

LE2005a における StarBED を利用した 大規模ストリーミングサーバの構築

下川 俊彦[†] 池野 桂司[‡] 池上 恵介[‡] 東 正美^{*} 青木 哲郎⁺

† 九州産業大学情報科学部 〒813-0011 福岡市東区松香台 2-3-1

‡ 北海道総合通信網株式会社 〒060-0031 札幌市中央区北1条東2丁目 5-3

*松下電器産業株式会社 〒541-0043 大阪市中央区高麗橋 1-6-8

[†]情報通信研究機構 〒184-8795 小金井市貫井北町 4-2-1

E-mail: [†]toshi@is.kyusan-u.ac.jp, [‡]{ikeno, ikegami}@hotnet.co.jp, ^{*}mazuma@pana.net, ⁺aoki@nict.go.jp

あらまし StarBED と JGN2 を用いて、大規模ストリーミングサーバを構築した。このサーバは 175 台の PC サーバと 10Gbps の対外接続を持つ。このサーバを LE2005a 日食中継で利用し配信実験を行い評価した。

キーワード ストリーミング、大規模配信、StarBED、JGN2

Construction of Large Scale Streaming Server using StarBED at LE2005a

Toshihiko SHIMOKAWA[†] Keiji IKENO[†] Keisuke IKEGAMI[†]

Masami AZUMA^{*} and Tetsuo AOKI[†]

† Faculty of Information Science, Kyushu Sangyo University 2-3-1 Matsukadai, Higashi-ku, Fukuoka, 813-0011 Japan

‡ Hokkaido Telecommunication Network Co., Inc. Kita 1-jo Higashi 2-chome 5-3, Chuo-ku, Sapporo, 060-0031 Japan

* Matsushita Electric Industrial Co., Ltd. 6-8, Koraibashi 1-Chome, Chuo-ku, Osaka, 541-0043 Japan

[†]National Institute of Information and Communications Technology 4-2-1 Nukui-kita, Koganei, Tokyo, 184-8795 Japan

E-mail: [†]toshi@is.kyusan-u.ac.jp, [‡]{ikeno, ikegami}@hotnet.co.jp, ^{*}mazuma@pana.net, [†]aoki@nict.go.jp

Abstract We construct a large scale streaming server using StarBED and JGN2 network. This server contains 175 PC servers and 10Gbps external link. We test and evaluate this server in LE2005a Solar eclipse live streaming.

Keyword Streaming, Large Scale Contents Distribution, StarBED, JGN2

1. はじめに

インターネットが普及し、職場や家庭でも広帯域なネットワーク環境が利用可能になってきた。この環境を活かす応用として動画像配信の重要性が高まってきている。

しかし、大規模な数の利用者に対する動画像配信は容易ではない。単一の ISP 内の利用者に対する動画像配信では Content Delivery Network(CDN)を利用するところが一般的である。しかし、複数の ISP を対象とした大規模配信においては、単純な CDN ではその処理性能が不足する。この問題を解決する手法の一つが Content Routing Network (CRN)アーキテクチャである。これは、複数の CDN をピアリング（接続）することで処理性能の拡大を狙ったものである。筆者らが活動を進めている CRN フォーラム[1]では、実際に CRN アーキテクチャに基づいたネットワークを構築し、大規模配信に関する研究開発を進めている。

さまざまな大規模配信実験を行っていく中で、CRN アーキテクチャの問題点の一つとして管理運用の困難さという点が明らかになった。これは、CRN アーキテクチャが分散型のシステムであることに起因する。

そもそも CRN アーキテクチャが考案された理由は単一の CDN の処理性能の不足である。そこで本研究では、大規模な処理性能を持つ単一 CDN (ストリーミングサーバ) による大規模配信技術に関する検討を行った。この大規模 CDN は、情報通信研究機構の大規模ネットワーク実験施設 StarBED[2][3]と、超高速・高機能研究開発テストベッドネットワーク JGN2[4]を組み合わせて構築した。

