# 既存のメール分類機構の組み合わせを可能にする機構の提案

山本泰隆　　乃村能成　　谷口秀夫

岡山大学大学院自然科学研究科

多様化し巧妙化する迷惑メールに対し，数多くの迷惑メール対策が提案されている．これら対策のうち，メール分類機構は，受信側メールサーバ上で行う方式であり，利用者単位での導入が容易であることから，普及が進んでいる．ここでは，メール分類機構の問題点を説明し，その対処の基本的な考え方を示す．さらに，その考え方を実現する機構として，複数のメール分類機構を併用可能にする機構を提案する．提案する機構は，既存の迷惑メール対策システムを分類処理の一部として利用可能にする．また，分類処理にメール分類機構の組み合わせのみを設定することで，利用者の設定作業を簡略化する．

# A Mechanism for Mixing up Existing Spam Filters

YASUTAKA YAMAMOTO,　YOSHINARI NOMURA　and　HIDEO TANIGUCHI

Graduate School of Natural Science and Technology, Okayama University

Since spam mails have been getting cleverer and various in kind, many anti-spam techniques have been developed. Among these anti-spam techniques, spam filters would be one of the most popular techniques. Because spam filters are usually implemented as receiver-side tools, and therefore, it is easy for e-mail users to introduce these tools on their own demand. In this paper, we describe the problems on existing spam filters, and show some ideas to cope with these problems. Also, we propose a mechanism for mixing up many existing spam filters based on our ideas. Our mechanism is able to include existing spam filters as a part of e-mail classification process. It looks simple and easy to understand for e-mail users. All they have to do is the choice of spam filters. It will reduce their troubles on setting up spam filters.

## 1 はじめに

電子メール（以降，メール）は，インターネットにおいてWWW（World Wide Web）と並んで最も普及しているサービスの一つである．一方で，通信コストの低さから，不特定多数の利用者に対して，同意を得ずに一方的に送信される迷惑メールが増加し問題になっている．迷惑メールには，宣伝や広告を目的としたものから，詐欺行為を目的としたものまで，様々な種類がある．迷惑メールの増加による利用者への被害として，メール受信量の増加による通信費用の増大，迷惑メールとそれ以外の正当な目的を持ったメール（以降，正当メール）との分類に要する時間の浪費がある．さらに，フィッシングに代表される詐欺行為に迷惑メールが利用されることもあり，利用者の被害はより深刻になっている．

これらへの対処として，現在では数多くの迷惑メール対策が提案され，利用されている．既存の迷惑メール対策は，迷惑メールの送信を防止する送信側における対策と迷惑メールの受信を防止する受信側における対策に分けられる．それぞれの対策の概略を以下に説明する．

送信側における対策は，メール送信者のメール投稿時，またはメールサーバ間でのメール配送時に行われる．具体的な対策としては，メール投稿時における送信者認証の追加[1]，意図的な配送遅延によるメール配送量の制御[2]，および外部ネットワークへのメール配送の制限がある．送信側における対策は，メール配送経路の送信者に近い位置での対策であるため，利用者が受信する迷惑メールの量を削減するだけでなく，通信路上を流れる迷惑メールの量を削減できる．しかし，送信側における対策の効果は，それぞれの対策の普及状況に依存し，単独のメールサーバのみの対応では，充分な効果は得られない．

受信側における対策は，受信側メールサーバのメール受信時，または利用者のメール取得時に行われる．受信側における対策では，利用者ごとの迷惑メールの分類規則に基づき，正当メールと迷惑メールとの分類処理を行う．代表的な分類処理としては，ホワイトリストやブラックリストによる送信者の受信許可や受信拒否の判別，判定規則に基づくメール本文の検査，統計的な解析手法の

