

Web を利用して電子メールの添付文書の転送量を軽減する機構の実現

矢野 大介† 乃村 能成‡ 谷口 秀夫‡

九州大学大学院システム情報科学府† 九州大学大学院システム情報科学研究院‡

インターネットを利用した代表的なサービスとして電子メールがあり、その利用は爆発的に増加している。このため、一人が受信するメールの大きさも増加しており、添付文書がメールの大きさの増加に拍車をかけている。一方、利用者が携帯型計算機で無線回線を使ってメールを受信する場合が増加している。しかし、無線回線を使った場合は、低速回線であるため受信に時間がかかる。そこで、本稿では、メールの大きさの増加に大きな影響を与えている添付文書に注目し、Web を利用して電子メールの添付文書の転送量を軽減する機構について述べる。また、本機構の実現例を紹介し、本機構を組み込むことによる処理量の増加は 1 秒以下であり、メール受信時間を約 16 分の 1 にできることを示す。

Implementation of a Mechanism to Reduce Attachment Size of E-mail Using Web

Daisuke YANO†, Yoshinari NOMURA‡ and Hideo TANIGUCHI‡

Graduate School of Information Science and Electrical Engineering, Kyushu University†
Faculty of Information Science and Electrical Engineering, Kyushu University‡

Internet e-mail facility is explosively getting more and more popular today. The size of e-mails has also been growing, and attachment of e-mail influences the size. Meanwhile, we often use a handy mobile computer to fetch e-mails from a mail server through a narrow wireless communication line. In case of using wireless telephone line, it takes a long time to fetch e-mails because the network bandwidth would be narrower than usual. In this paper, we propose a mechanism to reduce the attachment size of e-mails using Web and its applications. Also, we describe our experimental implementation of the mechanism and show the effect of it. We found that the overhead of our implementation is at most 1 seconds for each e-mail. We also show that our mechanism can make e-mail fetch time 1/16 shorter.

1 はじめに

技術の進化に伴い、インターネットの利用はますます増加し、人々の日常生活において重要な役割を果たしている。インターネットを利用した代表的なサービスとして、電子メール(以下 メール)や World Wide Web(以下 Web)がある。

メールは、遠隔の相手と非同期に通信を行う代表的な通信手段である。郵便に比べ、非常に速く届き、即時性がある。また、電話のように相手の都合に合わせる必要がないため、自分の好きな時間にメールを送信できる。これらの特長から、メールは、通知や連絡の便利な手段として多くの人々に利用されている。しかし、元来、メールには、メールを送信する際にさまざまな制約があった。例えば、メールではテキスト形式のデータしか送信できなかった。この制限を解消するために、1993 年、MIME(Multipurpose Internet Mail Extensions)^{[1][2]} 規格が公開された。

現在では、MIME の普及によって、画像や音声といったテキスト以外のデータをメールに添付して容易に送信できるようになっている。つまり、メールは一般的な電子文書やデータの受渡し手段として、利用者にとってさらに重要な役割を果たすようになっている。

このように、メールの利用者が増加し様々な形で利用されるようになり、また、MIME によって、メールに一般的な電子文書を添付した添付文書付きメールが増加している。このため、一人が受信するメール数が増加すると共に、一つのメールの大きさも増加している。

一方、個人のメールの利用形態が多様化している。例えば、オフィス内の居室での高速な専用線や自宅での ADSL 回線といった比較的高速な通信環境でメールを受信できる場合と、移動先での PHS や携帯電話の回線といった低速な通信環境でメール

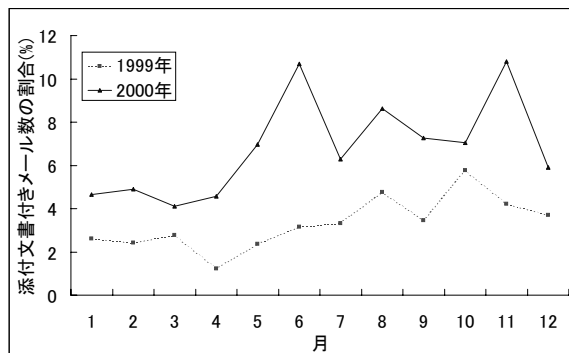


図 1 添付文書付きメール数の割合.