2. StarBED

StarBED の正式名称は独立行政法人情報通信研究機構北陸 IT 研究開発支援センターである。512 台の PC およびそれらを結合するバックボーンネットワーク、さらにファイアウォール経由でのインターネット接続

環境などを持つ。これら 512 台の PC は、装備されたネットワーク I/F の違いなどから、いくつかのグループに分けられている。本研究ではこれらの中のうち A 群と呼ばれるグループの機材 208 台を利用した。この PC の性能を表 1 に示す。

表 1 StarBED A 群の PC のスペック

CPU	Pentium III 1GHz
メモリ	512Mbytes
NIC	1000BASE-T x 1, 100BASE-T x 2
OS	Windows 2000 Server Turbo Linux FreeBSD

表 1 に示したように、これらの PC の OS はマルチブートになっており、制御用端末から利用する OS を選択することができる。また、ブートに利用するディスクイメージをあらかじめ作成しておくことにより、StarBED が標準で用意しているのとは異なるシステムを用いて利用することも可能である。

またこれらの PC は、二つのネットワークで相互接続されている。一つは制御用のネットワークである。上述の制御用端末との通信などに利用される。このネットワークへは 100BASE-T の NIC を介して接続されている。もう一つのネットワークは実験用ネットワークである。これは VLANなどを用いて実験用仮想ネットワークを構築するためのもので、1000BASE-T の NIC を介して接続されている。

3. StarBED と JGN2 を用いた大規模ストリーミングサーバの設計

前章で述べた StarBED を用いて構築した大規模ストリーミングサーバについて述べる。

ストリーミングサーバの構築には以下の二つの方法がある。

- 一般的な PC を用いる方法
- 専用ストリーミング配信機器を用いる方法

前者は比較的安価に構築できる一方性能向上やメンテナンス性に問題がある。後者はこの反対の性格を持つ。

一般的に ISP でストリーミングサーバを構築する場合、配信性能とメンテナンス性を重視し、後者の方法を選択する場合が多い。しかし前述のようにこの方法で大規模ストリーミングサーバを構築するには大量に専用ストリーミング配信機器を用意する必要がある。CRN アーキテクチャに基づく配信の場合、各 ISP がそれぞれ機器を用意することでこの問題を回避可能である。しかし、本研究のように単一の大規模配信ストリーミングサーバを構築する際にはこの方法は現実的ではない。

そこで我々は一般的な PC を用いた大規模ストリーミングサーバの構築を検討した。この場合、性能面とメンテナンス面に問題が生じる。性能面に関しては、CPU 的な処理性能の問題と配信用ネットワークの帯域の二つの問題がある。

CPU 的な問題については、StarBED に注目した。StarBED を利用することで数百台の PC が利用可能になる。従って、投入する台数を増やすことにより解決を試みることにした。ネットワークの帯域の問題については JGN2 に注目した。JGN2 を利用すると 10Gbps のネットワークが利用可能になる。JGN2 の利用に関しては、いわゆるラストワンマイル、足回り回線敷設が問題となることが多い。しかし、StarBED は情報通信研究機構の設備と言うこともあり、この問題は解決可能という見込みがあった。これらより以下のような大規模ストリーミングサーバを設計・構築した。

3.1. 配信サーバ

配信サーバを設計するにあたり、対応プロトコル、サーバ用ソフトウェア、サーバの台数、負荷分散方式について検討を行った。これらについて以下で説明する。

3.1.1. 対応プロトコル

ストリーミングに利用されるプロトコルには Windows Media Technology(以下 WMT)、Real System、QuickTime などさまざまなものがある。本研究では、利用者の多さおよびライセンス形態の点から WMT にのみ対応することにした。

3.1.2. 配信サーバ用ソフトウェア

配信サーバ用の OS としては、Windows 2000 Server を選択した。これは WMT の配信用 OS としては、実質的に Windows 系 OS 以外に選択肢がないこと、および StarBED の PC がすでにライセンスを保持していたことが理由である。

