポンスの長所を保ったまま，ベイジアンフィルタリングによる精度で，迷惑エラーメールの排除も行うことができる点である．また，Jakobsson らの方式[8] は迷惑メール送信者による盗聴に耐えることを目標に設計されているので，エラーメールの判別もまた盗聴に耐える方式である必要がある．本方式ではエラーメールの判別においてメールの内容に基づく検査を行っており，盗聴によって迷惑メール対策を回避することは難しい．

本提案方式はチャレンジ–レスポンスとベイジアンフィルタリングを組み合わせた方式である．提案方式と，チャレンジ–レスポンスのみを単独で用いる方式を比較した場合，チャレンジ–レスポンス単独ではエラーメールを正当なエラーメールと迷惑エラーメールに判別する手段を持たない．提案方式はエラーメールに対してベイジアンフィルタリングを行い，正当なエラーメールと迷惑エラーメールを判別しているので，明らかに提案方式のほうが高い精度で正当な電子メールと迷惑メールを判別できる．

次に，提案方式と，ベイジアンフィルタリングのみを単独で用いる方式を比較する．正当な送信者が必ず提案方式における登録作業を行うならば，提案方式は通常の電子メールに対して誤検出や見逃しを行わないので，より高い精度で正当な電子メールと迷惑メールを判別できる．一方エラーメールに対してはともにベイジアンフィルタリングによって判定を行う．ここで，通常の電子メールに対する両方式の判別精度の差が原因となり，エラーメールの判別精度に差を生じることが考えられるが，この点については必ずしも明らかではない．

このため，提案方式と，ベイジアンフィルタリングのみを単独で用いる方式について，エラーメールに対する精度の評価を行う．

6. 実　　験

6.1 実験方法

本論文で提案した方式によって得られる効果を確認するためにいくつかの実験を行う．本論文の実験ではチャレンジ–レスポンスについて以下のモデル化を行い，本論文で提案する方式の，迷惑エラーメールに対する有効性を評価する．本実験では迷惑エラーメールのうち特に間接送信によるもののみを扱う．

本実験では，以下のような仮定を行った．

- 非応答率を p(%) とする．つまり，正当な通常の電子メールのうち p(%) については，送信者が登録作業を行わなかったために迷惑メールとして学習されるものとする．
- 迷惑メールの送信者は，チャレンジ–レスポンスの登録要求にまったく応じない．つまり，すべての通常の迷惑メールは迷惑メールとして学習される．
- 正当なエラーメールには，通常の正当な電子メールと同様の内容が添付される．また迷惑エラーメールには，通常の迷惑メールと同様の内容が添付される．

実験に用いた電子メールは，正当な電子メール 773 通，迷惑メール 176 通である．どちらも主に本文が日本語の電子メールであり，また若干の英文メールが含まれている．本実験においては**表 1** に示すように電子メールに占めるエラーメールの割合を 2 通り設定し，それぞれについて実験を行った．

本実験で使用するエラーメールは 3 パートからなる MIME マルチパート形式の電子メールであり，ヘッダおよび第 1・第 2 パートには MTA の 1 つである postfix [12] が送信するエラーメールと同様の一定の内容を，第 3 パートには通常の電子メールをそれぞれ含む電子メールを作成した．また，エラーメールの作成時にエラーメール内に挿入した通常の電子メールは，以後の実験において通常の電子メールの集合から除外した．

表 1 実験に用いる電子メールの数および初期学習を行う電子メールの数
Table 1 Number of messages used in tests and number of messages learned initially.

条件	初期学習比率	電子メールの形式・種類			
		非エラー		エラー	
		正当	迷惑	正当	迷惑
A	（全電子メール数）	*731*	*121*	*42*	*55*
	1/2	365	60	20	27
	1/5	146	24	8	10
B	（全電子メール数）	*754*	*121*	*19*	*55*
	1/2	376	60	9	27
	1/5	150	24	3	10

実験は以下のような手順を繰り返して行い，10 回繰り返したときの平均値を結果として用いた．

(1) 正当な電子メールのうち一定割合の電子メールをランダムに選び，ベイジアンフィルタリングに正当な電子メールとして学習させる．

(2) 迷惑メールのうち一定割合の電子メールをランダムに選び，ベイジアンフィルタリングに迷惑メールとして学習させる．

(3) 上で選ばれなかった正当な電子メール・迷惑メールをランダムに並べ替える．

(4) **ベイジアンフィルタリング単独方式の場合**：(3) でまとめた電子メールを 1 通ずつベイジアンフィルタリングに判定させ，同時にベイジアンフィルタリングの判定に基づいて学習させる．