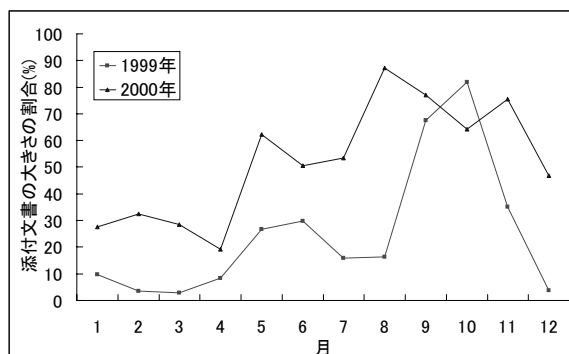


図 2 添付文書の大きさの割合.

を受信しなければならない場合がある。このため、オフィス内の居室にある計算機でメールを受信する場合は短時間で受信できるが、移動先での PHS や携帯電話などの無線回線を使ってメールを受信する場合は、受信時間が長時間になってしまい、通信の利用価格も高くなる。さらに、多くの場合、居室にある計算機に比べ移動先で使用する携帯型計算機は性能が低いため、移動先ではメールの受信にさらに時間を費やしてしまう。

つまり、メールの数、大きさが増加しており、また、低速回線でのメールの受信には時間がかかる。このため、メールの転送量を小さくすることが必要である。

そこで、本稿では、メールの大きさの増加に大きな影響を与えている添付文書に注目し、Web を利用してメールの添付文書の転送量を軽減する機構について述べる。具体的には、メールの添付文書量が増加し、メールの大きさに影響を与えている状況を説明し、メールの大きさを小さくする手法を述べる。さらに、Web を利用してメールの大きさを小さくする手法である転送量軽減機構を示し、その基本機構について述べる。最後に、提案した機構の実現例を紹介し、評価により機構の有効性を示す。

表 1 添付文書付きメールの総数.

1999 年	212 通
2000 年	454 通

表 2 添付文書の総量.

1999 年	17496 KB
2000 年	45615 KB

2 メールの添付文書の増加と対処

2.1 添付文書の量

メールの添付文書の増加状況を分析するため、著者の一人が受信したメールについて分析を行った。1999 年と 2000 年について受信したメールのうち、添付文書付きメールに関する分析結果を図 1 と図 2 に示す。

図 1 は、全受信メール数に対する添付文書付きメール数の割合を月毎に示したものである。全ての月で 2000 年の方が受信メールが添付文書を含むメールである割合が高くなっている。また、表 1 に示すように、添付文書付きメール数は、2000 年の方が総数が多く、全体で約 2.14 倍に増加しており、添付文書付きメールの受信数は、増加する傾向にあるといえる。図 2 は、全受信メールの大きさの総和に占める添付文書の大きさの割合を月毎に示したものである。10 月を除く全ての月で、2000 年の方が添付文書の大きさの占める割合が多くなっている。月によっては、約 90% を記録するものもあり、添付文書がメール量に与える影響は非常に大きいといえる。また、表 2 に示すように、受信した添付文書付きメールに含まれる添付文書の大きさは、2000 年の方が総量が多く、全体で約 2.61 倍に増加しており、添付文書の大きさが増加傾向にあるといえる。

これらの分析から、添付文書付きメール数の割合は約 10% 以下であるにもかかわらず、添付文書がメールの受信量に占める割合は、1999 年では約 40%、2000 年では実に約 60% であることがわかった。したがって、添付文書は、メールの総量を約 2 倍にしているといえる。さらに、分析に利用した個々の添付文書付きメールから、添付文書部分を取り除くと、大きさが平均で約 44 分の 1、最小では約 2400 分の 1 になることがわかった。

つまり、受信する添付文書付きメールの数は増加傾向にあり、受信する添付文書付きメールの大きさも増加傾向にある。さらに、添付文書はメールの総量を非常に大きくしている。

上記は、大学教官の一例であるが、多くのメール利用者において、同様の傾向にあると推察できる。

表 3 添付文書を本文から分離する方式の比較.

方式	概要	長所	短所
Web の利用	添付文書を Web サーバ上に分離して置き、その URL のみをメール本文に追加する.	(1) 既存のメールソフト、Web ブラウザを利用できる. (2) 現状との親和性が高い.	(1) ファイルとして格納する記憶域や Web サーバが必要になる.
external-body の利用	MIME の external-body サブタイプを利用する.	(1) 参照情報のみを持つ MIME パートを受信者側のメールソフトが自由に加工できる利点がある.	(1) ほとんどのメールソフトが対応していない. (2) データを格納する記憶域や Web サーバ、FTP サーバが必要になる.