WMT の配信用ソフトウェアの最新のものは Windows Media 9 Series である。配信用サーバには Windows Media サービス 9 を利用する。しかし、これは Windows 2003 Server 以降でないと利用できない。従って、今回構築するシステムでは Windows Media サービス 4.1 のサーバを利用した。

3.1.3. サーバの台数

ストリーミングの配信を行う場合に考慮しなければいけない事項の一つに、配信コンテンツの帯域がある。今回は、4. で述べるように LE2005a 日食中継において実証実験を行うことを前提に、300kbps, 150kbps, 45kbps の 3 種類のコンテンツを用意することにした。いわゆるブロードバンド利用者向け、ナローバンド利用者向け、その中間という位置づけである。

本節冒頭で述べたように、本研究では大規模配信を

実現するための処理能力を得るために、多数の PC を並列に利用するというアプローチを取った。このため、各コンテンツに何台の PC を利用するかを決定しなければならない。今回は 3.2 で述べるように 10Gbps のネットワークを利用することを前提に設計を行った。

PC の割り当て台数を決定するために、Windows Media Load Simulator を用いて、今回利用する StarBED A 群の PC の性能測定を行った。この結果、300kbps のコンテンツであれば 1 台あたり 300 ストリーム程度の配信が安定して運用できる限界であることが明らかになった。すなわち、1 台あたりの配信可能帯域は 90Mbps ということになる。

利用者からのアクセスの合計が配信用ネットワークの帯域である 10Gbps に達した場合とサーバの障害に備えるため、300kbps 用のサーバは 120 台と決定した。現在の日本のネットワーク利用者の傾向から判断して、ほとんどの利用者は 300kbps のコンテンツへアクセスしてくると予想した。従って、150kbps, 45kbps 用のサーバの台数は、それぞれ余裕を見て表 2 のように決定した。

表 2 各コンテンツ用配信サーバ台数

コンテンツ 帯域	サーバ 台数
300kbps	120
150kbps	20
45kbps	10

3.1.4. 負荷分散方式

複数の PC を用いた場合、それらの PC 間での負荷分散が必要となる。従来、CRN アーキテクチャに基づく配信においては、さまざまなポリシーに基づく負荷分散を行ってきた。例えば、経路情報を用いて最近傍のサーバを選択や、利用者が所属する AS 番号に基づく選択である。これらのポリシーは、配信サーバが広域に分散していることを前提としたものである。

しかし、今回は一箇所に集中している。従って、単純なラウンドロビンを用いることにした。ラウンドロビンを実現するために DNS を用いた。クライアントサイドにおける DNS キャッシュの影響を排除するために TTL を 0 に設定した。さらに DNS 返答では一つだけ IP アドレスを返すように設定した。このように DNS 返答に詳細な設定を行うために Tenbin [5] を利用した。

3.2. 配信ネットワーク

大規模なストリーミングサーバの構築においては、安定した配信を行うために二つのネットワークを用い

ることが一般的である。一つは、利用者がアクセスしてくる利用者用ネットワーク。もう一つは、配信する映像を各配信サーバに分配する映像分配用ネットワークである。

今回利用した StarBED の A 群の PC はすでに二つのネットワークを持っていました。そこで、広帯域が必要となる利用者用ネットワークに 1000BASE-T の NIC による実験用ネットワークを利用し、映像配信用ネットワークとして 100BASE-T の NIC による制御用ネットワークを用いた。

利用者用ネットワークのインターネットへの接続には、前述のように JGN2 を用いた。StarBED は、元々インターネット接続環境を有している。しかし、既存のネットワークは StarBED が立地する「いしかわサイエンスパーク」内の各施設が共同で利用するネットワークであるため、我々が実験で大量のトラフィックを発生させることは困難である。また、もしも可能であったとしても大規模配信実験に利用するには帯域が不足することが考えられた。そこで JGN2 を利用し 10Gbps の回線を設置した。