提案方式の場合：(3) でまとめた電子メールを以下の手順で 1 通ずつ処理する．

(a) 通常の正当な電子メールならば，確率 p（＝非応答率）でベイジアンフィルタリングに迷惑メールとして学習させ，そうでなければ正当な電子メールとして学習させる．提案方式の実際の運用においては，未登録の送信者からの最初の電子メールは表示されず，登録作業の後同じ電子メールが再送されることによって，登録済みの正当な送信者からの電子メールとしてベイジアンフィルタリングに学習され，受信者に表示される．最初の電子メールはベイジアンフィルタに影響を与えないことから，本実験においては簡単のため，再送された電子メールのみを考慮することによって，すでに登録され

表 2 方式による誤検出数・見逃し数の変化
Table 2 Number of false-positives and false-negatives by Bayesian-only scheme and our proposed scheme.

条件	初期学習	方式	非応答率 p(%)	誤判定の回数（確率）			
				通常のメール		エラーメール	
				誤検出	見逃し	誤検出	見逃し
条件A	1/2	単独	-	0.8(0.21%)	8.9(14.51%)	0.0 (0.00%)	4.2 (14.94%)
		併用	0	−*	−*	0.0 (0.00%)	3.2 (11.43%)
	1/5	単独	-	0.6(0.10%)	31.5(32.44%)	0.0 (0.00%)	17.6 (39.00%)
		併用	0	−*	−*	0.0 (0.00%)	11.4 (25.33%)
条件B	1/2	単独	-	0.7(0.19%)	9.0(14.67%)	0.0 (0.00%)	3.8 (13.69%)
		併用	0	−*	−*	0.0 (0.00%)	2.9 (10.36%)
	1/5	単独	-	0.4(0.07%)	32.4(33.42%)	0.0 (0.00%)	17.6 (39.15%)
		併用	0	−*	−*	0.0 (0.00%)	9.1 (20.22%)

∗ · · · チャレンジ–レスポンスを併用する方式では，通常の電子メールはチャレンジ–レスポンスのみで処理されるため，非応答率 0 の場合，通常の電子メールに対するベイジアンフィルタリングによる誤検出・見逃しは存在しない．

表 3 非応答率による誤検出数・見逃し数の変化
Table 3 Number of false-positives and false-negatives for unresponsive rate.

初期学習	方式	非応答率 p(%)	エラーメールに対する誤判定の回数（確率）			
			条件 A		条件 B	
			誤検出	見逃し	誤検出	見逃し
1/2	併用	10	0.0 (0.00%)	2.8 (10.00%)	0.0 (0.00%)	2.4 (8.57%)
		20	0.0 (0.00%)	1.8 (6.43%)	0.0 (0.00%)	1.8 (6.43%)
		30	0.0 (0.00%)	1.8 (6.43%)	0.0 (0.00%)	0.8 (2.86%)
		40	0.1 (0.46%)	1.5 (5.36%)	0.2 (2.00%)	0.7 (2.50%)
		50	0.2 (0.91%)	1.0 (3.57%)	0.2 (2.00%)	1.0 (3.57%)
1/5	併用	10	0.0 (0.00%)	6.8 (15.11%)	0.0 (0.00%)	5.8 (12.89%)
		20	0.0 (0.00%)	6.2 (13.78%)	0.4 (2.50%)	3.2 (7.11%)
		30	0.0 (0.00%)	3.2 (7.11%)	0.5 (3.13%)	2.2 (4.89%)
		40	0.3 (0.88%)	4.2 (9.33%)	0.4 (2.50%)	3.5 (7.78%)
		50	0.2 (0.59%)	3.0 (6.67%)	1.0 (6.25%)	4.4 (9.78%)

ている送信者と新たに登録される正当な送信者を同一視している．

(b) エラーメールならばベイジアンフィルタリングに判定させた後，判定に基づいて学習させる．

(c) 通常の迷惑メールならばベイジアンフィルタリングに迷惑メールとして学習させる．

ここで (c) は送信者による非応答をモデル化したものである．

実験ではベイジアンフィルタリングとして ruby を用いた実装の 1 つである bsfilter（revision 1.34）[13] を使用し，迷惑メール確率の計算には Graham の方式（単語ごとのベイズ確率が 1/2 から最も遠い単語 15 個から迷惑メール確率を求める）を用いている．閾値は文献 4) において提案された経験的な閾値である 0.9 をそのまま用いている．

