

デジタルビデオを用いたリアルタイムマルチメディア通信システム

中内清秀*，金子晋丈**，齊藤 昭***（東京大学大学院工学系研究科）
森川博之****（東京大学大学院新領域創成科学研究科）

*nakauchi@mlab.t.u.-tokyo.ac.jp **kaneko@mlab.t.u.-tokyo.ac.jp ***saito@mlab.t.u.-tokyo.ac.jp
****mori@mlab.t.u.-tokyo.ac.jp

インターネットの広帯域化や端末の高性能化が加速的に進み、さまざまなメディアがインターネット上で配信されるようになった。また、広帯域を必要とする高品質な映像や音声のリアルタイム配信に対する要求が高まっている。同時に、映像や音声をリアルタイムに多地点に対して配信に対する要求も急速に増加している。そこで本稿では、高品質映像、音声を伝送するためのデータフォーマットとして、広く一般に普及しているDV（Digital Video）を取り上げ、DVをIPネットワーク上で配信するリアルタイムマルチメディア配信システムについて、その要素技術とともに解説する。さらに本システムの実ネットワーク環境における運用例を示す。

はじめに

次世代インターネットに向けて、IPv6技術、高速広帯域で低遅延な転送、IP上での多様なメディアの転送、IPマルチキャストなどの新しい技術の研究開発が活発に行われ、多様なアプリケーションに対する期待が高まっている。その中でも特に、IPネットワークで高品質な映像や音声をリアルタイムで多地点に配信するアプリケーションが注目されている。

バックボーンネットワークにおけるWDM（Wavelength Division Multiplexing）技術やローカルネットワークにおけるギガビットイーサネット技術などの進歩によるネットワークの広帯域化、MPLS（Multi-Protocol Label Switching）技術などによるネットワークの高速化、IPマルチキャストによる多地点配信の効率化、さらに端末のメモリ容量や処理能力の大幅な向上などの背景のもと、このようなリアルタイムマルチメディア配信システムが比較的安価に構築できるようになってきた。

そこで本稿では、高品質メディアであるDVをインターネットを介して配信するリアルタイムマルチメディア配信システムについて解説する。まず、要素技術としてDV over IP、IPv6、IPマルチキャストを取り上げ、それぞれ簡潔に解説する。さらに配信した複数メディアのデジタルデータを統合的に蓄積、再生する高機能アーカイブシステムの要素技術であるSMIL（Synchronized Multimedia Integration Language）についても解説する。次に、リアルタイムマルチメディア配信システムの概要と構成について述べる。最後に、リアルタイムマルチメ

ディア配信システムの運用例として、WIDE GIOSプロジェクトにおける国内での双方向多地点遠隔講義と東京大学大学院情報理工学系研究科設立シンポジウムにおける海外との双方向中継について述べる。

