

Comet による IEEE1394 を利用した計算機ネットワークの構築

古賀 久志、陣崎 明

新情報処理開発機構 並列分散システム富士通研究室

〒211-8588 川崎市中原区上小田中 4-1-1

E-mail: {koga,zinjin}@flab.fujitsu.co.jp

Cometを接続して並列ネットワークサーバを構築するための高速ネットワークとして、我々はIEEE1394に注目している。一方で、IEEE1394は125 μ 秒毎にデータを遅延なく届けるQoS機能を持つ。今回、Cometのプロトコル処理性能の評価を目的としてIEEE1394同期パケットをIPカプセル化してEthernetとゲートウェイするシステムを構築した(IEEE1394/IP)。従来のネットワーク処理では割り込みオーバーヘッド等で上記プロトコル変換を125 μ 秒単位で行えないが、Cometではネットワークアダプタにプロトコル処理をオフロードすることにより約40 μ 秒で1パケットを処理する。IEEE1394/IPを用いて、WIDEバックボーンでのビデオストリーム中継実験及び最新ルータで生じるジッタの定量測定を行った。その結果、(1) Cometのみで構成されるネットワークでは125 μ 秒のパケット間隔を維持できるのに対し、(2) 最新の商用ルータでもジッタが2倍(250 μ 秒)に膨らむ、ことが確認された。

広域ネットワーク、IEEE1394、ジッタ、ネットワークサーバComet、IP通信、ルータ

IEEE1394 Is Intergrated into Computer Networks by Comet

Hisashi Koga, Akira Jinzaki

Parallel and Distributed Systems Fujitsu Laboratory, RWCP

1-1, Kamikodanaka 4-Chome, Nakahara-ku, Kawasaki, 211-8588 Japan

E-mail: {koga,zinjin}@flab.fujitsu.co.jp

We consider IEEE1394 as a high speed network to connect Comets and form them into parallel network servers. Apart from that, IEEE1394 provides QoS function in which all data arrive at the destination without any delay. To evaluate the protocol processing performance of Comet, we construct a system which transforms IEEE1394 isochronous packets to IP packets and forwarded them to Ethernet (IEEE1394 over IP). While the previous network processing mechanism cannot process this protocol conversion every 125 micro-seconds due to the interruption overhead, Comet can process it at 40 micro-seconds by offloading it onto a network adapter. With the use of this IEEE1394 over IP function, we made DV-stream relay experiments on the WIDE backbone network. We also measured how much jitter is generated when a packet passes a router. As the result of these experiments, the following two facts are obtained.

- (1) When the network consists only of Comet machines, the 125 micro-second packet intervals are maintained.
- (2) Even the latest commercial router expands the jitter two times (i.e. 250 micro-seconds).

1. はじめに

Comet ネットワークサーバ [1]はプロトコル処理をネットワークアダプタにオフロードすることにより高速なネットワーク処理を実現するシステムであり、具体的にはプロトコル処理エンジンを搭載した PCI カードを標準の PC、WS に搭載したものである。我々は Comet ネットワークサーバを Gbps クラスの標準ネットワークで接続してクラスタ化し、Web サーバのようなサーバ機能と高速のルータ機能を兼ね備えた並列ネットワークサーバを構築することを目標にしている。

Comet ネットワークサーバを接続する高速ネットワーク媒体として IEEE1394 に注目した。IEEE1394 の転送速度は数年後には 3.2Gbps になり、しかも ATM などと比べて大変安価という利点がある。また、IEEE1394 はバスプロトコルであり、read、write transaction を基本としたメモリアクセスイメージの通信機構を持ち、並列分散システムでよく使われる分散共有メモリ型通信を無駄なく実現できる。

一方で、IEEE1394 は 125 μ 秒毎に遅延なくパケットを送り届ける QoS 機能を備え（同期モード）、この機能を LAN や WAN に応用することが期待されている。しかし、IEEE1394 ではケーブル長が最大 4.5 m に制限され、遠距離間通信を行えない。この制限を回避する為に IEEE1394 パケットを IP 化して Ethernet とのゲートウェイを用いて遠隔の IEEE1394 ネットワークを結合する方法が考えられる。

ここで問題になるのは IEEE1394 と IP とのプロトコル変換を行いつつ、いかに 125 μ 秒毎というパケット処理間隔を守るかである。もし、125 μ 秒のパケット間隔が維持できなければ、折角の IEEE1394 の QoS 機能を殺してしまう。当研究室の目指す「超高速インターネット」の構想においては、今後のインターネットではこのように μ 秒の精度で IP 通信のジッタを抑える機能もインフラとして提供すべきであると考えている。