JGN2 からインターネットへの接続については、WIDE プロジェクトの協力を得た。WIDE プロジェクト北陸 NOC を経由し、DIX-IE および NSPIXP-3 へ接続した。

4. 配信実験

構築した大規模ストリーミングサーバを用いた配信実験を LIVE!ECLIPSE 2005 Annular (以下 LE2005a)において実施した。LE2005a は LIVE! UNIVERSE プロジェクト [6] が実施する日食中継実験である。

4.1. LIVE! ECLIPSE 2005 Annular

LIVE! UNIVERSE プロジェクトは、天文および宇宙科学などに関するさまざまな現象やイベントをネットワークを通じて広く世界に紹介し、社会に貢献することを目的とする任意団体である。2002 年の太平洋金環日食中継を皮切りに今回の LE2005a までに 10 回の中継を行ってきてている。また、前身の LIVE!ECLIPSE / Leonids 実行委員会の実績を加えると、その中継回数は 20 回を超える。

LE2005a は、2005 年 10 月 3 日の 16:40 (日本時間) から 19:23 (日本時間) にかけて、スペインなどで観測可能な金環日食を世界に向けて中継するものである。LIVE! UNIVERSE プロジェクトにおける日食中継では、観測可能な地点で撮影した映像を一旦日本に伝送し、そこから世界に向けて配信を行う。観測地によっては十分なネットワークが利用できないことが多い。そのような場合には、いわゆるブロードバンド用コンテンツの提供はできない。

しかし、今回の金環日食においては、スペインのマ

ドリッドという都市部での観測が可能であった。このため研究用ネットワークなどを利用し、観測地から日本まで高速回線を利用することができたため、ブロードバンド用の高画質コンテンツを利用者に対して提供することが可能であった[7]。

4.2. 映像配信ネットワーク

LE2005aにおける映像配信の流れは以下のようにになっている（図1）

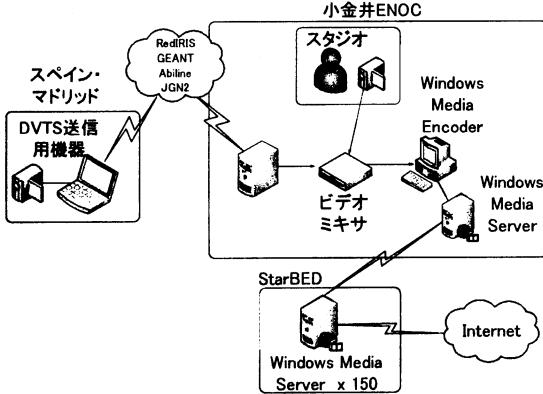


図1. LE2005a配信ネットワーク

1. マドリッドから小金井への伝送

観測地であるマドリッドで撮影された映像は、研究ネットワークなどを経由して、情報通信機構小金井オフィスに設置した Encoding NOC (以下 ENOC)に伝送される。この伝送には DVTSなどの高画質映像伝送システムを利用した。

2. 小金井でのエンコード

ENOCにはスタジオも設置され、マドリッドから伝送されてきた映像を素材として、ストリーミング中継を行う番組を作成する。この番組を入力として ENOCで WMT 形式にエンコードする。

3. 小金井から小松への伝送

小金井でエンコードしたデータは、本研究で構築したストリーミングサーバで配信を行うために、StarBEDが立地する小松へ伝送される。この伝送路としてもJGN2を用いた。

4. 小金井からの配信

小松では、ストリーミングデータ分配用の機材(NetCacheC1100)を用いて、150台の配信用サーバにデータを分配する。分配を受けた各配信用サーバは利用者からの要応じて、ストリーミングデータの配信を行なう（図2）。