要素技術

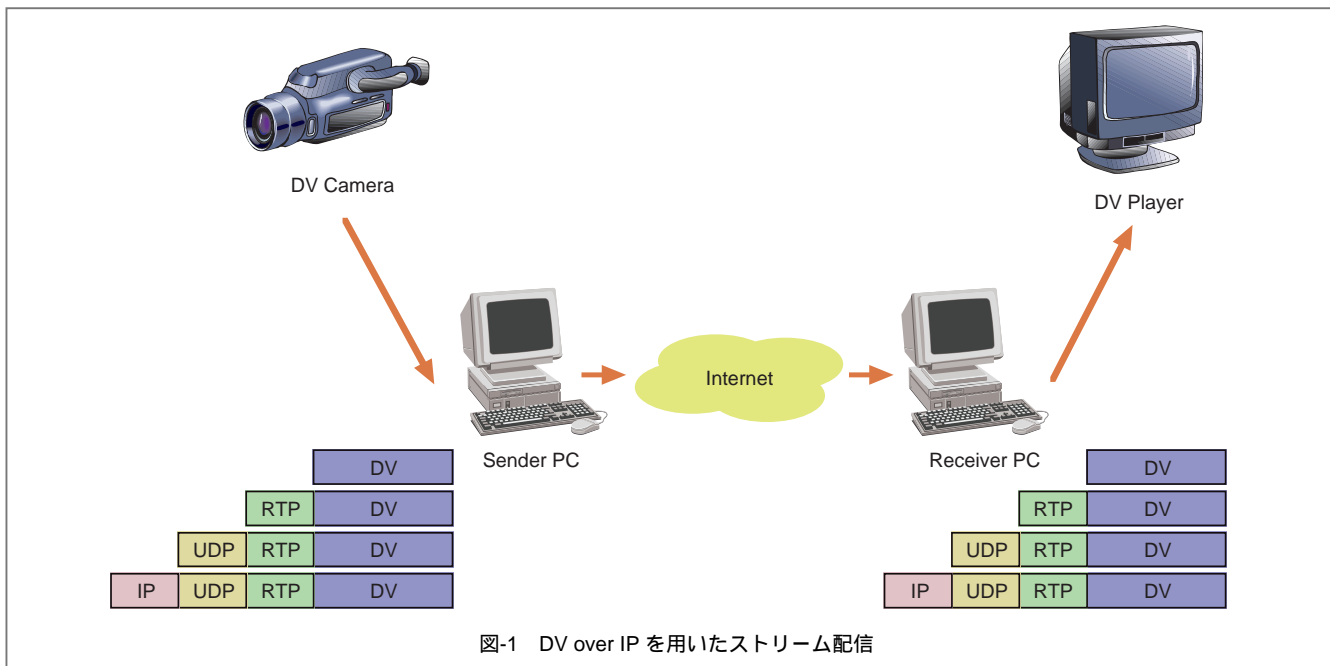
リアルタイムマルチメディア配信システムでは、リアルタイムに高品質映像や音声を安価に伝送するため、現在民生用ビデオのフォーマットとして広く普及しているデジタルビデオ（DV）を採用している。DVの伝送には、次世代インターネットプロトコルであるIPv6を用い、DV送受信アプリケーションとしてDVTS¹⁾を採用している。多地点伝送する場合にはIPv6 Multicastを用いる。さらに、アーカイブ化された複数のメディアの再生を同期させるためにSMILを用いる。

IPによるDVの伝送

現在、IEEE1394インタフェースを持つ民生用DV機器を用い、通信媒体としてインターネットを利用することにより高品質な映像と音声の中継を低価格で実現できるようになった。

DVフォーマットの基本単位は80バイトのDIFブロックで、これを多数並べたDIFシーケンスを複数用いてDV映像の1フレームを構成する。音声もDIFフレームに符号化され、対応するビデオと同一のフレーム内で伝送される。

DVフォーマットではMPEGなどと異なりフレーム間圧縮は行わず、DCT（Discrete Cosine Transform）とRun Length



Encodingを用いてフレーム内圧縮のみを行っている。そのため、帯域が不足する場合、映像フレームを間引くことで使用帯域を減少させることができる。この場合でも、音声品質は保持されるよう制御される。間引きせずにフルレートでDVを伝送する場合、伝送レートは約35Mbpsである。

民生用DV機器を用いてインターネット上で映像と音声を送るためには、民生用DV機器からDV信号を受信し、IEEE1394インタフェースからDVパケットとしてインターネットへ送出する機能、およびインターネットからDVパケットを受信し、DV信号として民生用DV機器へ出力する機能を具備したアプリケーションが必要となる。

このようなアプリケーションとしてDVTS (Digital Video Transmission System)¹⁾が開発されている。DVTSを用いたDV送受信の様子を図-1に示す。DVTSは以下のような特徴を持つ。

- RTP (Real-time Transport Protocol) を用いたリアルタイム伝送
- RTCP (RTP Control Protocol) を用いたパケットロスの監視や動的なフレームレートの調整
- 次世代インターネットプロトコルIPv6への対応
- IP マルチキャストへの対応

DVTSではdvsendというプログラムにより、IEEE1394インタフェースから読み込んだDVパケットにRTPヘッダを付与し、UDPでカプセル化してインターネットに送出する。DVTSではUDPパケットに含まれるDIFブロックの数を自由に設定できるため、レート制御が容易に実現できる。そして、dvrecvというプログラムにより、インターネットから受信したUDPパケットからDV信号を取り出し、IEEE1394インタフェースへ出力する。また、dviewというプログラムにより、受信したDV映像をXウ

ィンドウシステム上で表示することもできる。

IPv6

IPv6は、拡大するネットワーク環境への対応、IPv4からの移行の容易性、セキュリティ保証、ネットワーク設定の容易性、モバイルホストへの対応、ルーティングポリシーの簡略化を目的として、1995年にIETF (Internet Engineering Task Force) のIPNG WGによって基本仕様が策定された²⁾。IPv4と大きく異なるIPv6の特徴として、豊富なIPアドレス空間、経路情報の集約、およびネットワーク層におけるセキュリティ機能が挙げられる。