今回、Comet の高速プロトコル処理性能を評価する事を目的として IEEE1394 同期モードで流れるデジタルビデオ（以下 DV）ストリーム（30Mbps/1 チャンネル）を IP カプセル化してインターネットに流し、それを

受信してビデオ再生するシステムを開発した（IEEE1394/IP）。その結果、

- 送信側 Comet と受信側 Comet を直結したネットワークでは、ノイズをのせずに DV を安定再生できる。

という良好な結果が得られた。Comet ボード内での IP カプセル化及びパケット forwarding に要する処理時間の合計は 1 パケットあたり 40 μ 秒であった。

本 IEEE1394/IP システムを用いて日本国内のインターネット研究の中心研究組織である WIDE のバックボーンネットワーク上での DV ストリーム通信実験（定性評価）を行った。すると、

- 最新のルータで構成されたバックボーンネットワークであっても DV ストリームを 2 チャンネル流すとビデオ画像が激しく乱れる。

という結果になった。この結果を受けて、受信側 Comet でパケット処理開始時刻の間隔を計測し、Comet および現在の最新ルータを通ることによるジッタの膨みを定量的に比較したところ、

- やはり Comet を直結した場合はパケット処理間隔が 125 μ 秒を中心とした正規分布に従う。
- ルータを一段通ると 250 μ 秒にジッタが膨らむ。

ということが判明した。このように Comet を使えば従来方式では不可能な μ 秒単位の高精度な通信が実現できる。例えば、今後発展する広域並列分散処理においてもデータ到着時刻の予測が可能になると期待される。

以下では 2 節で Comet 上での IEEE1394/IP 実装について説明し、3、4 節でそれぞれ WIDE バックボーンでの定性実験、ルータを通った場合のジッタの定量測定について詳しく述べる。5 節でまとめを述べる。

2. IEEE1394/IP

2.1. IEEE1394 の概要

IEEE1394 とは PC とディスクやビデオ装置などの周辺機器を接続する目的で開発された高速シリアルバス規格である。IEEE1394 には同期モードと非同期モードの 2 つの転送モードがある。このうち、特に同期モードでは 125 μ 秒毎に遅延なく定期的にパケットを転送する QoS 機能が有り、マルチメディアデータを途切

れなく転送するのに向く。この性質から IEEE1394 は DV データ転送規格に採用されている。

また、転送性能も大変高速で、現在 400Mbps まで規格化され、将来的には 3.2Gbps まで拡張される。また、コントローラチップも大変安価であり、200Mbps のチップセット (LLC チップと PHY チップ) の価格が \$20 を下回った[5]。これらの理由から最近では PC クラスタを構成するための高速ネットワークとしても注目されている。

しかし、スケーラビリティの面では

- 1 リンクの最大長が 4.5m。
- 1 物理ネットワークに存在できるノード最大数が 64。

と制限があり、遠距離通信は出来ない。

2.2. Comet による IEEE1394/IP 実装

我々は IEEE1394 同期モードで流れる DV ストリームパケットを Comet で IP カプセル化してインターネットに送出してビデオ再生するシステムを開発した (IEEE1394/IP)。これは前節で述べた IEEE1394 のスケーラビリティがないという欠点を克服する。

2.2.1. Comet

Comet は通信プロトコル処理をネットワークアダプタにオフロードしてハードウェア処理する事により、通信の高速化を目指したシステムである[1]。Comet ネットワークアダプタはプロトコル処理専用エンジンを搭載した PCI カードであり、PC、WS の PCI スロットに挿して使用する(図 1)。現在の試作[2]では、プロトコル処理エンジンを Comet ボード上の 166MHz RISC によってソフトエミュレーションしている。1 枚の Comet ネットワークアダプタには PMC (PCI Mezzanine Card) 規格の NIC (Network Interface Card) を 2 枚ドータボードとして装着可能である。

2.2.2. IEEE1394/IP 実装

Comet ネットワークアダプタに IEEE1394 と 100Base-TX を装備し、DV 装置を IEEE1394 で接続する。DV カメラは 125 μ 秒毎に DV ストリームパケット (約 500byte)を同期転送モードで流すので、これを Comet

