平成 13 年度

学士学位論文

パケットアセンブリにおけるエッジルータ負荷に関する研究

A study on the load of packet assembly on edge routers

1020331　　山田　敦

指導教員　　島村 和典 教授

2002 年 2 月 8 日

高知工科大学 情報システム工学科

要 旨

パケットアセンブリにおけるエッジルータ負荷に関する研究

山田 敦

近年インターネットの普及に伴い，ネットワークを流れるパケットの数は，増加の一途をたどっている．このため中継ノードには，多くのパケットを効率よく処理し，高速に転送することが求められている．

しかし一方で，現在の IP によるパケット転送には多くの非効率な面がある．その一つとして，アクセス網から基幹網へと流れる IP パケットは，その多くが基幹網の MTU（Maximum Transfer Unit）に比べてはるかに小さい，という点が挙げられる．アクセス網の小さな MTU に合わせて小さく分割化されたパケットが，MTU の大きな基幹網内でもそのままのサイズで転送されているのである．データがたくさんの小さなパケットとして転送されると，中継ルータが処理するヘッダの数が増加するため，その処理負荷が増大する．また，ヘッダの総量が増えて，実データ以外の部分が大きくなってしまうと，無駄に多くのデータを送ることとなり，帯域使用効率が低下する．

これらの点に着目し，転送の効率化を図る方法として，筆者らはパケットアセンブリ [2] を提案する．本方式では，小さなパケットをエッジルータにおいて複数個を結合（アセンブリ）し，一つの大きなパケットとして基幹網へ送出する．そして効率よく基幹網ルータを中継し，再び目的のアクセス網の直前にあるエッジルータ上で，パケットを再構成（リアセンブリ）する．以上のようにして基幹網内のコアルータにかかる処理負荷を軽減することで，転送プロセス全体としての効率化を図る．

本方式では特に，パケットの結合・再構成処理による，エッジルータへの負荷の集中が懸念される．そこで本論文では，評価環境を作成しアセンブリ時の，エッジルータにおけ

る CPU 負荷の測定を行った．これらの測定結果を元に，本方式の有効性について検証を行った．

キーワード　パケットアセンブリ，MTU，IP，アクセス網，基幹網，負荷

Abstract

A study on the load of packet assembly on edge routers

Yamada Atsushi

With growing rate of Internet access around the world, the number of packets transferred over network goes on increasing. So the trunk nodes are required to handle the packets efficiently, and transfer them fast.

On the other hand, the present IP packet transmission has some inefficiency. In fact, the size of the packets transferred from the access network to the backbone network is smaller than the backbone network MTU (Maximum Transfer Unit). The packets adapted to the size of the access network's MTU are very small. These are transmitted even on the backbone network without changing size. If many small packets carry the data, the trunk nodes must process a lot of IP header, so increasing the processing overhead on the trunk nodes increase. And because of the increasing amount of packet header, we can't utilize the bandwidth in network efficiently. So we focus attention on that fact, and are studying the efficient method for transmitting the packets. We call the method as "Packet Assembly". In "Packet Assembly" method, the small packets are assembled at edge routers, and are transferred to backbone network as the large packet. After that, it is transmitted efficiently in backbone network. Finally, at the edge router just before the destination access network it is reassembled. In this way, we aim to reduce the load of the router on backbone network and to make the process of transferring packets more efficient.

In our system, the load on edge router may increase. So while assembling, we

measured the CPU utilization on there by conducting experiment.

From the result of the experiment we thought about the availabilities of Packet Assembly.

key words Packet Assembly, MTU, IP, access network, core network, load

目次

図目次

表目次

第 1 章

研究の目的

本研究の目的は，既存インターネットにおけるパケット転送を見直し，中継ルータにかかる負荷を低減することである．そのために筆者らは，**パケットアセンブリ**転送方式を提案する．本方式では，小さなパケットをエッジルータにおいて複数個を結合（アセンブリ）し，一つの大きなパケットとして基幹網へ送出する．そして効率よく基幹網ルータを中継し，再び目的のアクセス網の直前にあるエッジルータ上で，パケットを再構成（リアセンブリ）する．以上のようにして基幹網内のコアルータにかかる処理負荷を軽減することで，転送プロセス全体としての効率化を図る．

第 2 章

研究の背景

2.1　ネットワークの拡大

1969 年に 4 台のコンピュータで構築された「ARPA net」以来，30 年以上が経過し，現在のインターネットは約 1 億 1 千万台のコンピュータが接続される巨大なネットワークに成長している．Internet Software Consortium の Web サイト資料 [5] から作成したグラフを，図 2.1 に示す．これによると接続ホスト数は増加の一途をたどっており，それとともに