現在、これらの傾向に加えて、携帯型計算機を利用して受信することが多くなっている。例えば、携帯電話を利用した場合、9600bps の低速回線を利用するため、受信時間が長くなり、通信の利用価格が高くなるため好ましくない。今後のさらなる携帯型計算機の普及により、この問題は大きなものになる。

以上のことから、受信するメールの添付文書量を軽減することは非常に重要である。

2.2 対処

メールの添付文書の大きさを小さくする方式としては、大きく 2 つの方式が考えられる。一つは添付文書を圧縮して送信し、受信時に伸張する方式である。もう一つは添付文書を本文から分離する方式である。さらに、添付文書を本文から分離する方式として、Web を利用する方式^[3]と MIME の external-body サブタイプを利用する方式の 2 つが考えられる。ここでは、添付文書を本文から切り離す 2 つの方式について、その特徴を表 3 に示し、以下に説明する。

Web を利用する方式は、添付文書をメールから分離して格納し、Web 操作できるようにして、メールには添付文書の格納位置を URL (Uniform Resource Locator) として追加する方式である。この場合、添付文書を格納する新しい記憶域や Web サーバが必要になる。しかし、受信側では、必ず添付文書の大きさ分だけメールの大きさを軽減できる。また、受信には既存のメールソフトを使うことができる。さらに現在の PC には Web ブラウザが標準的にインストールされていることから、Web ブラウザを使うことで、受信側に特別なソフトウェアが必要ないという利点がある。つまり、この方式は、既存のメール機能と Web の機能を組み合わせた方式であるため、現状との親和性が高いといえる。

external-body を利用する方式は、MIME の message メディアタイプの external-body サブタイプを利用して添付文書を分離する方式である。external-body とは、参照情報のみを持つ特殊な形

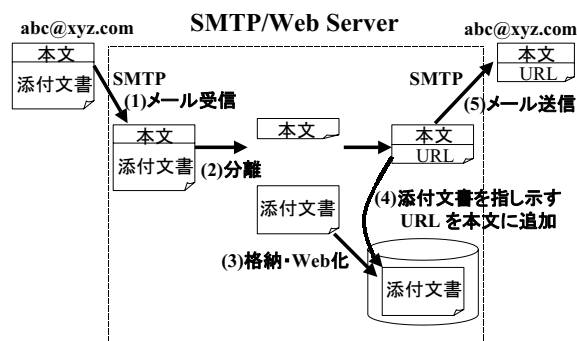


図 3 基本機構.

式の MIME パートである。external-body を利用する場合、Web などで公開されているデータへの参照情報を MIME パートの 1 つとして送信できるので、受信者側のメールソフトが自由に加工できる利点がある。例えば、添付文書のメディアタイプが text ならそのままメールソフト内で表示し、image ならば Web で参照させるといようなことをメールソフト側で操作できる。しかし、external-body をサポートしているメールソフトが非常に少ないという問題がある。また、データを格納する記憶域や Web サーバ、あるいは FTP サーバが必要になる。

以上の比較により、メール文書量を非常に小さくでき、かつ受送信側の利用インタフェースを特殊化せず、現状との親和性が高い Web 利用方式が優れているといえる。

3 基本機構

Web を利用してメールの添付文書の転送量を軽減する機構では、送信者と受信者の利用インタフェースを特殊化しないことを前提とする。これにより、既存のメール機能と Web の機能のみで本機構を導入することができる。本機構の基本機構を、図 3 に示す。基本機構では、以下の処理を行う。

- (1) 添付文書付きメールを受信する
- (2) メール本文と添付文書を分離する
- (3) 分離した添付文書を Web 公開用ディレクトリに格納し、Web サーバから参照できるようにする

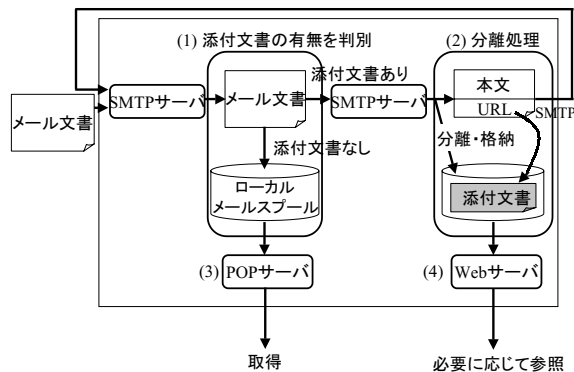


図4 実装内容.