豊富なIPアドレス空間

IPv4アドレスは32ビットで記述されるため、10進数に直すと約43億のアドレス空間となる。これは世界人口の61億よりも少なく、あらゆるデバイスにIPアドレスが付き、インターネットが世界的な通信基盤になるにはアドレス数が絶対的に不足する。しかし、IPv6ではアドレスは128ビットで記述されるため、約340潤のアドレス空間となる。

経路情報の集約

一般に、グローバルアドレスが割り当てられるネットワークの増加に伴ってバックボーンルータの経路情報が増加し、ルータに搭載されるメモリの増加や経路情報の検索速度の低下を招く。そこで、広大なアドレス空間を持つIPv6のアドレス割当てはIPv4と異なりネットワークの論理トポロジーと密接な関係を持ったプロバイダ型アドレス割当てを採用しており、世界的に統一された階層構造となっている。

ネットワーク層におけるセキュリティ機能

一般に、ネットワークのセキュリティを強化するためにはファイアウォールが導入され、ファイアウォールを経由してインターネットにアクセスするという手

法がとられている。しかし、ファイアウォールの導入はネットワークの透過性を阻害し、インターネットの基本原則であるエンド・ツー・エンド議論に反することになる。そこでIPv6では、ネットワーク層でのセキュリティ機能が導入された。具体的にはIPレベルでの認証と暗号化に関する枠組みが規定されている。IPv6のセキュリティ機能は、特にファイアウォール間、モバイルホストとその無線基地局間において、潜在的需要があると考えられている。

IP マルチキャストによる広域展開

多地点遠隔講義やイベントの多地点中継などでは、同一のデータを同時に複数のホストに配信する必要がある。IPマルチキャストはこの多地点通信を効率的に実現する技術である³⁾。

マルチキャストの基本コンポーネントは、グループ管理プロトコルとマルチキャストルーティングプロトコルの2つである。グループ管理はIGMP (Internet Group Management Protocol) によって行われる。IGMPにより、ルータは自身に接続されるLAN上にマルチキャストグループに参加するホストが存在するかどうかを確認する。また、マルチキャストルーティングプロトコルにより、パケットを配信するためのマルチキャストツリーが形成される。

マルチキャストルーティングプロトコルはデンスモードとスパースモードの2種類に分類される。デンスモード・ルーティングでは、マルチキャストツリーとして一般に送信者を根とした最短木が形成される。一方、スパースモード・ルーティングでは、コアまたはランデブーポイントと呼ばれるルータを中心とした共有木が形成される。

スケーラビリティの観点では、デンスモードよりもスパースモードの方が優れているといわれている。第一の理由は、マルチキャストツリーの形成過程において、前者が定期的にフラッディングを行い、結果としてネットワークに大きな負荷をかけるのに対し、後者はフラッディングを行わず、受信者からの要求があって初めてマルチキャストツリーの枝を形成することである。第二の理由は、ルーティングテーブルとして、前者が(S,G)のエントリーを持つのに対し、後者は(*,G)のエントリーだけを持つため、ルーティングテーブルがグループ単位に集約できることである。ここで、Sは送信者のIPアドレスを、Gはマルチキャストアドレスをそれぞれ表し、(*,G)とは、送信者に関係なくマルチキャストグループ単位にエントリーを1つだけ持つ、という意味である。

35Mbpsもの帯域を必要とするDVを伝送するリアルタイムマルチメディア通信システムにおいては、特にス

```
<smil>
<head>
<meta name="title" content="Class 20000002-01"/>
<layout>
<root-layout width="640" height="240"/>
<region id="pix" top="0" left="0" width="320" height="240"/>
<region id="main" top="0" left="320" width="320" height="240"/>
</layout>
</head>
<body>
<par>
<seq>
<video src="rtsp://stream.soi.wide.ad.jp/soi/class/20000002/video/01/01.rm"
region="main" clip-begin="01:08:45" end="1025s"/>
</seq>

</par>
</body>
</smil>
```

図-2 SMIL の記述例

ケーラビリティを重視し、代表的なスパースモード・ルーティングプロトコルであるPIM-SM (Protocol Independent Multicast - Sparse Mode) を採用した。PIM-SMでは、マルチキャストセッションの開始時に形成されるマルチキャストツリーはランデブーポイントを中心とした共有木であるが、トラフィックが増加すれば、より効率的な最短木に移行する。

