ビジネス開発者必須講座

P2P/ブロードバンド時代の新・TCP/IP入門

村上 健一郎　法政大学ビジネススクール イノベーション・マネジメント研究科 教授

第7回　なぜ宣伝どおりのスピードが出ない？　－その2－

8月号では、最大ウィンドウサイズによる通信速度の上限に言及し、巧妙な輻輳ウィンドウ制御によるネットワーク負荷の最適化について説明しました。今回は、MTUと呼ばれるIPパケットの最大長が、どのような影響を及ぼすかについて考えていくことにします。

[Q1]

MTUを調整すると速度が向上すると聞いたことがあるのですが、MTUとは一体何ですか？　また、どのような理由でこれが転送速度に影響を与えるのでしょうか？

[A1]

MTU＝最大IPパケット長

MTU（Maximum Transmission Unit）とは、最大のデータグラム長、つまり、最大IPパケット長です。これが大きければ、1つのパケットで送れるデータ量も増えます。同じデータ量ならば少ないパケット数で転送できるので、パケットごとの処理のオーバーヘッド（付加的処理）が減り、転送速度が向上します。

7月号（第5回）で、エンキャプシュレーション（カプセル化）については触れましたが、ネットワーク層のIPパケット（IPデータグラム）は、その下位にあるリンクレイヤーのパケット、たとえばイーサネットパケットのデータ部分に入れられて転送されます。これを図1に示します。

一般的に使用されているイーサネットのデータ部の最大長は、1500バイトです。ですから、MTUの上限も1500バイトになります。ここで上限と言ったのは、1500バイト以下にMTUを設定して使用してもいいからです。しかし、効率の面から、通常は、MTUを利用できる最大の長さに設定して使用します。

イーサネットを使用してTCP/IPで通信を行うコンピュータでは、MTUからIPヘッダー長である20バイトを引いた値以上の長さのデータを転送する場合には、それを複数のIPパケットに分割して送らなければなりません。それらのパケット長もMTU以下でなければならないのは言うまでもありません。なお、IPプロトコルで規定されているMTUの最大値は65,535、つまり64Kバイトです。

表1に、さまざまなリンクレイヤーの伝送媒体と最大のMTUについてまとめました。

最大データ長はどのようにして決まる？

すでに述べたように、MTUは使用するリンクレイヤーのパケットで許される最大データ長によって制約を受けてしまいます。

それでは、どのような理由でデータ長の制限が決められたのでしょうか？　ここでは、主な2つの理由について説明しましょう。

(1) 転送の待ち時間が長くなるから

イーサネットや無線LANなどでは、CSMA方式が用いられてきました。CSMAとは、Carrier Sense Multiple Accessの略です。これは、1つの物理レイヤーを多数のコンピュータで共有するときに使用される転送方法です。たとえば、オリジナルのイーサネットの場合は同軸ケーブルであり、無線LANの場合には、特定周波数の電波です。

CSMAでは、データを送りたいコンピュータは、まず、他のコンピュータによって転送が行われているかどうかを

チェックします。転送中であれば、それが完了するまで待ちます。転送中でなければ、データの転送を開始します。このとき、転送できるパケット長が長ければ長いほど、同じデータ量でも少ないパケットで転送できます。

しかし、いったん、あるコンピュータが転送を開始すると、それが終了するまで他のコンピュータはデータを転送することができませんから、データ長が大きくなればなるほど待ち時間が長くなるという問題が発生します。

10Mbpsの速度の場合、1500バイトのデータを転送するために1ミリ秒以上かかります。100台のコンピュータがあったとすると、順番にパケットを転送するだけでも順番が回ってくるのに100ミリ秒(0.1秒)以上かかります。イーサネットでは、ネットワークが空いたと思って複数のコンピュータが同時にデータの送信を開始することもあります。

前出の数字には、この衝突によるパケットの損傷と再送を想定していません。それを想定すると1桁以上長い待ち時間が発生すると思うべきです。ですから、むやみにデータ長を大きくすることができません。