6.2 考　　察

表 2 は，ベイジアンフィルタリングだけを用いた場合と提案方式の迷惑エラーメールに対する判定精度の実験結果である．表 2 より，チャレンジ–レスポンスとベイジアンフィルタリングの併用方式を使用することで，正当なエラーメールに対する誤検出は増えずに，迷惑エラーメールに対する見逃しのみが減少していることが分かる．チャレンジ–レスポンスとベイジアンフィルタリングの併用により，エラーメールの判別精度は向上しているといえる．

次に，**表 3** は正当な電子メールの非応答率 p を変化させた場合の実験結果である．非応答率 p が増加するにつれて誤検出は若干増加しているが，見逃しは減少している．

正当な電子メールが非応答により迷惑メールとして学習されることで，その電子メールに含まれていた各単語について，その単語に対する迷惑メール確率が上昇する．これらの単語を含む他の正当な電子メールは，これらの単語の迷惑メール確率が上昇することで誤検出を受けやすくなる．逆に，従来見逃していた迷惑メールのうちこれらの単語を含むものについては，そ

れらの単語に対する迷惑メール確率が上昇することで見逃しが起きにくくなる．誤検出の増加と見逃しの減少は以上のような理由によるものと考えられる．いずれの条件でも誤検出率は非応答率に比べ十分小さいので，チャレンジ–レスポンスとベイジアンフィルタリングを併用する方式では，非応答率の増加による誤検出の増加は無視できるといえる．

また，表 2，3 を通じて，初期学習が少ない場合に見逃しが大きく増えることが分かる．初期学習が少ない場合，

- 迷惑メール確率を求めることのできる単語が少なくなる，
- 各単語に対して計算される，単語の迷惑メール確率の精度が低くなる，

という問題が生じるため，誤った判定は全体的に増加している．

Graham の計算方式では，単語の迷惑メール確率の精度の低下を避けるため，過去に 5 回以上出現した単語のみを確率計算に用いているが，誤検出を少なくするために，正当な電子メールにおける単語の出現回数を 2 倍にして計算している．つまり，迷惑メールにのみ現れた単語は 5 回出現するまで確率計算に用いられないのに対して，正当な電子メールにのみ現れた単語は 3 回現れた時点で確率計算に用いられるようになる．初期学習が少ない場合に見逃しが大きく増加するのは，この差によるものと考えられる．

ベイジアンフィルタリングでは，正当な電子メールと迷惑メールの区別は，計算された確率と，正当な電子メールと迷惑メールを分類する閾値との比較によってなされる．見逃しを減らすためにこの閾値を下げると誤検出が増え，誤検出を減らすためにこの閾値を上げると見逃しが増える．また，誤検出と見逃しでは，必要な電子メールが失われる危険性のある誤検出のほうがより重大な誤りであるので，誤検出と見逃しを同列にとらえて，その和が最小とする方法は最適とはいえない．Bogofilter[14] などの実装では，1 つの閾値で正当な電子メールと迷惑メールに分けるのではなく，正当な電子メールと断定するための閾値と，迷惑メールと断定するための閾値の 2 つを用い，2 つの閾値の間を「不確実」として計 3 種類に分類することによって，迷惑メール確率が高く計算された電子メールが大量の迷惑メールの中に埋没する危険を軽減している．ベイジアンフィルタリングをチャレンジ–レスポンスと併用する提案方式においても，このような手法が有効であることが考えられる．

7. ま と め

本論文では，チャレンジ–レスポンスに基づく迷惑メール対策だけではエラーメールの形式をとる迷惑メールに対応できない問題を指摘した．そして，チャレンジ–レスポンスとベイジアンフィルタリングを併用する迷惑メール対策方式を提案し，実験によりこの方式によって得られる電子メール判別の精度について評価を行った．チャレンジ–レスポンスとベイジアンフィルタリングを併用することにより，エラーメールについてもベイジアンフィルタリングの精度は向上しているといえる．