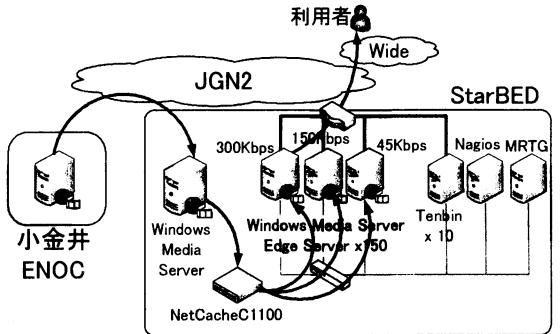


図2. StarBED内配信ネットワーク

また、本配信実験では運用支援用としてStarBED内に、以下のようなサーバを立ち上げた。

表3. 運用支援用サーバ

サーバ名	目的	OS	台数
MRTG	ネットワーク利用帯域監視用	Turbo Linux	1
Nagios	システム稼働状況監視用	Turbo Linux	1
FTP	ログデータ収集用	FreeBSD	1
ntp	時刻同期用	FreeBSD	2
Tenbin	負荷分散用	FreeBSD	10
配信サーバ予備	予備機	Windows 2000 server	10

サーバによりOSが異なっている。予備機に関しては、配信用サーバにトラブルが生じたときの予備であるので、本番機と同じWindows 2000 serverとした。それ以外についてはそれぞれのサーバ作業担当者が使い慣れたOSを選択した。従って、これらのOSを選択した技術的な理由は特にない。

4.3. 大規模ストリーミングサーバの構築

4.3.1. Windows 2000 Server サーバ構築

表1、表3から分かるように、本配信実験では、合計175台のPCを利用した。このうち予備機を含む配信用PC160台の構築について述べる。

StarBED A群のPCには、標準でWindows 2000 Serverがインストールされていたのは前述の通りである。しかし、標準のシステムにはWindows Mediaサーバは未インストールであった。そこですべてのPCにインストール作業を行う必要がある。StarBEDにおいては、その運用支援システムの一環として、PC群へのディスクイメージ配布機能がある。これは、あらかじめ作成しておいたディスクイメージを、各PCにダ

ウンロードし、そのディスクイメージを元に PC を起動するものである。

今回はこの機能を用いて、160 台の Windows Media サーバマシンを構築した。配信用のディスクイメージを構築するのは、比較的容易な作業であった。通常の Windows Media サーバマシンを構築するのと同様の手順で作業を行い、その後ディスクイメージ作成システムを起動するだけである。ただし、このディスクイメージの作成には十分な注意が必要である。今回は、ディスクイメージ配布作業後、各 PC の設定変更を行う必要が生じたため、160 台全部を手動で GUI から設定変更することになってしまった。なお、StarBED の PC はターミナルサーバに接続されているため、単一のターミナルから順次作業を行うことができる。

ディスクイメージの配信段階でもいくつかの問題が明らかになった。まず、システム上の都合で、同時にディスクイメージを配信可能な台数が 5 台に限られていた。本配信実験用に構築したディスクイメージの場合、1 回のイメージ配信に約 20 分掛かった。また、配信時に、各 PC をフロッピーディスクから起動する必要もあり、手間と時間が掛かる作業となつた。全 160 台にイメージを配信し終えるのに 10 時間以上が必要であった。

また、配信されたディスクイメージは、すべて同一である。これは、すべて同じホスト名で起動することを意味する。Windows OS は、起動時にホスト名が衝突していると問題が生じる。従って、ディスクイメージ配信後、1 台 1 台設定を変更する必要がある。この作業は、ディスクイメージ配信と並行して行うことはできるものの、自動化することができなかつた。従てこれも全 160 台手動で GUI から作業を行つた。

4.3.2. UNIX 系サーバ構築

表 3 に示すように、UNIX 系のサーバとしては、Tenbin 用に 10 台のサーバを構築した。これは、DNS の NS ラウンドロビンの挙動を調査するためである。LIVE! ECLIPSE 2001 Lunar における中継実験の際に、複数の DNS サーバを設置した場合、各サーバへの問い合わせの数にばらつきが発生することが明らかになつた。その原因が、ネットワーク環境や各サーバの性能の問題なのかが不明であった。そこで、本配信実験で、同一箇所に複数のサーバを配置することで、挙動を観察することにした。