(2)ノイズによるパケット損傷の問題

パケットのデータ長が長ければ、同じデータ量を少ないパケットで送ることができるのですが、逆に問題の発生する確率が高まります。それは、ノイズによるパケットの損傷です。たとえば、ランダムにノイズが発生すると仮定しましょう。そうすると、パケットの長さに比例してノイズによって損傷を受ける確率も高まります。その場合、再転送することになりますが、長ければ長いほど再転送による無駄が大きくなります。

以上のように、転送待ち時間やノイズの問題を考慮してリンクレイヤーのパケット(フレーム)サイズが決定されています。このサイズの制限により、MTUも制約を受けるというわけです。なお、インターネットでは、転送元のコンピュータから転送先のコンピュータまで、いくつものルーターを経てIPパケットが転送されます。

ルーター間は伝送媒体(イーサネットや専用線など)を介して接続されており(図2)、それぞれ異なったMTUが使用されている場合があります。ですから、自分のコンピュータが使用している伝送媒体だけを考慮してMTUを設定しても、必ずしもそれが最適なものとはなりません。

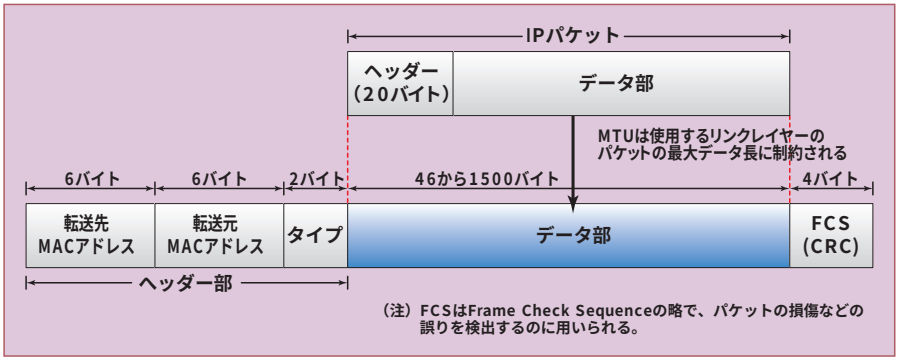

図1　MTUとイーサネットパケットのデータ部との関係

リンクレイヤーの伝送媒体やプロトコルの名前	設定できる最大のMTU	備考
イーサネット	1500バイト	IEEE 802.2の規格LLC/SNAPを使用する場合には1492バイト。
100Mbpsイーサネット	1500バイト	ー
1Gbpsイーサネット (ギガビットイーサネット)	1500バイト ジャンボフレームは通常フレーム6個分の9000バイト	ー
無線LAN	2296バイト	2312バイトのデータ長があるが、8バイトのWEP用のオーバーヘッドと8バイトのLLC/SNAPを使用するために2296バイトとなる。
ADSLのPPPoEoA (PPP over Ethernet over ATM)	1454バイト	ATMのセル化のため、1448バイトのほうが効率が良い。以下、順次48バイトを減算した1400、1352などでも効率が良くなる。
ADSLのPPPoE (PPP over Ethernet)	1492バイト	PPPヘッダー8バイトがイーサネットのデータ部の先頭に入るため。

LLC：Logical Link Control、論理リンク制御
SNAP：Sub-Network Access Protocol、サブネットワーク・アクセス・プロトコル
WEP：Wireless Equivalent Privacy、無線LANの暗号化方式。IEEE 802.11標準