今後の課題として，安全かつ正当な送信者の負担が少ないチャレンジ–レスポンスの設計が考えられる．本論文で用いた Jakobsson らのチャレンジ–レスポンスは盗聴に対して安全である一方，送信者は送信に用いる計算機にソフトウェアをインストールする必要があることから，正当な送信者に要求される負担は必ずしも小さいとはいえない．非応答が多い場合，受信者は，破棄されるメールの中に正当な電子メールが含まれていないかを頻繁に確認することになり，チャレンジ–レスポンスによって得られる利点が大きく損なわれるからである．

謝辞 本研究の一部は，財団法人セコム科学技術振興財団平成 15 年度研究助成「インターネット妨害障害に対する暗号論的対策技術の研究」と文部科学省科学研究費補助金学術創成研究課題番号 14GS0218「社会基盤を構築するためのシステム LSI 設計手法の研究」（研究代表安浦寛人九州大学システム LSI 研究センター長）の支援を受けている．

参 考 文 献

1) 山井成良，山外芳伸，宮下卓也，大隅淑弘：発信者詐称 SPAM メールに対する対策手法，情報処理学会研究報告，分散システム/インターネット運用技術，Vol.22, No.9, pp.51–56 (2001).
2) 情報処理推進機構：セキュリティセンター：「W32/Mydoom」ウイルスに関する情報 (2004). http://www.ipa.go.jp/security/topics/newvirus/mydoom.html
3) Spamassassin. http://www.spamassassin.org/
4) Graham, P.: A Plan for Spam. http://paulgraham.com/spam.html
5) Graham, P.: Better Bayesian Filtering, *2003 Spam Conference*, Cambridge, Massachusetts. http://spamconference.org/proceedings2003.html
6) Gabber, E., Jakobsson, M., Matias, Y. and

Mayer, A.: Curbing Junk E-Mail via secure Classification, *Financial Cryptography '98*, Anguilla, British West Indies, International Financial Cryptography Association, LNCS 1465, pp.198–213, Springer (1998).
7) Hall, R.J.: Channels: Avoiding unwanted electronic mail, *1996 DIMACS Symposium on Network Threats*, Piscataway, New Jersey, pp.85–103, American Mathematical Society (1997).
8) Jakobsson, M., Linn, J. and Algesheimer, J.: How to Protect Against a Militant Spammer, Cryptology ePrint archive, report 2003/071 (2003).
9) Mailblocks. http://www.mailblocks.com/
10) Dwork, C. and Naor, M.: Pricing via Processing or Combating Junk Mail, *Cryptology-Proc. Crypto '92*, Santa Barbara, California, LNCS 740, pp.139–147, Springer-Verlag (1993).
11) Klensin, J. (Ed.): Simple Mail Transfer Protocol, Internet. RFC-2821 (Apr. 2001).
12) Postfix. http://www.postfix.org/
13) Bsfilter. http://www.h2.dion.ne.jp/˜nabeken/bsfilter/
14) Bogofilter. http://bogofilter.sourceforge.net/

(平成 15 年 12 月 4 日受付)
(平成 16 年 6 月 8 日採録)

岩永　学

2003 年九州大学工学部電気情報工学科を飛び級のため中退．同年より同大学大学院システム情報科学府情報工学専攻修士課程，現在に至る．迷惑メール対策，ネットワークセキュリティの研究に従事．電子情報通信学会会員．

田端　利宏（正会員）

1998 年九州大学工学部情報工学科卒業．2000 年同大学大学院システム情報科学研究科修士課程修了．2002 年同大学院システム情報科学府博士後期課程修了．2001 年日本学術振興会特別研究員．2002 年九州大学大学院システム情報科学研究院助手．博士（工学）．オペレーティングシステム，コンピュータセキュリティに興味を持つ．電子情報通信学会会員．

櫻井　幸一（正会員）

1988 年九州大学大学院工学研究科応用物理専攻修士課程修了．同年三菱電機（株）入社．現在，九州大学大学院システム情報科学研究院情報工学部門教授．1997 年 9 月より 1 年間コロンビア大学計算機科学科客員研究員．2001 年 4 月より九州大学システム LSI 研究センター併任．暗号理論・情報セキュリティ・社会情報工学の研究に従事．博士（工学）．2000 年情報処理学会坂井特別記念賞受賞．2000 年・2004 年情報処理学会論文賞受賞．電子情報通信学会，日本数学会，ACM 各会員．