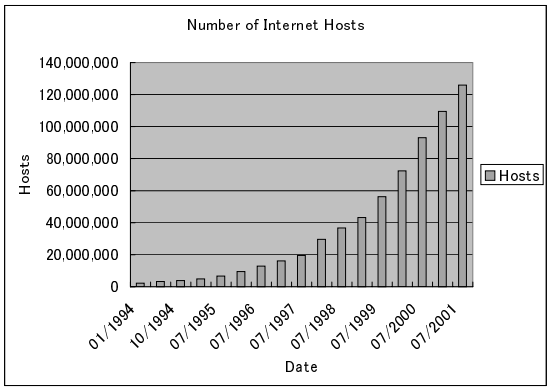

図 2.1　インターネットのホスト数

今後もトラフィック量が増加していくことは，まず間違いない．

2.2 現在の IP 転送における問題点

2.2.1 ヘッダ処理による転送負荷

トラフィックの増加はすなわち，ネットワーク中を転送されるパケット数の増加を意味する．パケットを転送する中継ルータは，受け取ったパケット一つ一つ（per-packet）について IP ヘッダを参照し，処理していく．per-packet で行われる処理には，ルーティングの決定，パケットヘッダの処理，チェックサム計算，入出力キュー間の移動がある．

ある調査 [4] によれば，現在 end-to-end でのホップ数は約 16 であり，中継する十数個のルータはすべて，per-packet でルーティング処理を行う．

トラフィックの増加に伴い，過大な数のパケットがルータへ集中すると，それぞれのルータでは，処理が間に合わずに遅延が生じたり，あるいはそのまま破棄されるなどということが起こる．ルーティング時のヘッダ処理にかかる負荷は per-bit，つまりデータ量によってではなく，per-packet で決まる．そのため同じデータ量であれば，できるだけ大きなサイズのパケットで転送をした方が，効率がよい．

図 2.2[3][6] は，1500B のパケットと 9kB に拡張されたパケットについて，サーバでのスループットと CPU 利用率を比較したものである．このグラフより，パケットサイズが大きい方がスループットは向上し，特に CPU 利用率は半減していることがわかる．このように，ルータにかかる負荷はパケットの処理個数，すなわちヘッダの処理回数に依存する．今後，増加を続けるトラフィックによって，転送処理にかかる負荷はますます増大するだろう．

2.2.2 基幹網内の小パケット

図 2.3 は，1998 年に InternetMCI バックボーンにおいて行われたトラフィック調査 [1] によるものである．この図は基幹網内でのパケットサイズの分布を示している．このグラフから，小サイズのパケットが圧倒的に多く，また 44，552，576，1500B にピークが集中していることがわかる．40〜44B の小パケットには TCP の確認応答（ACK）パケット，SYN，FIN，RST といった制御パケットや，telnet のキー入力キャラクタを運ぶパケットなどがあ