表1　さまざまな伝送媒体と最大のMTU

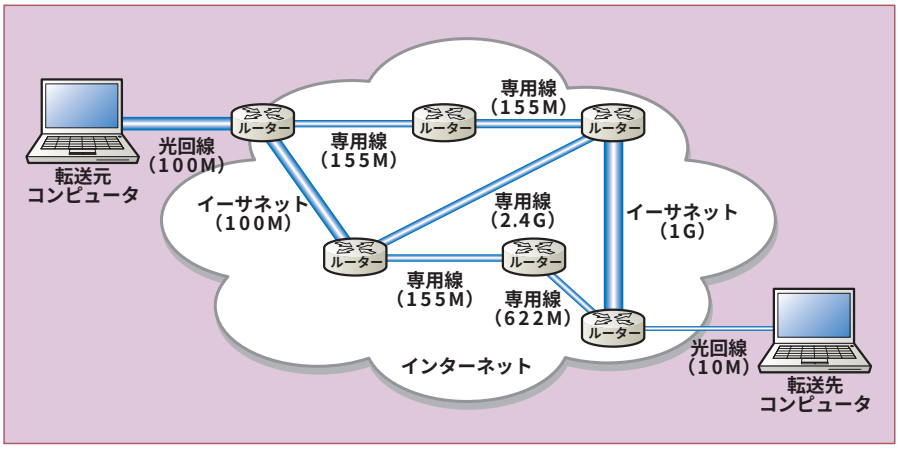

図2　コンピュータ間には多数の伝送媒体が介在する

これについては、後述します。

細分化はどのようにして行うのか？

コンピュータやルーターが、隣接するルーターにIPパケットを転送するときには、それを送り出すイーサネットや専用回線などのインターフェイスのMTUをチェックします。もし、IPパケットの長さがそのMTUを超えていた場合には、細分化して各IPパケットがMTU以下の長さになるように分割し、それらを転送します。これを細分化（フラグメンテーション）と呼びます。パケットがいったん細分化されると、転送先のコンピュータにおいてのみ元のデータに組み立てられます。転送中に再組み立てが行われることはありません。

細分化の様子を図3に示します。IPパケットの長さは、ヘッダー部のパケット長フィールド（7月号の「図2 IPヘッダー」を参照）に入っています。これを検査すれば、MTUを超えているかどうかがわかります。超えている場合には、IPパケットを分割します。

細分化の副作用

細分化によってIPパケットは分割されるため、たとえ転送元のコンピュータと転送先のコンピュータの間にIPパケットよりも小さなMTUの伝送媒体があっても、目的のコンピュータまで転送できるようになります。しかし、細分化には通信のオーバーヘッドを増大させるいくつかの副作用があります。

1つは、分割によって処理のオーバーヘッドが増大するということです。分割されたIPパケットは転送先のコンピュータまで送られ、そこで元のIPパケットに再構成されます。このとき、複数のIPパケット部分をつなぎ合わせるためにメモリー間のコピーが発生したり、ヘッダー処理の時間が分割されたデータの数に比例して増大したりして、処理が遅くなるのです。

2つ目は、すべての細分化されたデータが揃うまで、到着したIPパケットをすべてメモリー上に保存しておかなければならないという問題です。

細分化されたIPパケットのうち1つが途中で破棄された場合、確認応答が返ってこないことに転送元のコンピュータが気づいて元のIPパケット全体を再転送するまで、分割されたデータは保存されることになります。これは、メモリーの利用効率から言っても望ましいことではありません。また、分割されたデータが1つでも欠けると、元のIPパケット全体が再転送されることになるため、ネットワークの帯域が無駄に使用されることにもなります。

3つ目は、細分化された一連のIPパケットが一度に転送されると、どれかが失われる確率が高まるという問題です。たとえば、もともと8Kバイトだったパケットを1500バイトごとに分割すると想定すると、6個に細分化されることになります。これらがまとまってルーターに到着した場合、待ち行列がしきい値以上になるかもしれません。そのとき、前号で説明したTD（Tail Drop）によるパケットの破棄が採用されているとすれば、後ろの一部のパケットが破棄される可能性があります。しかも、再転送するごとに同じことが繰り返される可能性もあります。前のフラグメントは必ず転送されるのに、常に後ろのフラグメントが捨てられるために、通信不能となる恐れもあります。

Path MTUとその検出方法

細分化による副作用を防止するためには、分割されたデータが発生しないようにしておけばよいということになります。つまり、転送元から転送先までの間にあるすべての回線のMTUを知り、その中で最も小さいものをMTUにすればよいということになります（図4）。このMTUをPath MTUと呼びます。しかし、インターネットは多数の回線とルーターから構成されるため、転送元から転送先のコンピュータまでにあるすべての回線や通信媒体のMTUをあらかじめ知るのは不可能です。