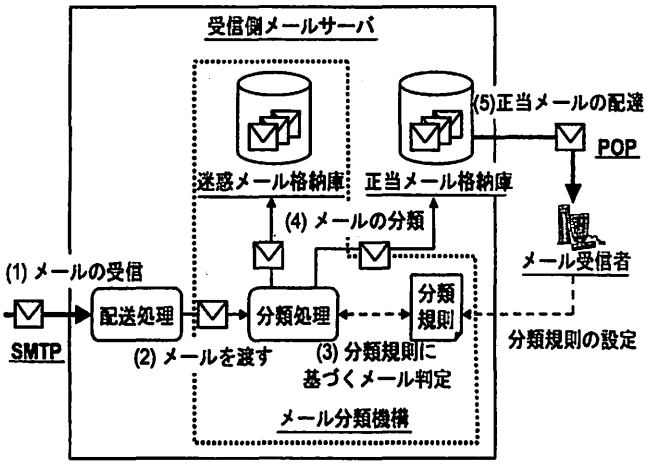


図1 メール分類機構の処理の流れ

利用[3],[4]，およびこれら分類処理の併用[5]がある．受信側における対策は，メール配送経路の終端での対策になるため，通信路上を流れる迷惑メールの量を削減する効果は少ない．しかし，受信側における対策は，効果が他のメールサーバの対応状況に依存せず，導入が容易なことが利点である．さらに，利用者のメール環境や好みに合わせた対策が可能である．このため，受信側における対策は広く普及している．

ここでは，既存の迷惑メール対策のうち，受信側メールサーバ上で行う対策であるメール分類機構について，利点と問題点を示す．さらに，その問題点への対処として，複数のメール分類機構を併用可能にする組み合わせ機構を提案し，その基本機構と期待される効果を説明する．なお，メール分類機構について，分類処理に関する研究や評価は，数多く行われているが，メール分類機構を効果的に併用可能にする研究はまだない．本機構は，メール分類機構を併用可能にすることで，迷惑メールの分類精度を維持しつつ，利用者の利便性が向上すると期待できる．

## 2 メール分類機構と問題点

### 2.1 処理の流れ

メール分類機構は，既存の迷惑メール対策の一つであり，受信側メールサーバ上で動作する．メール分類機構は，分類処理，分類規則，および迷惑メール格納庫から構成される．メール分類機構の処理の流れを図1に示し，以下に説明する．

(処理1) 配送処理は，利用者宛のメールをSMTP (Simple Mail Transfer Protocol) により受信する

(処理2) 分類処理は，配送処理からメールを受け取り，処理を開始する

(処理3) 分類処理は，利用者の設定した分類規則に基づき，受信メールが迷惑メールか否かを判定する

(処理4) 分類処理は，判定結果によりメールを分類する

(処理5) 利用者は，POP (Post Office Protocol) に代表される手続きにより，正当メール格納庫に格納されたメールのみを取得する

分類処理において迷惑メールと判定されたメールは，専用の迷惑メール格納庫に格納され，利用者に配達されない．これにより，メール分類機構は，利用者の迷惑メール受信を防止する．また，利用者による分類規則の設定を可能にすることで，利用者の好みに合わせたメールの判定を行う．

### 2.2 利点

メール分類機構は，受信側における対策であり，送信側における対策と比較して，以下の利点がある．

(利点1) 効果が他のメールサーバの対応状況に依存せず，導入が容易である

(利点2) 分類規則を利用者が設定できるため，利用者のメール環境や好みに合わせた対策ができる

さらに，メール分類機構は，受信側における対策のうち，利用者の計算機上で行う方式と比較して，以下の利点がある．

(利点3) メールサーバ上での利用のため，利用者の計算機環境を限定しない

(利点4) 利用者への配達前に分類処理を行うため，利用者のメール取得時の通信量を削減できる

### 2.3 問題点

迷惑メールの送信手口は，年々巧妙化している．さらに，迷惑メールの種類は，広告や宣伝を目的とした商用メールから，詐欺行為を目的としたフィッシングメールまで様々であり，「迷惑」の判断基準は，幅広いものになっている．このため，単純な分類規則や単一の分類処理によるメール分類機構では，充分な効果が得られなくなっている．そこで，迷惑メールの分類精度向上を図るために，複数の分類処理を併用し，さらにそれぞれの分類規則を高度化したメール分類機構が登場している．