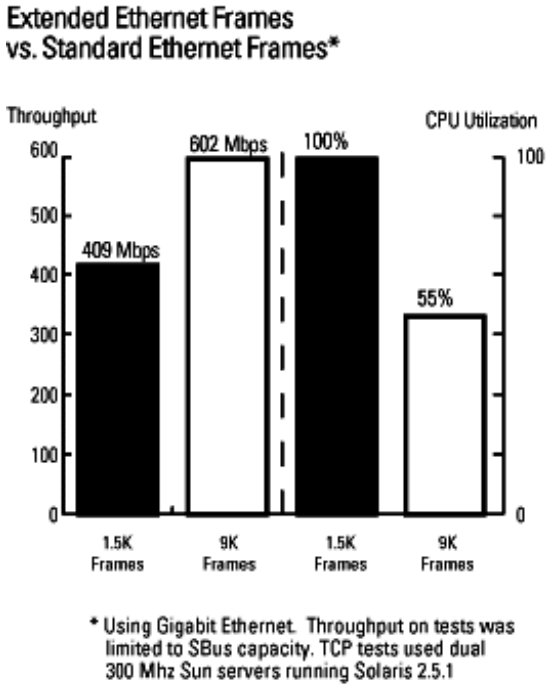

図 2.2 Extended Ethernet Frames vs. Standard Ethernet Frames

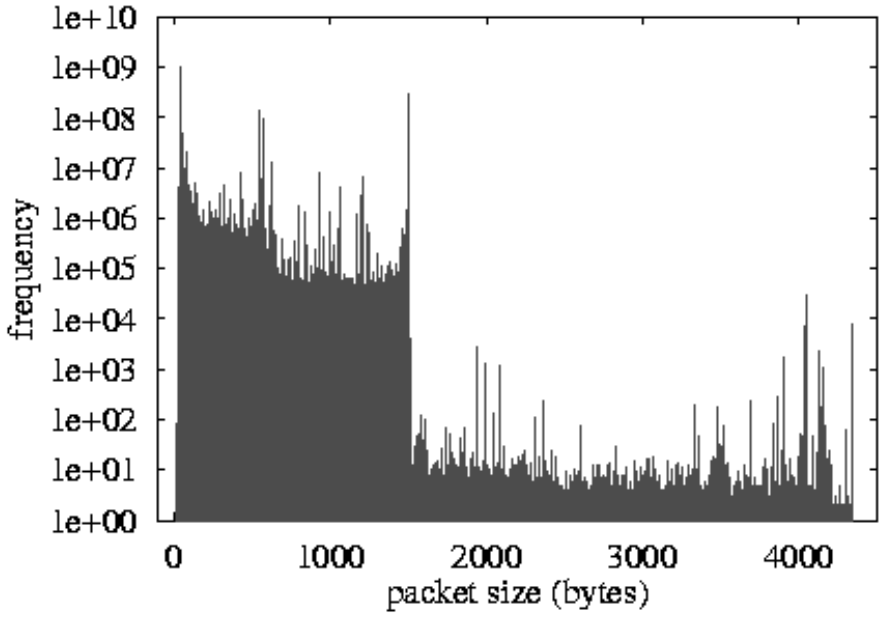

図 2.3 パケットサイズの分布

る．また，多くの経路 MTU 探索を実装していない TCP 実装系では，デフォルトの MSS（Maximum Segment Size）として 512，あるいは 536B を使用しているので，これに TCP と IP のヘッダが付加された，552B や 576B のパケットが多い．

1500B 以下のパケットが多いのは，アクセス網で使われる Ethernet の MTU（Maximum Transfer Unit）に関係がある．MTU とは，転送できる IP データグラムの最大サイズのことである．MTU はデータリンクによって異なる．ここで，各データリンクの MTU を表 2.1[9, Page: 135] に示す．この表に示されるように，Ethernet の MTU は 1500B である．

表 2.1　いろいろなデータリンクの MTU

データリンク	**MTU（オクテット）**	**Total Length（単位はオクテット，FCS 込み）**
IP の最大 MTU	65535	-
Hyperchannel	65535	-
IP over HIPPI	65280	65320
16MB IBM Token Ring	17914	17958
IP over ATM	9180	-
IEEE 802.4 Token Bus	8166	8191
IEEE 802.5 Token Ring	4464	4508
FDDI	4352	4500
Ethernet	1500	1518
PPP（Default）	1500	-
IEEE 802.3 Ethernet	1492	1518
IP の最小 MTU	68	-

Ethernet の MTU に合わせて送られたパケットが，基幹網内でもそのままのサイズで転送されているのだ．また，MTU に合わせて IP データグラムは分割化されるが，途中経路では再構築されない，ということも小サイズパケットが増える要因となっている．

2.3 既存技術

中継ノード負荷を低減する，既存の転送効率化技術を紹介する．

2.3.1 Jumbo Frame

Jumbo Frame は，イーサネットフレームのサイズを拡大することで，転送の効率化と端末側の処理負担の低減を実現する技術である．これは，米 Alteon WebSystems[*1]によって開発された．

Ethernet は作られた当時からずっと，1518B のフレームサイズを使ってきた．従来の 10Mbps や 100Mbps の Ethernet では，あまり大きなデータフレームが流れるとコリジョンが発生しやすくなるなど，問題があった．しかし，今日の Gigabit Ethernet（1000Mbps）は高速かつ低ロスであるため，1518B のフレームサイズでは，そのキャパシティを生かし切れない状況となった．そのため Jumbo Frame では，Ethernet のフレームサイズを 1518B から 9000B に拡張している．現在販売されている Gigabit Ethernet の NIC（Network Interface Card）や，スイッチの製品の多くは，この技術をサポートしている．

2.3.2 MPLS

MPLS（Multi Protocol Label Switching）は，シスコ社のタグスイッチング技術をベースに誕生した，IETF（Internet Engineering Task Force) 標準のテクノロジーである．