(4) 格納した添付文書を指し示す URL をメール本文の末尾に追加する

(5) URL を追加したメールを元の宛先に送信する
図3に示すように、本機構では、添付文書付きメールを入力とし、メールヘッダ、本文、格納した添付文書を指し示す URL を1つにして出力する。このように添付文書の転送量を軽減する機構をメール変換機構とすることで、インターネットメールシステムにおける本機構の配置に関し、高い自由度を持たせることができる。

4 実装と評価

4.1 実装内容

本機構の配置に関して、本機能を導入するかどうかを受信者単位で決定できる利点があることから、受信側の SMTP サーバで実現した。実装内容を図4に示す。図4は、(1)添付文書の有無を判別する処理、(2)メール本文と添付文書を分離する処理、(3)受信者がローカルマシンにメールを取得するための POP サーバ、(4)格納した添付文書を受信者の必要に応じて閲覧させる Web サーバからなる。これらの処理を受信側のメールサーバ上で稼働させ、それぞれの処理を連携させることによりシステムを構成する。以下に、本システムの動作を説明する。

(1) 受信側 SMTP サーバはメールを受け取り、受信者のホームディレクトリの .forward ファイルを読み、procmail を起動する。procmail とは、メールに関する処理を行う既存のプログラムである。procmail は、添付文書の有無を判別し、

(a) 添付文書がない場合、そのままローカルスプールに格納する。

(b) 添付文書がある場合、同一サーバ内の特定のユーザ (separate ユーザとする) にメールを転送する。

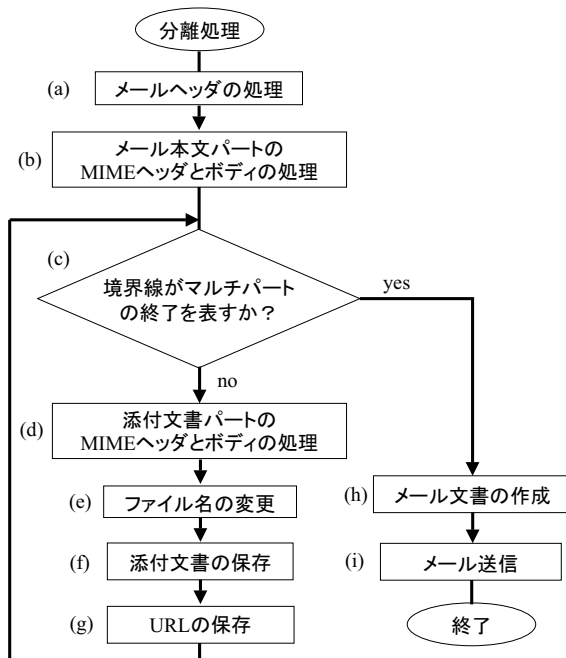


図5 分離処理.

(2) 転送された添付文書付きメールを受け取った SMTP サーバは、separate ユーザのホームディレクトリの .forward ファイルを読み、分離処理を起動する。分離処理では、添付文書をメール本体から分離して格納し、Web サーバから参照できるようにする。さらに、添付文書を指し示す URL をメール本文に追加し、元の受信者宛に送信する。

(3) 受信者は、POP サーバにアクセスすることによって添付文書なしのメールを取得する。

(4) 受信者は、必要に応じて、メールに記載された URL に基づき添付文書を Web ブラウザで参照する。

さらに、図4の(2)で起動される分離処理の流れを図5に示し、以下に詳細な実装内容について説明する。なお、(2)で受け取る添付文書付きのメールは、ヘッダ部分とボディ部分からなり、さらにボディ部分は複数の MIME パートで構成されている。ヘッダ部分には、ボディ部分の属性を示す Content-Type ヘッダがあり、この場合、その値は、マルチパート・メディアタイプとなる。また、この値のパラメータとして、各 MIME パートの境界を表す識別子 (boundary) が示されている。さらに、1つの MIME パートは、そのパートの属性を表す MIME ヘッダ部分と実際の内容であるボディ部分から構成されている。