SMILによるメディア同期

SMIL (Synchronized Multimedia Integration Language) とは、W3C (World Wide Web Consortium) によって勧告されたマルチメディアプレゼンテーション記述言語であり、1セットの独立したマルチメディアオブジェクトを1つの同期マルチメディア表現に統合することを可能にする⁴⁾。SMIL を利用することにより、製作者は以下のようなプレゼンテーションを行うことができる。

- 表現の時間的な挙動を記述する
- スクリーン上のレイアウトを記述する
- ハイパーリンクをマルチメディアオブジェクトと結び付ける

SMILとHTMLを比較すると、HTMLだけではオブジェクトの空間的な記述しかできないのに対し、SMILではオブジェクトの空間的配置に加えて時間的配置が可能になる。図-2にSMILの記述例を示す。また、代表的なタグについて以下に述べる。

- <root-layout>: コンテンツ全体の表示サイズの設定
- <region>: オブジェクトの表示位置の設定
- <par>: 同時に (並列に) 表示するオブジェクトの設定
- <seq>: 順に (直列に) 表示するオブジェクトの設定

図-2のSMIL文書を再生した時のスクリーンショットを図-3に示す。図-3では、講義の映像(01.rm)と複数のスライドから構成される講義資料の静止画(01.rp)の2つのオブジェクトが同時に01:08:45の部分から再生されている。

RealNetworks社のRealPlayerやApple社のQuickTimeなどの再



図-3 SMIL の表示例



図-4 リアルタイムマルチメディア通信システムの外観

	ラックマウント型PC	ノート型PC
CPU	Intel PentiumIII 600MHz	Intel PentiumIII 600MHz
メモリ	256MB	128MB
IEEE1394カード	PCI Lynx IEEE1394 Card	PCI OHCI IEEE1394 Card
NIC	Intel 10/100Base-TX	Realtek 10/100Base-TX
OS	FreeBSD 4.1	FreeBSD 4.1

表-1 DV送受信用PCの仕様

生アプリケーションがSMILに対応しており、今後ますます普及することが予想される。

リアルタイムマルチメディア通信システム

リアルタイムマルチメディア通信システムの主な機能は、音声および映像のフォーマット変換とDV信号のインターネットを介した送受信である。民生用DVカメラから取り込まれたアナログ映像信号（以下NTSC信号）およびマイクから取り込まれたアナログ音声信号は、1つのDV信号に混合／変換される。さらに、DV信号は送信用PCによりIPパケットに変換され、インターネットを介して他のサイトに送信される。一方、他のサイトからインターネットを介して受信されたIPパケットは受信用PCによりDV信号に変換される。さらにDV信号は、NTSC信号およびアナログ音声信号に変換／分離され、映像はプロジェクトを通じて会場のスクリーンに、音声は会場のスピーカにそれぞれ出力される。

本システムはIPマルチキャストにも対応しており、最大3サイトで同時に双方向通信を行うことができる。また、DV受信用PCは遠隔サイトの数だけ用意すればよいので、基本的な構成を変更することなく、より多くのサイト間での双方向通信に容易に拡張可能である。

システムの構成

本システムは、DV送信用PC1台、DV受信用PC2台、DVプレーヤ3台、DVコンバータ1台、映像ミキサ、音声ミ

キサ、4画面分割装置、および管理用モニタ2台から構成される。DVプレーヤはDV映像のモニタリング、および、DV信号からNTSC信号／アナログ音声信号への変換を行う。DVコンバータはNTSC信号／アナログ音声信号からDV信号への変換を行う。映像ミキサと音声ミキサは、それぞれ複数の系統から入力される映像、音声を合成して同一系統に出力する装置である。4画面分割装置は、4つの異なる系統から入力される映像を合成し、各入力を同時に1つのモニタに出力する装置である。本システムの外観を図-4に示す。図-4では、1対1の双方向DV送受信を行っているため、映像管理用モニタには送信側サイト、受信側サイトの2つのサイトの映像だけが映っている。

次に、本システム内の配線を図-5に示す。送信側のNTSC信号およびアナログ音声信号はDVコンバータでDV変換されるとともに、DVプレーヤでモニタリングされる。一方、受信側のDV信号はDVプレーヤでモニタリングされるとともに、NTSC信号およびアナログ音声信号に変換される。

DV送受信用PCを所有しないサイトとの双方向DV送受信を行う場合は、DV送受信用PCとして携帯用のノート型PC2台をそのサイトに持ち込み、双方向通信を実現している。本システムに含まれるDV送受信用ラックマウント型PCとノート型PCの仕様を表-1に示す。IPv6スタックとしてはKAME⁵⁾を採用した。