MPLS では，MPLS 対応のスイッチやルータ (LSR：Label Switch Router) が最初のパケットを受信すると，パケット内のルーティング情報に「ラベル」と呼ばれる短い固定長の情報を付け，次のホップ先に転送する．次にそのパケットを受信したルータはラベルを参照するだけで次のホップ先へ転送できるので，パケットごとにルーティング処理を行なう必要がなくなる．LSR 同士は LDP(Label Distribution Protocol) というプロトコルを用いてルーティング情報の交換を行ない，経路情報が変更されるとラベルを割り当てる．

[*1] 旧 Alteon Networks．現在は Nortel Networks の CNBU(Content Networking Business Unit) 部門．

MPLS では，1. ルータにかかる負荷を軽減することで，高速化を実現する，2. IP 層の経路情報を元にするので，ATM の下位層のルーティングプロトコルが不要．ATM ／フレームリレーと IP の緊密な統合が可能になる，3. ラベルによるネットワークの分離によって IP VPN を実現する，4. IP ネットワークにおいて ATM での高度な QoS 制御を利用することが可能になる，といったメリットが得られる．

欠点としては，一つずつのパケットの処理の負荷軽減はできても per-packet で処理負荷がかかっていることに変わりはないことや，MPLS に対応するエッジルータ，コアルータが必要なことなどがある．

第 3 章

パケットアセンブリ

3.1　パケットアセンブリのシステム

3.1.1　概要

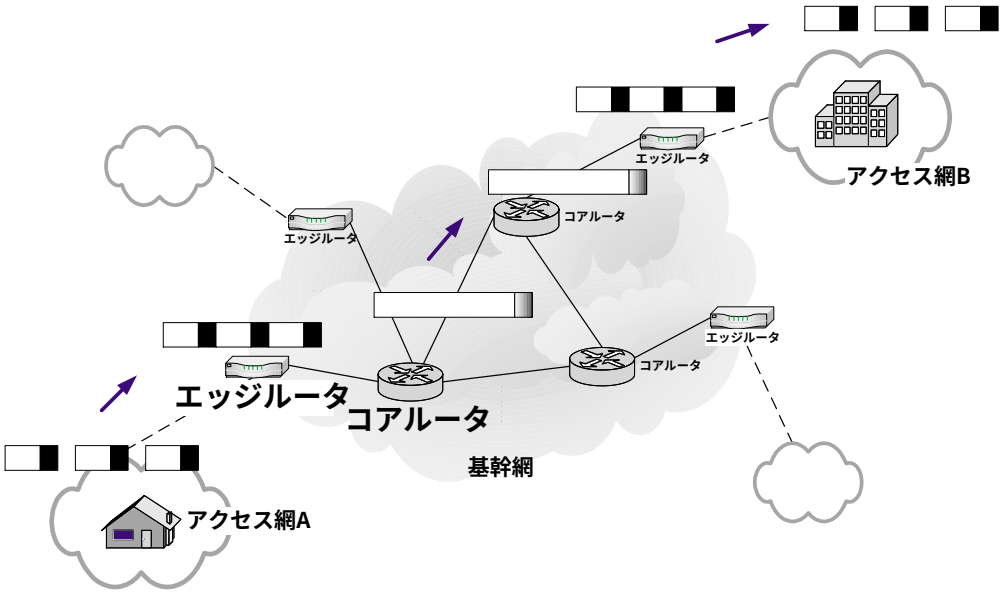

図 3.1　パケットアセンブリシステムの概要

本方式では，エッジルータ上で複数の IP パケットを大きな一つのパケットに合成する．ヘッダ数の減少によりこのパケットは基幹網を効率よく転送される．その後対向する側のエッジルータで元の小サイズのパケットに分割され，宛先ホストに届けられる．

たとえば図 3.1 の左下，アクセス網 A から右上のアクセス網 B へ，真中の基幹網を介し，パケットを送る場合を考える．まず，アクセス網 A を出た小サイズのパケットは基幹網に

入る前に，エッジルータで複数個が一つの大きなパケットにまとめられる．そしてこの合成されたパケットは，そのまま基幹網内を転送される．その後基幹網を抜け出る際に，パケットは再び元のサイズに分解され，アクセス網 B に送り届けられる．このような転送を行うことで基幹ネットワーク内のルータの負荷を下げ，転送効率の向上を図る．