で受けて IP カプセル化し、100Base-TX 側から Ethernet に送出する (図 1)。これを受信した Comet では、IP パケットから IEEE1394 同期パケットを取り出して IEEE1394 バスに転送し、DV 機器がそれをリアルタイムで再生する。

従来の計算機ネットワークでこうしたプロトコル変換を行う場合、アダプタからホストに割り込みを上げ、プロトコル変換を行い、アダプタにデータを送り返すという処理が必要で、これを 125 μ 秒という高精度で行うのは難しい。Comet ではネットワークアダプタ内に処理をオフロードすることで 1 パケットあたり 40 μ 秒で処理する (図 2)。ホストは Comet のコンフィグレーションを行うだけである。

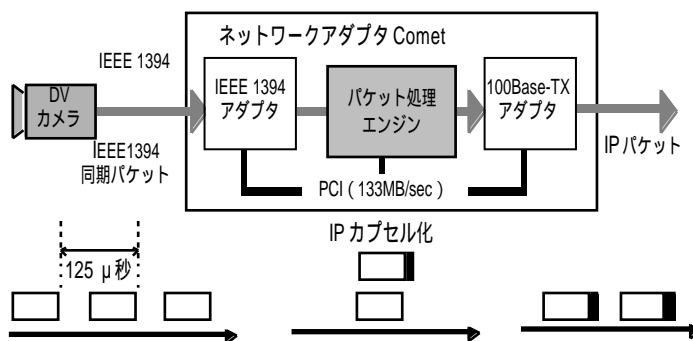
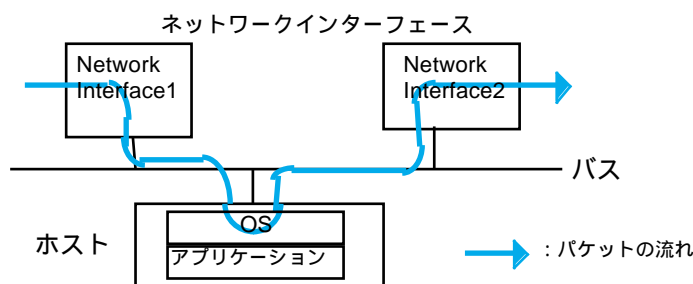


図 1 : Comet による IEEE1394 同期パケットの IP 化

(a)従来方式



(b) Comet

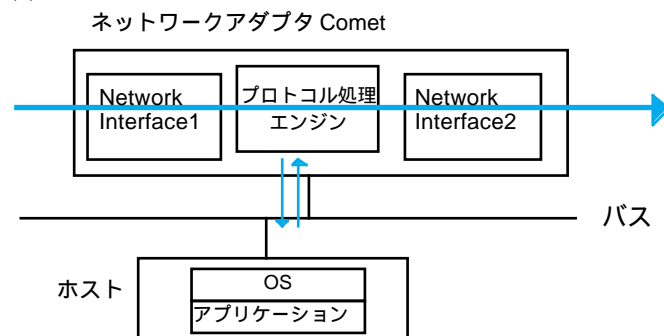


図 2 : 従来方式と Comet 方式

受信側 Comet に 125 μ 秒単位でパケットが到着しなければ安定した画像再生はできないが、送信側 Comet と受信側 Comet を直結させたところ、ノイズを一切出さずに安定した画像再生ができた。すなわち、Comet のみで構成したネットワークであればジッタを抑えられる。この場合については JSPP'98 のポスターセッションでデモを行った[3]。

この IEEE1394/IP は以下の 2 点からネットワークのジッタを測定するツールとしても使える。

- ビデオ画像の乱れ具合からパケット間隔が一定に保たれているかを定性的に容易に確認できる。
- IEEE1394 上を流れる DV ストリームは 1 チャンネルにつき 30Mbps、パケット長が固定長 492byte とあらかじめその性質が分かっており定量的に解析しやすい。厳密には 6% 強の割合で DV データを含まない 12byte 固定長パケット¹が混じる。

3. WIDE バックボーンでの実験(定性評価)

1998 年 5 月に Comet を WIDE プロジェクトの広域バックボーンに接続して、IEEE1394/IP 機能を利用し、実際に使われている広域網でジッタを抑えられるかの定性的な評価実験を行った。その結果を紹介する。以下のいずれの実験においても回線の太さは双方向に 40Mbps 以上ずつを確保し、DV ストリームを流しても回線の太さが問題にならないようにした。