UNIX 系のサーバの場合、コマンドラインから各種の操作が容易に行え、また ssh などのツールを使うことで、遠隔からの自動実行も容易である。従って、Tenbin 用のサーバ構築に関しては、台数は 10 台と比較的多かつたものの、特に問題は発生しなかつた。

その他の運用支援用サーバに関しては、台数が少な

かったこともあり、大規模ストリーミングサーバ構築に伴うような問題点はなかつた。

4.3.3. ネットワーク構築

StarBED 内の PC 群の利用者用ネットワークには /24 のサブネットを割り当てた。このサブネットは既設の 1000BASE-T で BigIronMG8 につながっている。グローバル IP アドレスを割り振りインターネットからアクセスを受け付け、ストリーミングのサービスを行う。映像分配用ネットワークとしては、IP アドレスなども含めて既存の制御用ネットワークをそのまま利用した。

BigIronMG8 は JGN2 を通じて WIDE 北陸 NOC に接続し、最大 10Gbps のトライフィックを配信可能とした。

5. 評価

5.1. システム構築

システムの構築に関しては 4.3.1. で述べたように、Windows 2000 Server を用いた配信サーバの構築において、さまざまな問題が発生した。これらの原因は以下のようなものがある

1. Windows 系サーバが GUI を使わずに設定変更する自由度が低いこと
2. 操作の自動化が難しいこと
3. ディスクイメージ配信機能に制限があること

以上のうち 1, 2 については Windows 系 OS 一般的の問題と考える。ただし、専用のツールなどを開発することなどで、作業量の軽減を行える可能性がある。2. については、ディスクイメージ配信サーバのディスク I/O 性能の向上により、同時配信台数の向上が計れるのではないかと考える。また、ディスクイメージ配信時に Windows のホスト名を変更する機能があるツールを利用することで、手間を減らすことも可能である。

5.2. システム運用

CRN アーキテクチャに代表される広域分散型の CDN の場合、運用対象となる機材が広域に分散することになる。これは、運用にあたる人員も多数必要ということになる。今回の配信実験では、配信サーバ担当者 1 名、ネットワーク担当者 1 名、システム監視担当者 1 名、負荷分散担当者 1 名、諸作業担当者 2 名の 6 名でシステムの運用を行うことができた。人的な面からは非常に有効であったと結論付けることができる。

また、システムの管理においても、配信システム全体の構成がシンプルになる。これも運用上は大きなメリットである。

一方で、集中型のシステムによる問題点もある。すなわち、本大規模ストリーミングサーバが、single point of failure となることである。今回の配信実験では、もしもの場合に備えて、ENOC (小金井) から配信する準備も行った。しかし、ENOC には十分な対外接続ネット

トワークを用意できなかった。

また、今回の配信実験用システムを構築している最中に、StarBED システム中のコアスイッチが異常発熱するという障害が発生した。今回は、強制的に外部から空冷することにより、問題を回避することができた。この問題については予備機材を準備することで回避可能である。しかし、今回の配信実験で用いた 10G のネットワークのように、利用にあたって高価な機材を用いる場合、予備機の準備が困難な場合もある。

5.3. 配信実績

今回の日食中継におけるストリーミングサーバへのアクセスを図 3 に示す

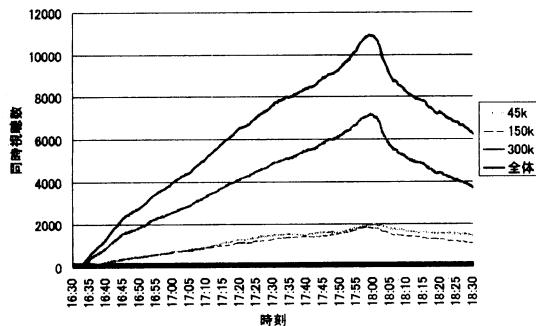


図 3. ストリームサーバへのアクセス

アクセス数の状況を表 4 にまとめる。

表 4. 最大アクセス数

コンテンツの帯域	最大同時視聴数	最大視聴時刻
全体	10944	17:58:52
300k	7165	17:59:06
150k	1872	17:58:04
45k	1987	18:00:18