本システムは、映像系統と音声系統が完全に分離された構成となっている。それぞれの系統について以下

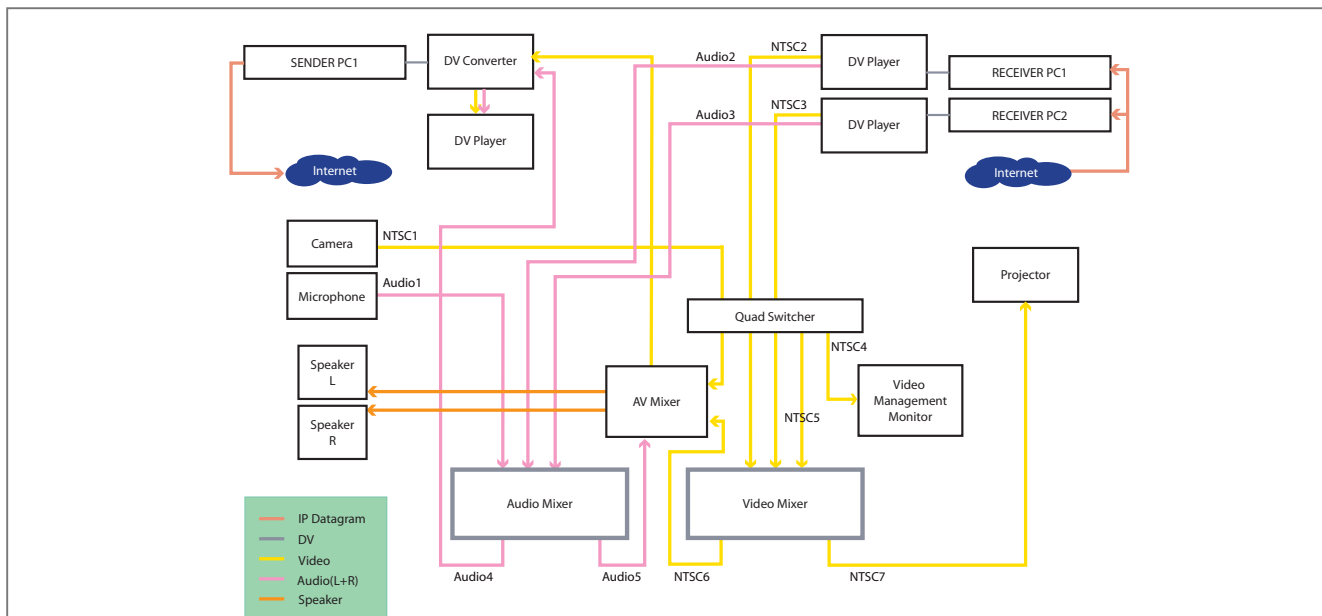


図-5 リアルタイムマルチメディア通信システムの配線図

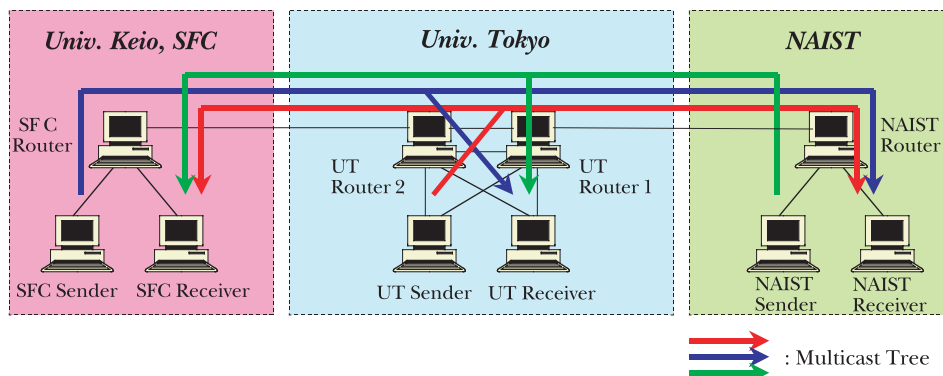


図-6 多地点双方向遠隔講義のネットワークポロジ

に解説する。

映像系統

図-5に示されるように、会場カメラから出力されるNTSC信号（NTSC1）、および受信側の2台のDVプレーヤから出力される2系統のNTSC信号（NTSC2,3）は、入力系統を4つ持つ4画面分割装置（Quad Switcher）に入力される。4画面分割装置には入力信号をスルーさせる機能と、入力信号を1つの画面に表示できるように合成する機能があり、入力された3系統のNTSC信号（NTSC1,2,3）、その3系統の信号から合成された1系統のNTSC信号（NTSC4）、そしてこの4系統から選択された1系統のNTSC信号（NTSC5）の合計5系統の信号を出力する。NTSC4は映像管理用モニタに入力され、講演者が会場カメラの映像および他のサイトの映像を同時にモニタリングするために使用される。また、NTSC1は送信用映像切替装置（AV Mixer）に入力される。