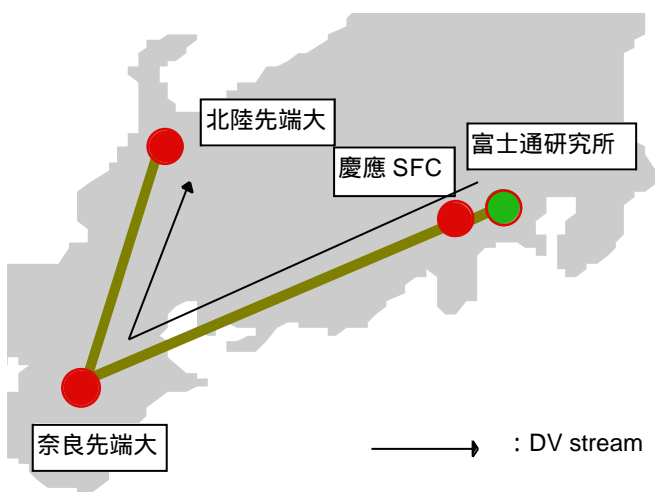


図 3: WIDE バックボーンでの DV 通信実験 (1 回目)

¹ 全パケットが 492byte だと、 $492 \times 8 \times 8000 = 31.488\text{Mbps}$ となり、DV ストリームの帯域(30Mbps)を上回る。

第 1 回目の実験では、富士通研から北陸先端大学に慶應大学湘南藤沢キャンパス (慶應大学 SFC)、奈良先端大学を經由して DV ストリームを片方向で 1 チャンネル流す実験を行った (図 3)。途中の経路には 5 段の IP ルータ、7 台の ATM スイッチが挟まる。それなりに安定した再生ができたが、Comet 直結の場合と異なり、映像が止まったためにビデオ再生用 NTSC モニタが真っ白になる場合があった。

第 2 回目の実験では、奈良先端大から慶應大学 SFC、北陸先端大に WIDE 研究会の中継を行いながら、同時に SFC からの画像を奈良先端大に送るという実験を行った (図 4)。各経路の途中に存在するルータと ATM スイッチの台数を表 1 に示す。

表 1 : 第 2 回目の実験での中継実験網

経路	IP ルータ	ATM スイッチ
奈良先端 ~ 北陸先端	3 台	3 台
奈良 ~ 慶應 SFC	4 台	2 台

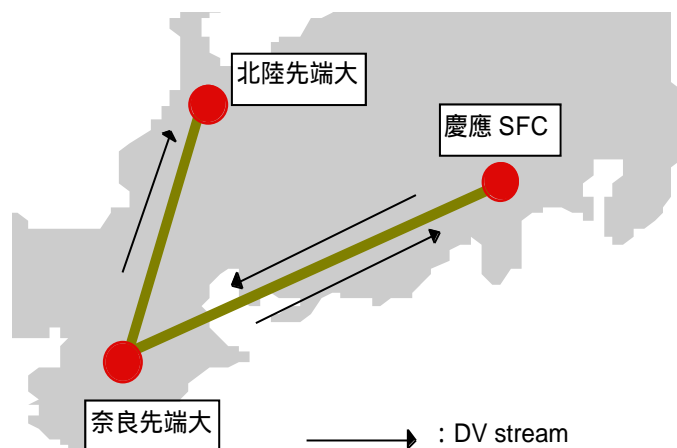


図 4: WIDE バックボーンでの DV 通信実験 (2 回目)

奈良先端大から北陸先端大、SFC への映像は第 1 回目の実験と同程度の品質で流せたが、SFC から奈良へ映像送信を開始すると何が映っているか判別できなくなった。最新ルータで接続されたネットワークでも、双方向に DV を 2 ストリーム流せないことが分かった。

この実験後、6 月に Interop'98 でも池袋と幕張会場間でビデオストリームを 2 チャンネル流し、最新ルータを

経由して中継する実験を行った。第2回目の実験と同様、双方向にDVストリームを流した場合は安定したビデオ再生が行えなかった。

4. 定量的なジッタ測定

Cometを直結した場合は高精度の通信を実現できるのに対し、最新ルータで構成されたバックボーンネットワークではジッタが制御できないという定性的な結果を示した。本節ではCometおよびルータを一段通るとどれだけジッタが発生するかを定量的に議論する。