3.1.2 特徴

変更を必要とするノードが少ない，MPLS との親和性など．

3.2 パケットアセンブリ転送方式の課題

3.2.1 エッジルータにかかる負荷

パケットアセンブリではエッジルータでパケットの合成・分割処理を行う．そのためコアルータのプロセッサにかかる負荷を低減する効果が期待できる．しかし一方で，エッジルータでのアセンブリ処理にかかる負荷の増加が予想される．そこで本研究では，エッジルータでのプロセッサ負荷を計測し，パケットアセンブリの影響について調べた．

第 4 章

アセンブリによるルータへの影響

4.1　エッジルータ負荷の測定

実験用ネットワークを構築し，エッジルータで擬似的なパケットアセンブリを行った．この時のエッジルータにおける CPU 負荷を測定した．

4.1.1　測定のねらい

エッジルータ負荷の測定を行うことで，以下に挙げることを調べる．

1. パケットアセンブリ処理によって，エッジルータではどのくらいの負荷上昇が起きるのか．
2. アセンブリ時に，コアルータ負荷が減少するか．
3. アセンブリ時，エッジルータ負荷はコアルータ負荷に比べ，増大するか．

4.1.2　測定方法

測定環境

図 4.1　測定ネットワーク構成

負荷測定を行うために 5 台の PC を使い，実験用ネットワークを図 4.1 のように構築した．これらの端末同士はすべて，Ethernet で接続されている．中心の 3 台の PC には NIC を 2 枚挿し，ルータとして機能するよう，設定した．この環境の中心にある端末は，実ネットワークのコアルータを，その両側の 2 台はエッジルータを模している．また，中央の 3 台のルータを基幹網と，そこからホスト A，B へとつながる部分をアクセス網と，見立てている．以後各端末を識別するため，左から順にホスト A，エッジルータ B，コアルータ，エッジルータ B，ホスト B と呼ぶことにする．

擬似アセンブリ

図 4.1 の環境を用い，エッジルータ A 上で擬似的にパケットアセンブリの操作を行う．“擬似的な” パケットアセンブリとは，IP の分割化（Fragmentation）と再構築（Defragmentation）処理を利用した，アセンブリ処理のことである．擬似アセンブリの動作を図 4.2 に示す．これは図 4.1 の左側 3 台の端末を拡大したものになっている．

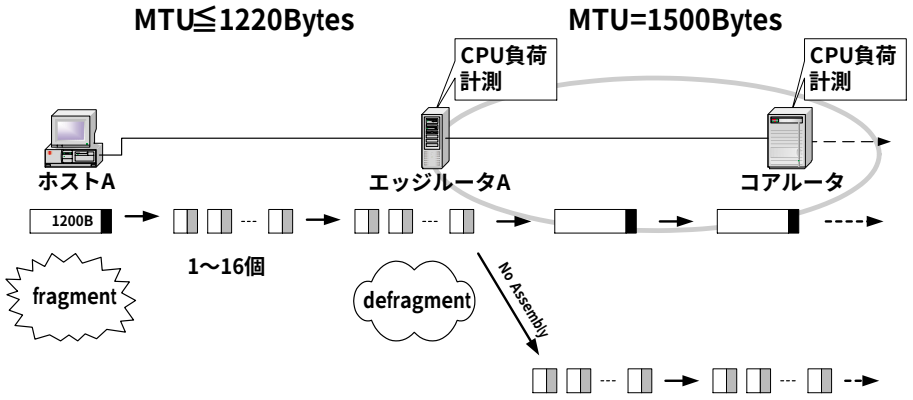

図 4.2　パケットの分割化と擬似アセンブリ

まず，基幹網の MTU を 1500 バイト，アクセス網を 1220 バイト以下に設定する．そして，1220 バイト一定サイズのパケットをホスト A から送出する．20 バイトは IP ヘッダの分である．そしてアクセス網の MTU を表 4.1 の通り変化させることで，パケットが送出時に 1〜16 個に分割化されるようにする．たとえば 4 つにフラグメントされるようにするた

表 4.1　送出時の分割化個数とアクセス網 MTU の対応

送出時の分割化個数（個）	アクセス網の MTU（Bytes）
1（分割しない）	1220
2	620
4	320
8	170
16	95

めには，MTU を 320 バイトにしておき，元の 1200 バイトのペイロードが，300 バイト×4 個になるようにする．300 バイトのパケットに 20 バイトの IP ヘッダが付加されるため，MTU は 320 バイトにしておく必要がある．このようにしてペイロードの総量は一定に保ち，エッジルータ A が受け取る**パケット数のみ**を変化させる．そしてエッジルータ A では，受信したパケットを常に再構築するよう，設定する．この再構築処理を本研究では「擬似アセンブリ」と呼んでいる．擬似アセンブリされたパケットは，送出された時の元の 1220 バイトのパケットに戻り，コアルータへと送られる．