しかし，こうしたメール分類機構について，分類処理の併用および分類規則の高度化による対処には，二つの問題がある．一つめは，分類処理の併用および分類規則の高度化による，メール分類機構の大規模化である．これは，運用するメールサーバの処理負荷の増加につながる．二つめは，設定すべき分類規則の高度化による，利用者の作業負担の増加である．これは，設定作業に不慣れな利用者にとって好ましくない．さらに，利用者ごとの分類規則の設定が可能であるものの，利用者同士が同様な分類規則を設定する場合，設定作業が冗長になる．

## 3 要求と対処

メール分類機構には，他の迷惑メール対策と比較して，利点があるものの，現状のメール環境での利用において，

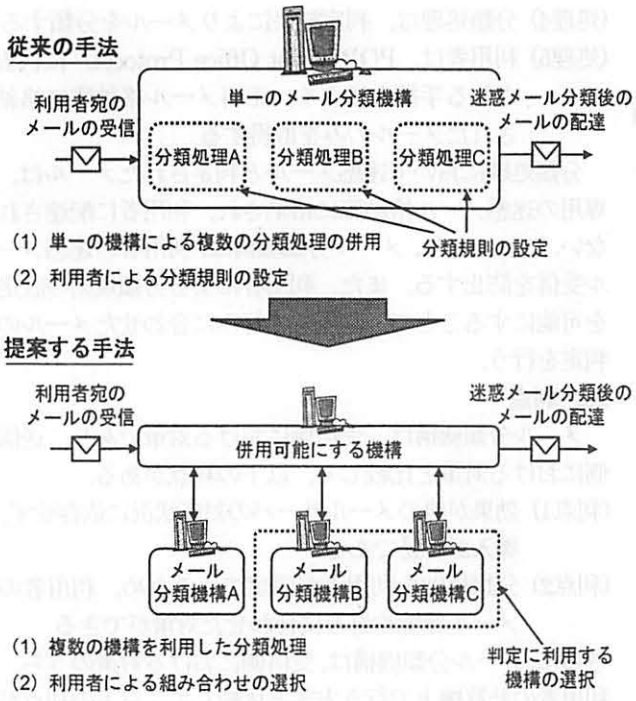


図2 基本的な考え方

先に述べた問題がある．そこで，メール分類機構の分類精度を維持しつつ，問題に対処する方法について述べる．

メール分類機構における問題から，以下の要求がある．

(要求1) 処理負荷の分散

(要求2) 利用者の負担の削減

それぞれの要求への対処について，基本的な考え方を図2に示し，以下に説明する．

(要求1)への対処として，メールの判定に複数のメール分類機構を利用する．例えば，分類処理A，B，Cを併用するメール分類機構と同等のメール判定を行う代わりに，それぞれの分類処理と同等なメール分類機構A，B，Cをメールの判定に利用する．

(要求2)への対処として，利用者は，分類処理に利用するメール分類機構の組み合わせのみを設定する．例えば，利用者は，分類処理B，Cを併用した方法によりメールの分類を行いたい場合，メール分類機構B，Cを組み合わせとして選択する．

## 4 組み合わせ機構

### 4.1 基本機構

前章で述べた対処を実現する機構として，複数のメール分類機構を併用可能にする機構である組み合わせ機構を提案する．組み合わせ機構の基本機構を図3に示し，以下に説明する．

組み合わせ機構は，メール分類機構と同様な利用形態であり，受信側メールサーバ上で動作する．組み合わせ機構は，分類処理，迷惑メール格納庫，および組み合わせ設定から構成される．さらに，分類処理において複数のメール分類機構をメールの判定に利用する．

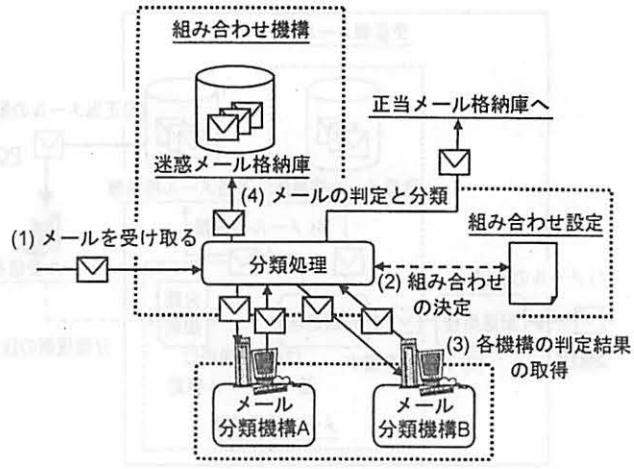


図3 基本機構

利用者は，分類処理に利用したいメール分類機構の組み合わせを設定することが可能である．組み合わせ機構は，利用者の設定したメール分類機構の組み合わせをもとに，メールの判定を行う．