4.1. 測定ツール

受信側のCometファームウェアにパケット処理間隔を測定する関数を埋め込んだ。CometファームウェアのIEEE1394ドライバのパケット処理開始時刻を測定し、それからパケット処理間隔を算出する。時刻の計測にはCometボード上にあるタイマを利用した。このタイマの1 tickは0.96 μ秒である。DVストリームに含まれる12byteの短いパケットについてはDMA転送やネットワーク伝送に要する時間の違いで125 μ秒単位でパケットが届かないので、その前後ではパケット処理間隔を測定していない。また、パケットロスが発生した場合もその前後では測定を行わないようにした。但し、今回のルータ一段を通した実験ではパケットロスは発生しなかった。

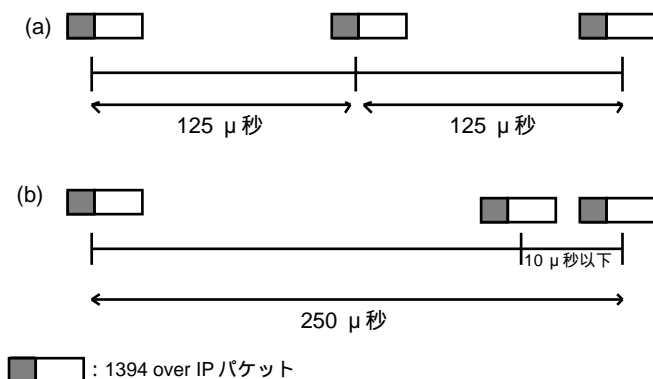


図 5：パケット処理間隔

図 5(a)のようにジッタが制御されていればパケット処理間隔は125 μ秒になる(ファームウェア main loop 1周程度の誤差は出る)。一方、図 5(b)のようにジッタがあるとパケット処理間隔が非常に大きくなったり、逆にファームウェアで2個一度にパケットを処理して

しまうことが起こる。後者ではパケット処理間隔が10 μ秒以下としてカウントされる。

4.2. 実験システム

ルータを通さずCometを直結した場合とルータを通した場合の遅延の揺らぎを計測した。以下ではルータを通る場合の実験システム構成について述べる(図 6)。

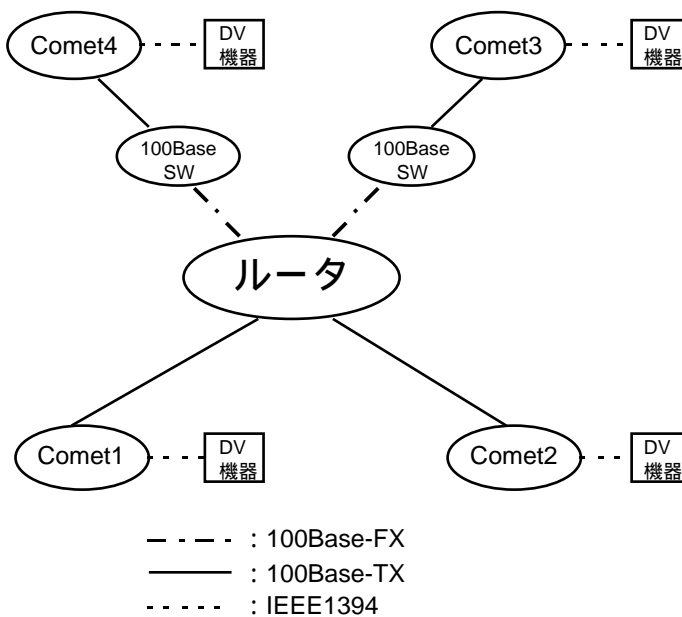


図 6：ルータを用いたジッタ測定実験システム構成

CISCO7500に100Base-TXと100Base-FXのネットワークカードを挿した、それぞれのカードには2つずつインターフェースがついている。こうして分けられる4個のネットワークに1台ずつCometを配置し、ルータを越えてDVストリームを流した時のパケット処理間隔を測定した。以下の3ケースについて計測した。

1. バックプレーンを通らない場合：100Base-TX から別の100Base-TXにDVストリームを1チャンネル流す(図 6のComet 2 から Comet 1 に向けてDVストリームを1チャンネル流す)。
2. バックプレーンを通る場合 100Base-FX から100Base-TXに向けてDVストリームを1チャンネル流した(Comet 3 から Comet 1 にDVストリームを1チャンネル流す)。
3. バックプレーンを通して2チャンネルDVストリームを通した場合(Comet 2 から Comet 4、Comet 3 から Comet 1 にDVストリームをそれぞれ1チ