(a) まず、添付文書付きのメールを受け取り、メー

ルヘッダ部分を処理する．具体的には，boundary を取り出し，その他のヘッダを保持しておく．

- (b) 次に，メールの最初の MIME パートを処理する．この部分は，添付文書処理後の新しい本文として再構成されるので，Content-Type ヘッダとボディ部分を再構成のために保持する．
- (c) 2 つ目以降の部分は添付文書部分である．ここで，MIME パートの境界がマルチパートの終了を表す識別子かどうかを調べる．
- (d) (c) において，境界がマルチパートの終了を表す識別子ではないとき，添付文書部分において，まず MIME ヘッダ部分を処理する．具体的には，ファイルのエンコード形式やファイル名等の値を取得する．さらにボディ部分を保持する．
- (e) 取得したファイル名から拡張子を抽出し，「時刻+プロセス ID+拡張子」という形式の一意なファイル名に変更する．
- (f) Web 公開ディレクトリ以下にある各受信者専用のディレクトリ以下に，(e) で得たファイル名で，(d) で保持しておいた添付文書の内容を保存する．この際，(d) で取得したエンコード形式に従って，デコードして書き込む．
- (g) 受信者がメール取得時にこのファイルを参照するための URL を保持しておく．ここで再び (c) に戻り，MIME パートの境界がマルチパートの終了を表す識別子かどうかを調べる．(c)～(g) を境界がマルチパートの終了を表す識別子となるまで繰り返す．つまり，添付文書がある限りこの処理を繰り返す．
- (h) (c) において，境界がマルチパートの終了を表す識別子であるとき，(a) で保持しておいたメールヘッダ，(b) で保持しておいた最初の MIME パートの Content-Type とボディ，そして (g) で保持しておいた添付文書を指し示す URL から新たなメール文書を作成する．
- (i) (h) で作成したメール文書を作成した SMTP クライアントに渡す．この SMTP クライアントは，メールヘッダを渡されたメールのヘッダに置き換えて送信する仕様であるため，(1) で受信したメールのヘッダと同じにすることができる．

以上の処理から，添付文書は Web から参照可能な形で分離・格納され，受信者のメールプールには，

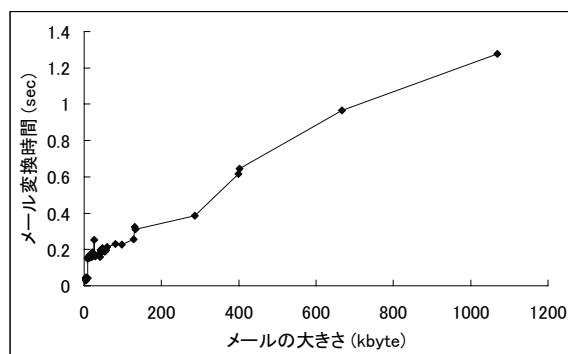


図 6 メール変換時間.

添付文書なしのメールのみが格納される．また，内容の異なる添付文書のファイル名が同じ場合でも，一意に参照できるように，格納する添付文書のファイル名は必ず一意である必要がある．そのため，(e) で，必ず一意になるファイル名に付け直している．

なお，分離処理を行うプログラムには，文字列処理を効率的に記述するために，インタプリタ言語である Ruby を用いた．

4.2 評価と考察

図 2 や図 1 の状態であった著者が，本機構を運用し，2001 年 7 月における 1 ヶ月間のデータを分析した．この結果から，本機構の有効性を評価する．実験には，SUN Microsystems 社の WorkStation, UltraSPARC 200MHz, メモリ 128MB の計算機を用いた．OS は，Solaris である．

2001 年 7 月のメール総受信数は 639 通で，そのうち添付文書付きのメールは 67 通であった．また，分離処理プログラムによって正常に処理されたメールは，67 通のうち 52 通であった．残りの 15 通は，MIME に違反している添付文書，本文が空のメール，MIME ヘッダの解析ミスなどの原因により，エラーとなったメールである．したがって，今回，67 通の添付文書付きメールのうち，これら 15 通のエラーとなったメールを除いた 52 通の添付文書付きメールについて分析を行った．

図 6 は，メールの大きさとメール変換時間（分離処理時間）の関係を示している．図 6 から，大半のメール変換時間は約 0.2 秒程度であり，非常に短い時間であるといえる．また，1MB を越えるメールについては，分離処理に 1 秒以上かかることがわかる．しかし，2.1 節で分析に利用したメールにおいて，1MB を大きく越えるものは全体のわずか 1%程度であり，ここで利用した 52 通の添付文書付きメールにおいても，1MB を越えるものは 1%程度

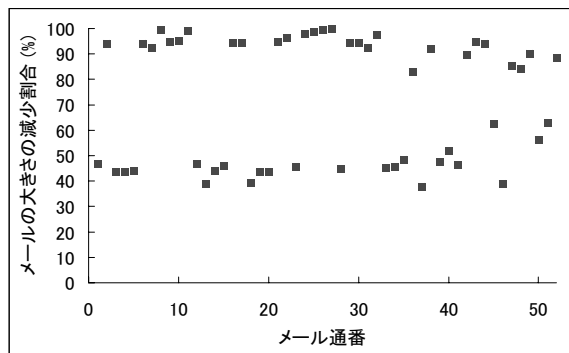


図 7 受信メールの大きさの減少割合.