そこで、これを転送元のコンピュータが自動的に検出する方法、つまり「Path MTU発見プロトコル」が使われるようになりました。

この方法は、IPヘッダーのフラグフィールド（7月号の「図2 IPヘッダー」を参照）のDF（Don't Fragment）フラグを利用します。これが「1」の場合は、分割禁止を示します。ルーターは、MTUによる制限でパケットを細分化しなければならな

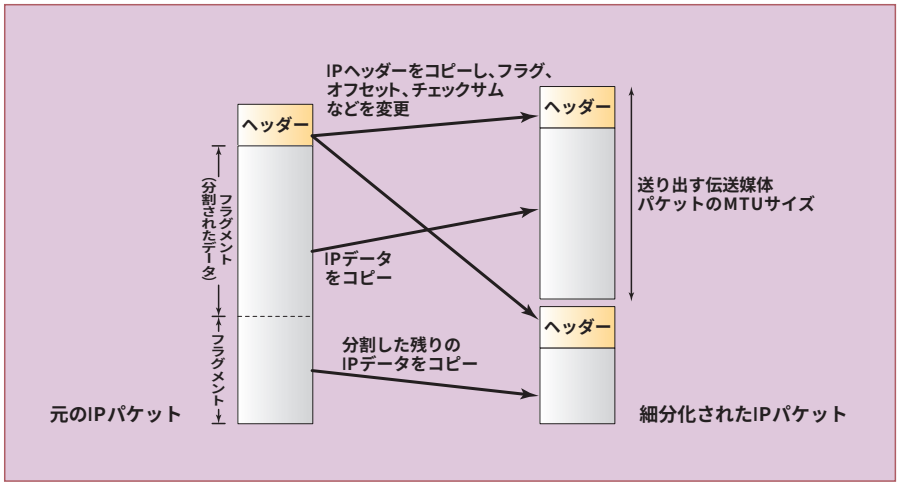

図3 IPパケットの細分化

いにもかかわらず、そのIPヘッダーにDFフラグが立ててあった場合には、送信元にそのエラーを通知することになっています。

通知には、ICMP（Internet Control Message Protocol）のパケットが使われ、“fragmentation needed but don't-fragment bit set”というエラーが返されます。そこで、この仕組みを応用してPath MTUを自動検出します。

送信元のコンピュータは、IPヘッダーのDFフラグでフラグメント（細分化）禁止を指定して長いパケットを送ります。ところが、経路上のどこかのルーターでIPパケット長がMTU以上になった場合、ICMPでエラーは返ってきますが、それ以下であればICMPは返ってきません。長いIPパケット長から始めてICMPが返ってこないようになるまで、次第にパケット長を短くしていけば、経路中の最小のMTUがわかるという仕組みです（図4）。この際、どのパケットにもDFビットを付けてフラグメント禁止にしておきます。そうすると、通信途中で経路が変わっても継続してPath MTUを見つけながら通信ができるというわけです。

ブラックホールに似た経路上の問題

一見うまくいくように見えるPath MTU発見の方法ですが、ルーターやファイアーウォールの設定ミスで、かえって通信ができないという問題が発生します。それはICMPのパケットを、むやみにフィルタリングして捨てるように設定されているファイアーウォールやルーターが存在するからです。

転送元ではICMPが返されないので、Path MTU以下だと信じて、DFビット付きのパケットを転送し続けます。ところが、それらは細分化が発生したルーターなどで捨てられ、エラーを示すICMPが実際には返されているのです。これに気づかない転送元のコンピュータは、これを繰り返し、結局、通信不能の状態になります。

悪いことに、Path MTUの発見ができる経路とできない経路が混在するために、どのような問題が発生しているのかを特定するのが困難になります。たとえば、あるウェブは見ることができるのに、別のウェブが見られなかったりします。たまたま、送るデータがPath MTUのサイズ以下であれば通信がうまくいくのですが、Path MTU以上の量のデータを送ろうとした途端にうまくいかなくなります。このため、ウェブの画面では、Path MTU以上のデータ量が大きい画像だけがエラーで表示されないにもかかわらず、データ量が小さいテキストの部分だけは問題なく表示されるという症状も発生します。メールの場合には、大きなメールは届かないのに、小さいメールだけは届くという現象になります。