チャンネルずつ流した)。

測定はすべての場合で Comet 1 で行った。3 番目のケースでは Comet 2 から Comet 4 に向けて流れる DV ストリームは Comet 1 で受信する DV ストリームに対してバックプレーンで衝突するノイズになる。

4.3. 結果

測定結果を(図7)に示す。20万回パケット処理間隔を測定し、16 tick(約15 μ秒)のきざみで分布を求めて%表示したものである。4.1節で記述した Comet ファームウェアで2個のパケットを1度に処理した場合のパケット処理間隔は10 μ秒未満なので、グラフでは0 μ秒として表示されている。

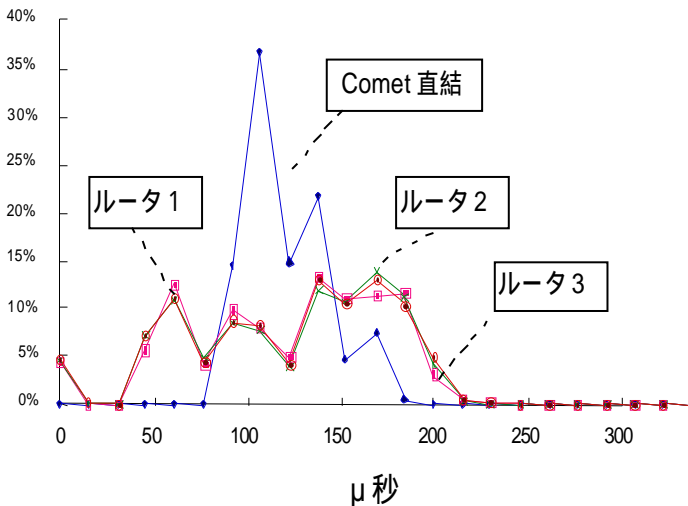


図7: パケット処理間隔の分布

1. Comet 直結の場合には 125 μ秒を中心とした正規分布に従う。また、0 μ秒の所が 0%になっており、受信側ファームウェアで2個パケットを1度に処理することはなく、ジッタが制御できていることがわかる。
2. ルータを1段通るとパケット処理間隔が250 μ秒まで膨らむ。グラフでは見づらいが300 μ秒を越えた場合もあり、ルータを5段も通れば、1.5m秒にも迫るジッタになる。また、パケットを2個同時に処理するということが起きた。しかし、一方でバックプレーンを通る、通らないやバックプレーンに30MbpsのDVストリームノイズが有る無いかの影響ははっきりと見えなかった。

但し、WIDE バックボーンでの実験では画像が激しく乱れたが、今回の測定中は DV ストリームを2チャネ

ル流しても時々ブロックノイズがのる程度でかなり高品質の再生ができ、現在のバックボーンネットワークの振舞いが完全に再現できたわけではない。

5. 結論

Comet はネットワークアダプタ内でプロトコル処理を行う事により、125 μ秒の高精度で IEEE1394 同期モードパケットの IP 化を実現する。現在はネットワークアダプタ上のプロトコル処理エンジンをソフトエミュレーションしているが、今後 FPGA 化、ASIC 化すれば、より高精度な通信を実現できる。逆に WIDE バックボーンネットワークでの DV ストリーム通信実験、ジッタ定量測定で示されたように、最新ルータを用いても Comet のような高精度の通信は実現できない。

今後は Comet にルータ機能を実装し、ルータを含め Comet のみでネットワークを構成すれば、広域 IP ネットワークでもジッタを抑えられ IEEE1394 同期モードが要求する高精度な通信を提供できることを実証する。また、IEEE1394/IP をジッタ評価ツールとして用いて、ルータを他段接続した時にジッタがどの程度膨らむのかを測定し、現在のバックボーンネットワークの振舞いを引き続き明らかにしたい。

参考文献

- [1] 陣崎、中村、村井：並列ネットワークサーバ Comet のアーキテクチャとその応用、信学技法 SWoPP'98 CPSY, 1998
- [2] 河合、下國、都筑、竹原、陣崎：Comet のハードウェア試作と性能評価、信学技法 SWoPP'98 CPSY, 1998
- [3] 古賀、陣崎：IEEE1394 バスの計算機ネットワークへの適用、並列処理シンポジウム JSPP'98 論文集 pp.158,1998
- [4] IEEE Std 1394-1995: IEEE Standard for a High Performance Serial Bus, IEEE Computer Society, 1996
- [5] <http://www.ti.com/sc/docs/msp/1394/products.htm>