組み合わせ機構における処理の流れを以下に示す．

(処理1) 利用者宛のメールを受け取る

(処理2) 組み合わせ設定を参照し，判定に利用するメール分類機構を決定する

(処理3) 判定対象のメールに対するメール分類機構の判定結果を取得する

(処理4) メール分類機構の判定結果をもとに，メールが迷惑メールか否かを判定し，メールを分類する

### 4.2 システム設計

#### 4.2.1 設計課題

組み合わせ機構の実現における設計課題を以下に示す．

(課題1) メール分類機構とのインタフェースの決定

(課題2) 各メール分類機構の判定結果の取り扱い

(課題3) 利用者間での情報の分離手法

(課題1)は，前節の基本機構で説明した処理の流れにおける(処理3)の具体的な処理手順を実現するためである．(課題2)は，処理の流れの(処理4)において，各メール分類機構から得た判定結果をどのようにメール判定時に取り扱うかを決定するためである．(課題3)は，一つのメール分類機構を複数の利用者で共有可能にするためである．

以降では，各課題の詳細とその対処について述べる．

#### 4.2.2 メール分類機構とのインタフェースの決定

組み合わせ機構では，メール分類機構をメールの判定時に利用する．ここで，メール分類機構について，SMTPを入力インタフェースとし，POPによるメールの取得の

有無を出力インタフェースとするメールの判定器と見なす．つまり，メール分類機構を利用するためのメールアカウントを用意し，そのメールアカウント宛に判定対象のメールを送信し，送信したメールをPOPにより取得できるか否かをそのメール分類機構の判定結果にする．これにより，組み合わせ機構は，メール分類機構を特殊化することなく利用可能にする．

ここで，メールを送信してから，メールを確認するまでの時間が短い場合，メールの確認がメール分類機構の処理終了前になり，正しい判定結果を確認することができない．一方で，メールを送信してから，メールを確認するまでの時間を充分に空けておけば，処理は終了し，正しい判定結果を確認できるものの，一通のメールに要する処理時間は長くなる．このため，メールの確認を行う契機が問題になる．

また，上記の方法を用いた場合，対象のメール分類機構が正常に動作しているか否かの確認ができない．このため，対象のメール分類機構が正常に動作しているのかを定期的に確認する別の手順が必要になる．

そこで，組み合わせ機構では，これらへの対処のため，メール分類機構の判定結果の取得を図4に示す流れで行う．この処理の流れを以下に説明する．

(処理1) メール分類機構に判定対象のメールを送信する
(処理2) POPによるメールの取得を一定の時間間隔で一定回数試行し，判定対象のメールを取得できれば，正当メールと判定し終了する
(処理3) (処理2)において，判定対象メールを取得できなかった場合，確実に正当メールに分類される確認用メールをメール分類機構に送信する
(処理4) POPによるメールの取得を一定の時間間隔で一定回数試行し，判定対象のメールと確認メールを取得できれば，正当メールと判定し終了する
(処理5) (処理4)において，確認メールのみを取得できた場合，迷惑メールと判定し終了する
(処理6) (処理4)において，確認メールを取得できなかった場合，判定結果を無効とし終了する

一定の時間間隔で一定回数メールの取得を試行する方法により，メール分類機構による分類処理の終了から，組み合わせ機構による判定結果の確認までの時間を最小限にできる．また，メールを取得できない場合に，確認用メールを送信することで，迷惑メールと判定されたことを確実に確認できる．さらに，確認用メールの送信は，メール分類機構の不具合の発生も検知できる．