擬似アセンブリが本来のアセンブリと異なるのは，受信した分割化パケットすべてを再構築する点である．擬似アセンブリでは，アセンブリ対象となるパケットを選択することは出来ず，送信時にフラグメントされたパケットを，元のサイズに戻すだけである．

測定

図 4.1，4.2 で示した環境で擬似アセンブリを実行し，エッジルータ A にかかる負荷を測定した．まず，擬似アセンブリする個数を変化させ，これによって CPU 負荷がどう変わるかを調べた．また，アセンブリを行わない状態でも同様に，処理個数による CPU 負荷の変化を測定し，擬似アセンブリ時との比較を行った．さらに，コアルータでの CPU 負荷測定もこれと同時に行い，エッジルータ負荷との比較を行った．以上をまとめると，次の通りで

ある．

1. 擬似アセンブリ使用時のエッジルータ A の CPU 負荷を測定した．
2. 通常転送時のエッジルータ A の CPU 負荷を測定した．
3. 擬似アセンブリ使用時のコアルータの CPU 負荷を測定した．
4. 通常転送時のコアルータの CPU 負荷を測定した．

このように擬似アセンブリ使用時／不使用時それぞれについて，エッジ／コアルータ双方で測定を行った．

具体的な測定手順を以下に示す．

1. 図 4.1 のホスト A からホスト B に対し，パケットを送信する．
2. MTU の調整により，送出されたパケットは分割化される．
3. 複数に分割化されたパケットがエッジルータ A に到達する．
4. エッジルータ A では擬似アセンブリを行う．（擬似アセンブリ使用時．）
5. エッジルータ A での CPU 負荷を測定する．
6. 同時にコアルータでの CPU 負荷を測定する．

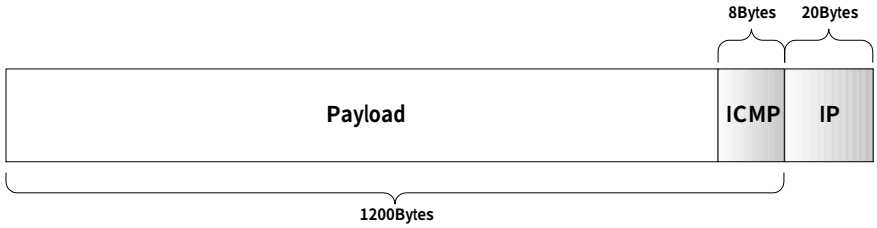

図 4.3　送信パケットの構成

パケットの送出には UNIX の ping コマンドを使用した．この時送信されるのは ICMP ECHO_REQUEST パケットで，図 4.3 のような構成となっている．また，ping の-f（flood ping）オプションを使用し，ECHO_REPLY パケットが戻ってくるとすぐに次のパケットが送出されるようにした．そして，1220 バイトの IP データグラムを 10000 個送出した．この

ときの負荷をエッジルータ A で測定した．これには UNIX の top コマンドを使い，100ms 毎にデータを採った．

4.1.3　測定結果

測定に使用した top コマンドの出力のうち “システム負荷” の値を集計し，グラフに表した．

エッジルータの CPU 負荷

エッジルータでの CPU 負荷測定の結果を，図 4.4 に示す．ここではアセンブリを行わず，受信したパケットをそのまま送出した場合のことを “非アセンブリ” と表現している．グラフの横軸は，ルータが一度に受信するパケットの個数で，縦軸は CPU 負荷である．このグラフでは，擬似アセンブリを行った場合と，非アセンブリ時を比較している．グラフは，アセンブリ個数（非アセンブリ時はルータを通過するパケットの個数）が増加するに従い，負荷が上昇することを示した．また，擬似アセンブリを行わない場合に比べ，行ったときの方が，負荷が減少するという結果を示した．