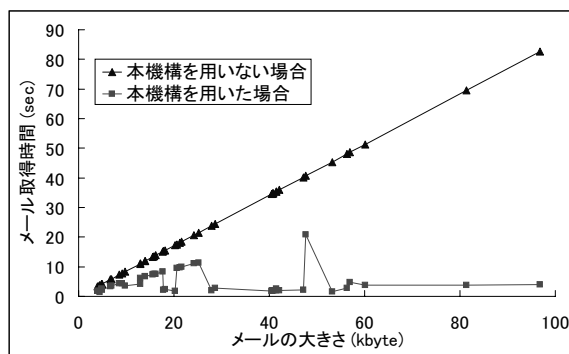


図 8 POP によるメール取得時間.

であった。したがって、大半の場合、1 秒以下で処理でき、分離処理によるメール配送のオーバーヘッドは小さいといえる。

図 7 は、メールの大きさの減少の割合をメール毎に示したものである。図 7 から、添付文書を取り除くことにより、最大で約 99%、最低でも約 40% 大きさが縮小されていることがわかる。90% 以上大きさが減少したメールには、数 KB の本文に対して添付文書が 1MB 以上のものもあり、添付文書がメールの大きさに与える影響は大きいことが確認できる。

次に、受信者がメールを POP サーバ経由で取得する時間を示す。9600bps の低速回線でメールを取得する際、本機構を用いない場合と用いた場合について、メール取得時間を図 8 に示す。図 8 は、分析に利用した 52 通のメールについて、メールの大きさが 100KB 以下のメール 44 通の取得時間を示している。本機構を用いない場合のメール取得時間は、受信するメールの大きさに比例して増加する。一方、本機構を用いる場合のメール取得時間は、メール大きさに比例せず、ほぼ一定で、大半は約 10 秒以下になる。図 8 には示していないが、100KB 以上のメールについても同様の結果を得た。10 秒から 20 秒かかるメールが 2,3 通あるが、これらのメールは、本文テキストだけで数十 KB あるものである。

本機構では添付文書のみを分離することで転送量の軽減を図っているため、本文テキストが数十 KB 以上あり増加する傾向にある場合は、それに比例してメール取得時間も増加する。しかし、分析に利用した 52 通のメールにおいて、本文テキストが 10KB を越えるメールは 1 割以下であり、本文テキストが数十 KB を越えることはほとんどないと考えられる。つまり、本機構により、メールの取得時間を非常に短縮できるといえる。具体的には、メール取得時間を平均約 16 分の 1 にできた。

5 おわりに

本稿では、メールの添付文書の転送量を軽減する機構について述べた。具体的には、メールに含まれる添付文書の量が増加している状況を説明し、添付文書の大きさを小さくする手法を述べた。さらに、Web を利用してメールの添付文書の転送量を軽減する機構を提案し、最後に実現例を紹介した。その結果、本機構を組み込むことによる処理量の増加は 1 秒以下であり、メール取得時間を平均約 16 分の 1 にできることを示した。なお、本機構は、メールから添付文書を分離して送信するため、添付文書のウイルス問題の解決も期待できる。

今後の課題として、変換ミスに対する分離処理プログラムの修正、Web 化した添付文書に対するアクセス保護機能および具体的なウイルス対策方法の検討がある。また、本稿では、本機能を導入するか否かを受信者単位で決定できる利点があることから、受信側のみでの実現例を紹介したが、本機構の実現は、SMTP 配送系のどの位置にでも置けることが望ましい。したがって、本機構の利用形態に関する考察も必要である。

参考文献

- [1] N.Freed & N.Borenstein, "Multipurpose Internet Mail Extensions(MIME) Part One: Format of Internet Message Bodies," RFC1521, September 1993.
- [2] K.Moore, "Multipurpose Internet Mail Extensions(MIME) Part Two: Message Header Extensions for Non-ASCII Text," RFC1522, September 1993.
- [3] 谷口秀夫, 乃村能成, 矢野大介, "Web を利用した電子メールの転送サイズ縮小機構の提案," 九州大学大学院システム情報科学紀要, 第 6 巻, 第 2 号, 2001.