#### 4.2.3 各メール分類機構の判定結果の取り扱い

組み合わせ機構は，各メール分類機構から得た判定結果をもとに，メールの判定を行う．

この判定方法は，以下の三つに方法に分類できる．

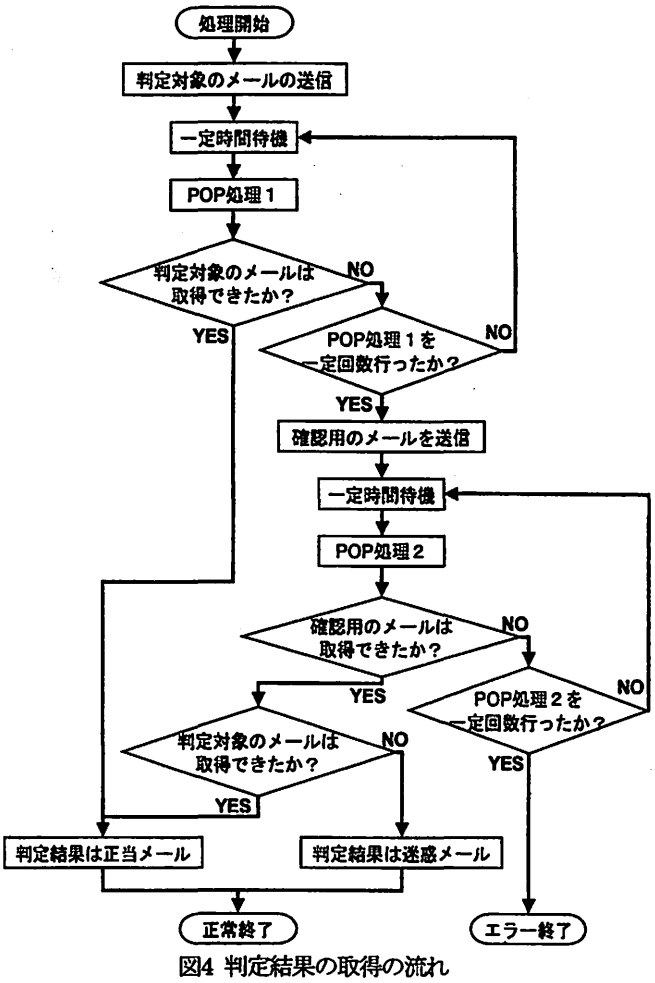

図4 判定結果の取得の流れ

(方法1) 判定結果に評価順序を設定
(方法2) 判定結果に重みを設定
(方法3) 特定のメール分類機構の判定結果を優先

(方法1)は，順番にメール分類機構の判定結果を確認し，一定数の正当メール判定，または迷惑メール判定が得られた時点で，判定処理を終了する．

(方法2)は，各メール分類機構の判定結果に重みを設定し，判定結果と重みに基づき，組み合わせ機構による判定を行う．重みの大きいメール分類機構の判定結果ほど，組み合わせ機構の判定結果に与える影響は大きくなる．

(方法3)は，特定のメール分類機構の判定結果により，組み合わせ機構の判定結果が決定される．例えば，ある優先するメール分類機構が正当メールと判定すれば，他のメール分類機構の判定結果に関わらず，組み合わせ機構では正当メールと判定する．

適切な判定方法は，利用するメール分類機構の種類に依存する．例えば，メール本文の検査や統計的な解析手法によるメール分類機構を併用しているならば，それぞれの機構の信頼性で重み付けした(方法2)が有効である．一方，ブラックリストを利用した受信拒否を行うメール分類機構を複数利用しているならば，それらの機構の判

定結果を優先する(方法3)が有効である．また，各機構の判定順序が重要になる場合は(方法1)で判定を行う必要がある．このため，各判定方法を必要に応じて適切に選択できることが望ましい．

#### 4.2.4 利用者間での情報の分離手法

組み合わせ機構では，複数の利用者が，同一のメール分類機構をそれぞれの組み合わせの一部として利用できる．このため，組み合わせ機構は，メール分類機構とやりとりするメールや判定結果が，どの利用者のものであるのかを識別できる必要がある．例えば，利用者X，Y，Zがメール分類機構Aを組み合わせに持つ場合，組み合わせ機構は，メール分類機構Aから取得した判定結果が，利用者X，Y，Zの誰のメールについての判定結果なのかを識別できる必要がある．