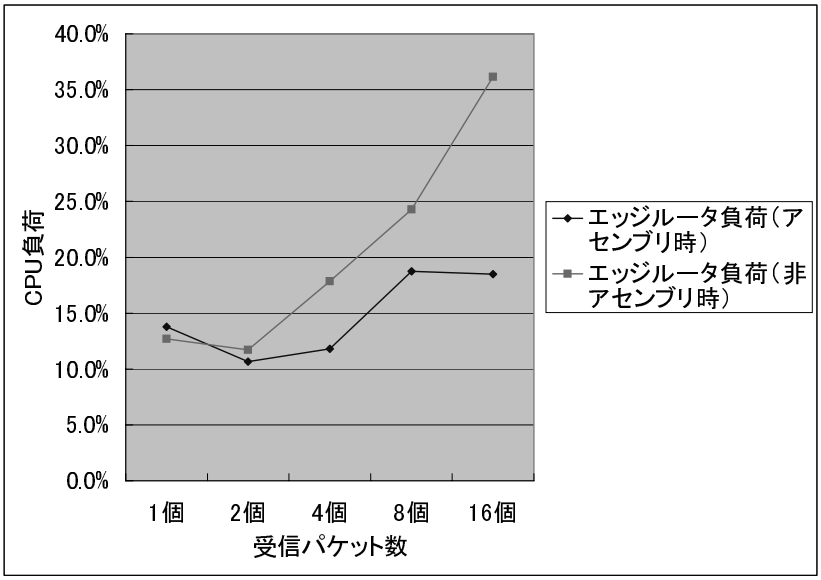

図 4.4　エッジルータの CPU 負荷

コアルータの CPU 負荷

コアルータでの CPU 負荷測定の結果を，図 4.5 に示す．このグラフでも，擬似アセンブリ時と，非アセンブリ時を比較している．全体の傾向として，非アセンブリ時に比べ擬似アセンブリ時の方が，低負荷という結果になった．

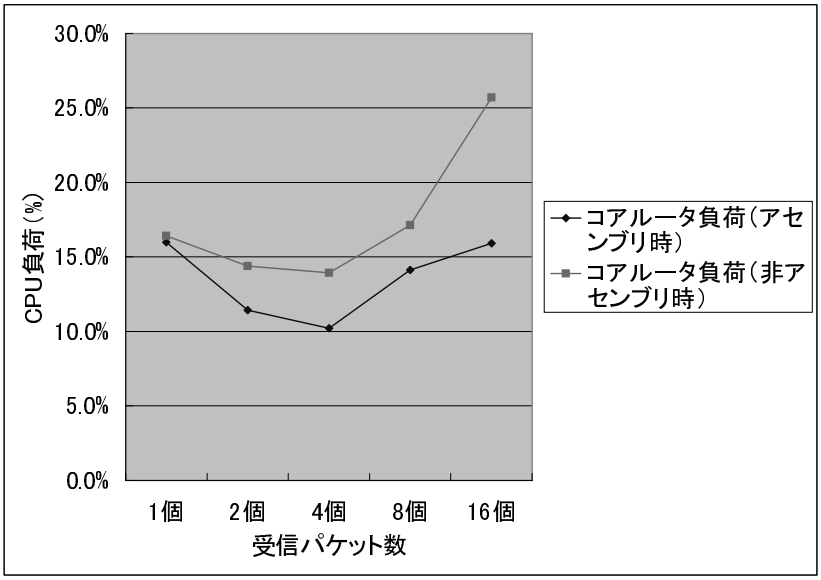

図 4.5　コアルータの CPU 負荷

擬似アセンブリ実行時の CPU 負荷

擬似アセンブリ実行時の CPU 負荷測定の結果を，図 4.4 に示す．

通常転送時の CPU 負荷

通常転送時の CPU 負荷測定の結果を，図 4.4 に示す．

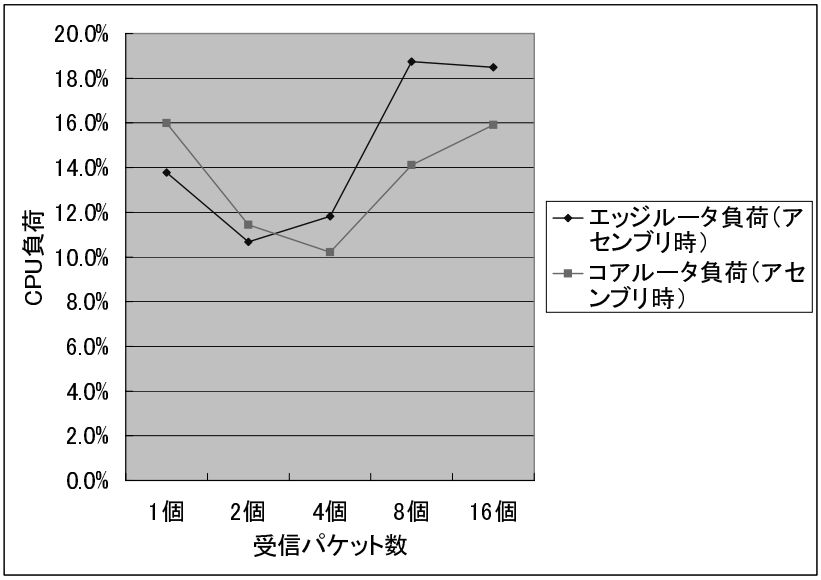

図 4.6　擬似アセンブリ実行時の CPU 負荷

4.1.4　考察

エッジルータ負荷の低減

エッジルータにかかる負荷は，擬似アセンブリの処理が余計にかかるにも関わらず，非アセンブリ時に比べ，低減される結果となった.