本来は、ファイアーウォールやルーターの管理者が、このような仕組みを考慮してフィルターを設定しておくべきものですが、残念ながら、そこまでできる管理者は少ないというのが現状です。

そこで、このような問題の対策として、最初からMTUのサイズを小さく見積もっておいて静的に設定しておき、Path MTU発見プロトコルを使用しない方法も採られています。また、ソフトウェアによっては、問題箇所を自動的に検出し、その相手だけ、Path MTU発見プロトコルを使用しないようにするものもあります。

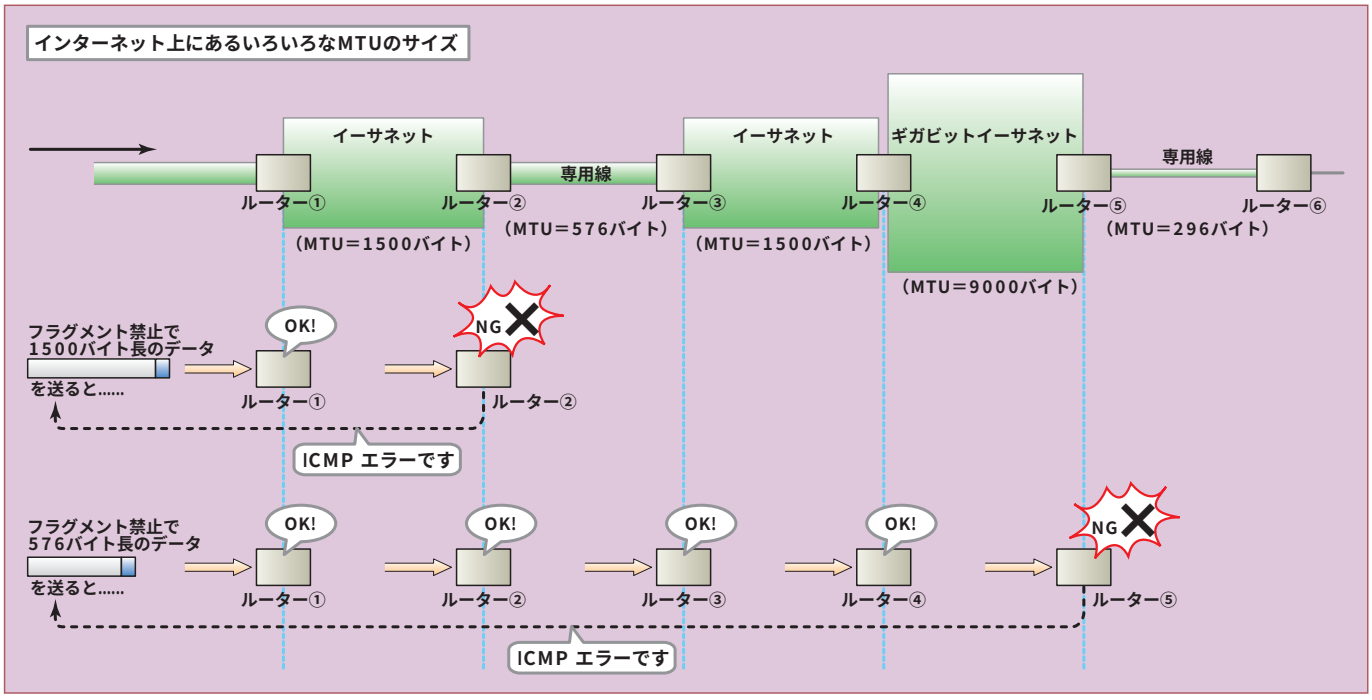

図4　Path MTUの検出の仕組み

［インターネットマガジン　バックナンバーアーカイブ］ ご利用上の注意

このPDFファイルは、株式会社インプレスR&D（株式会社インプレスから分割）が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン　バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

http://i.impressRD.jp/bn

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容（技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど）は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者（執筆者、写真の撮影者、イラストの作成者、編集部など）が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元（株式会社インプレスR&D）、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D（当時は株式会社インプレス）と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接的および間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先
株式会社インプレスR&D
All-in-One INTERNET magazine 編集部
im-info@impress.co.jp