この対処として，メールのメッセージIDを利用する方法がある．メールのメッセージIDを利用することで，メールに変更を加えることなく，それぞれの利用者のメールを識別できる．しかし，メールのメッセージIDは，送信されるメールごとに一意であり，同一メールを複数の利用者が受信する場合，メッセージIDの重複が発生する．このため，この方法は好ましくない．

そこで，組み合わせ機構では，メール分類機構に判定対象のメールを送信する際に，メールのヘッダにどの利用者のどのメールなのかを一意に識別できる情報を追加する．メールを取得する際には，このヘッダ情報を解析することにより，メールの識別を行う．この方法は，本来判定されるメールに対して，変更を加えるものである．しかし，このヘッダ情報の追加による変更はわずかなものであり，メール分類機構の判定結果に影響しないと推察できる．

### 4.3 期待される効果

組み合わせ機構は，メール分類機構の入出力インタフェースに着目し，メールの送受信によりそれらの機構を利用する．これには，以下の三つの効果が期待される．

一つめとして，利用する個々のメール分類機構は，他のメール分類機構の状態とは独立して処理が可能なため，異なる計算機上に分散できる．これにより，メールサーバの処理負荷を分散できる．

二つめとして，メールの送受信の形態で利用するため，メール分類機構を組み合わせ機構に合わせて特殊化する必要がない．これにより，既存の多くのメール分類機構を組み合わせ機構の一部として，容易に利用できる．

三つめとして，メール分類機構を利用するためのメールアカウントを用意することで，分類処理に利用するメール分類機構の追加や削除が行える．これにより，分類処理の変更を容易に行える．

また，メール分類機構は広く普及しており，メール分類機構には数多くの種類が存在する．ここで，メール分類機構の問題で述べたように，分類手法の併用や分類規則の高度化は，分類精度を向上させるものの，メール分類機構を利用者にとって扱いにくいものにする．一方で，利用者による分類規則の設定作業が少ないメール分類機構は，充分な分類精度を得られないものの，利用が容易である．そこで，組み合わせ機構では，利用の容易なメール分類機構を複数併用することで，利用の容易さを維持しながら，分類精度の向上を図る．さらに，組み合わせ機構の利用方法をメール分類機構の組み合わせのみの設定にすることで，分類精度を維持しつつ，利用者の作業負担を削減できる．

## 5 おわりに

複数のメール分類機構を併用可能にする組み合わせ機構について述べた．具体的には，既存の迷惑メール対策の一つであるメール分類機構について，その利点と問題点を説明し，問題点への対処を述べた．さらに，その対処を実現する機構として，組み合わせ機構を提案し，その基本機構と期待される効果を説明した．組み合わせ機構は，既存の多くのメール分類機構を判定処理の一部として利用可能である．これにより，分類処理を複数の計算機上に分散でき，さらに分類処理の変更を容易に行える．また，組み合わせ機構では，メール分類機構の組み合わせのみを設定可能にすることで，利用者の作業負担が削減される．

今後の課題として，提案する機構の妥当性を示すために，複数のメール分類機構を組み合わせることの有効性を検証する．具体的には，メール分類機構について，単独での利用の場合，およびそれらを組み合わせた場合の効果の評価を行う．また，実用性を考慮し，利用インタフェースを簡潔にしつつ，利用者による設定の自由度を向上させる方法について検討を行う．

### 参考文献


[1] R. Gellens, J. Klensin, "Message Submission," RFC2476, 1998.

[2] M. Tran, G. Armitage, "Evaluating The Use of Spam-triggered TCP/IP Rate Control To Protect SMTP Servers," Proceedings of ATNAC-2004, pp329-335, Australia, 2004.

[3] P. Graham, "A Plan for Spam," http://www.paulgraham.com/spam.html, 2002.

[4] X. Carreras, L. Màrquez, "Boosting Trees for Anti-Spam Email Filtering," Proceedings of RANLP-2001, pp.58-64, Bulgaria, 2001.

[5] 岩永 学, 田端 利宏, 櫻井 幸一, "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," 情報処理学会論文誌, Vol.45, No.8, pp.1939-1947, 2004.