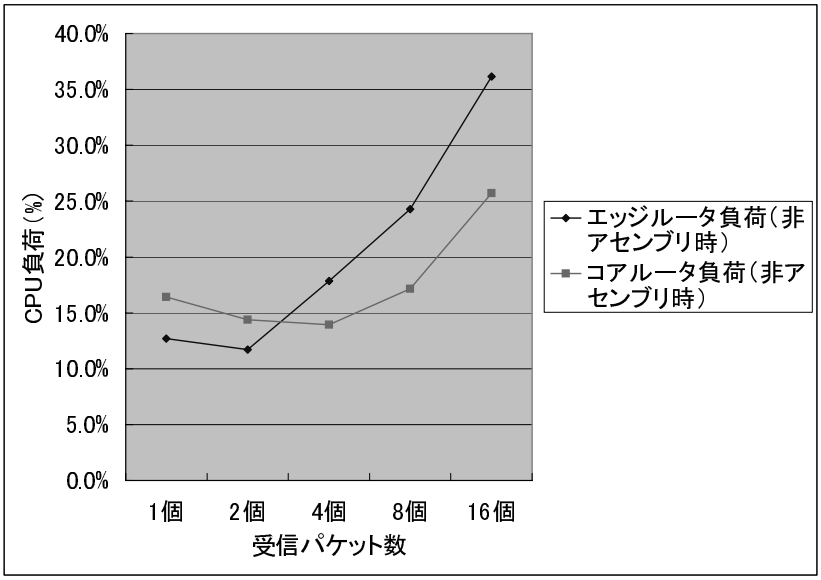

図 4.7　通常転送時の CPU 負荷

4.1　エッジルータ負荷の測定

コアルータ負荷の低減

第 5 章

考察

パケットアセンブリ処理によるルータの CPU 負荷は，エッジルータ，コアルータともに低減することが実験結果からわかった．このことから，パケットアセンブリが，中継ルータの負荷軽減に有効であることがわかった．

第 6 章

今後の課題

基幹網に ATM を使用した実験．現在開発を進めているアセンブリ機能を持ったルータでの実験．

謝辞

本研究を進める上で，貴重なご助言や多大なるご指導を頂いた高知工科大学工学部情報システム工学科の島村 和典教授，並びに通信・放送機構高知トラヒックリサーチセンターの神田 敏克氏に深く感謝したいと思います．また，島村研究室のみなさんと共に卒業研究に取り組めたことを，心よりうれしく思います．

参考文献

[1] K Claffy, Greg Miller, Kevin Thompson, The nature of the beast: Recent traffic measurements from an Internet backbone, INET'98 Conference, 1998.

[2] T.Kanda and K.Shimamura, "Load reduction for the node processors in core networks by packet assembly." IEEE CQR Technical Committee, CQR International workshop 2001 in Tucson, U.S.A.

[3] Phil Dykstra, Extended Frame Sizes for Next Generation Ethernets - a white paper by Alteon, 1999.

[4] DARPA ITO Information Technology Office.
http://www.darpa.mil/ito/research/ngi/supernet.html(February 5, 2002).

[5] Internet Software Consortium - Number of Internet Hosts
http://www.isc.org/ds/host-count-history.html(February 7, 2002).

[6] Phil Dykstra, Gigabit Ethernet Jumbo Frames, 1999.
http://sd.wareonearth.com/ phil/jumbo.html(January 15, 2002).

[7] 小林 寛征，神田 敏克，島村 和典，“高速パケットデフラグメント自律化方式の制御に関する研究”，2001.

[8] 浦西 慶規，神田 敏克，島村 和典，“ネットワークトランスファ性能評価とその高機能化に関する研究”，2001.

[9] 竹下 隆史，村山 公安，荒井 透，苅田 幸雄，“マスタリング TCP/IP 入門編 第 2 版”，オーム社，1998.