今回の中継では、最大の見所である金環食が 17:55～18:00 であったため、そのころ最大視聴数を数えたことになる。

配信における最大送信帯域は、約 2Gbps くらいであった。これは期待していたアクセスに比較すると必ずしも多くはない。しかし、1Gbps のネットワークでは不足であり 10Gbps が活かされたと言える。結果論ではあるが、より高画質なコンテンツを用意するべきであったかもしれない。

6. おわりに

本研究では、StarBED および JGN2 を用いて単一の大規模ストリーミングサーバを構築した。本ストリーミングサーバを用いて、LE2005a 日食中継において配信実験を行った。

この配信実験を通して、本研究で構築した大規模ス

トリーミングサーバによる単一サイトからの大規模配信が可能であることを実証できた。

StarBED や JGN2 のように大規模なシステムやネットワークを持つ研究施設は、さまざまな研究開発に大変有意義であると考える。

一方で、大規模ストリーミングサーバの問題点も明らかになった。構築においては、多数の機材を用いることに伴う作業量や Windows 系 OS による GUI 操作を強制されること。運用においては single point of failure になることである。

一方で、集中型のサーバであったことから、従来の分散型のシステムと比較すると、運用は容易であったと言える。特に人的な面では非常に大きなメリットがあった。

ネットワークに目を向けた場合、今回構築したサーバは JGN2 (あるいは WIDE) のシングルホームであった。これは従来の CRN アーキテクチャに基づく配信に比べると脆弱と言える。

今後の課題としては、サーバ構築手法の改善、およびネットワークの多重化がある。後者に関しては、5.2 で述べたような大規模ストリーミングサーバを構築する際の内部のネットワークの多重化と、上述の配信用ネットワークのマルチホーム化を検討したい。

謝辞

本研究を進めるにあたって、WIDE プロジェクトの皆さんには多大なる協力を頂いた。本実験を行う機会を与えてくださった LIVE! UNIVERSE プロジェクトの皆さんに感謝する。ハイホーブロードバンドシステムズの西川由佳梨氏、NTTスマートコネクトの遠藤論氏、九州産業大学の井上宗徳氏、坂田充氏、藤村英樹氏には配信実験に多大なる貢献をしていただいた。IIJ の山本文治氏からはさまざまな有意義なアドバイスを頂いた。埼玉大学の吉田紀彦先生には研究推進にあたってご支援を頂いた。

文 献

- [1] CRN フォーラム, <http://www.crnf.net>.
- [2] “実ノードを利用したネットワークシミュレーションにおけるノードへの OS の導入及びパラメータ設定機構の開発,” 三角 真、宮地 利幸、知念 賢一、篠田 陽一、情報処理学会研究報告書, DPS-116, p95-100, ISSN:0919-6072, Jan 2004
- [3] StarBED Project, <http://www.star-bed.net/>
- [4] JGN2 Home Page, <http://www.jgn.nict.go.jp/>
- [5] “多様な選択ポリシーを利用可能なサーバ選択機構,” 下川俊彦、吉田紀彦、牛島和夫、電子情報通信学会論文誌, J84-D-I:9, 1396-1403 Sep 2001
- [6] LIVE! UNIVERSE, <http://www.live-universe.org/>
- [7] “LE2005a の概要とマドリッドからの高速回線による高画質映像伝送,” 山本文治、中村一彦、尾久士正己、電子情報通信学会技術研究報告, IA-印刷中, Jan 2006