4画面分割装置から出力されるNTSC2,3,5の計3系統は、映像ミキサ（Video Mixer）に入力される。映像ミキサは、入力された3系統の中から選択された1系統、もしくは選択された2系統から合成された系統（NTSC6）を送信用映像切替装置に出力し、同時に、同一の方法で選択

（合成）したNTSC信号（NTSC7）を会場内のプロジェクタに出力する。送信用映像切替装置は入力される2系統（NTSC1,6）の中から1系統を選択し、送信側にあるDVコンバータに出力する。

音声系統

図-5に示されるように、会場マイクから出力されるアナログ音声信号（Audio1）、および受信側の2台のDVプレーヤから出力される2系統のアナログ音声信号（Audio2,3）は音声ミキサ（Audio Mixer）に入力される。音声ミキサはメイン出力とそれ以外の出力の2つの出力系統を持つ。音声ミキサは、入力された3系統（Audio1,2,3）の中から1系統を選択、または複数選択して合成した音声信号（Audio4）を、送信側にあるDVコンバータに出力する。同時に、同様の方法で生成した音声信号（Audio5）を会場内スピーカに接続されるアンプ（AV Mixer）に出力する。

音声系統において最も重要なことは、エコーバックの防止である。エコーバックへの対策を施していないシステムでは、受信した音声と会場内のマイクが同一の音声ミキサに入力されるため、他のサイトの音声を送り返してしまう。この場合、他のサイトでは音声

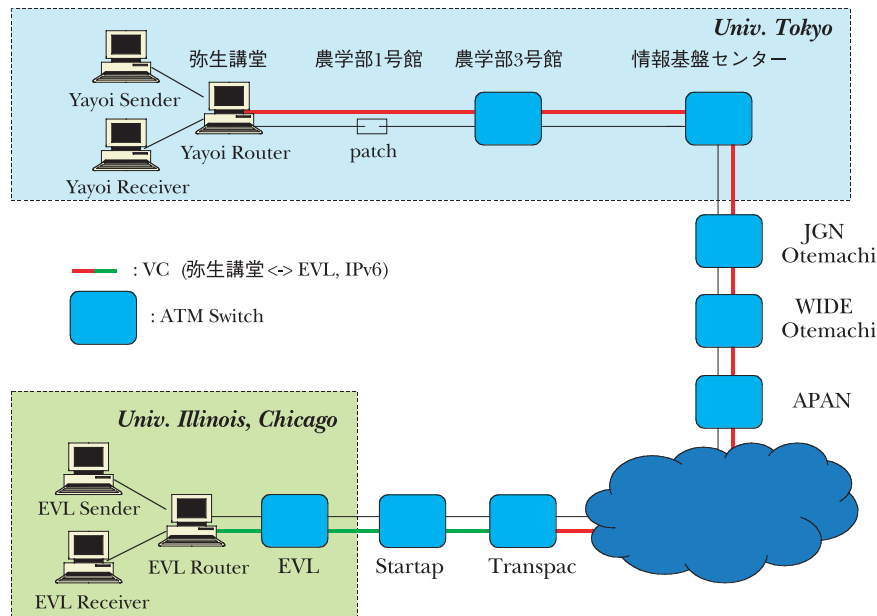


図-7 日米遠隔講演のネットワークトポロジー

エコーし、音声が非常に聞きづらい環境となる。エコーキャンセラを用いてエコーバックを回避する手法もあるが、通信遅延が非常に大きい地点間での配信には対応できない。そこで本システムでは、音声ミキサにおいて、送信される音声信号 (Audio4) に受信した音声信号 (Audio2,3) が流れ込まないように調整している。

リアルタイムマルチメディア通信システムの運用例

WIDE GIOSプロジェクト⁵⁾の一環として、2000年冬学期に、東京大学、慶應義塾大学、および奈良先端科学技術大学院大学の3地点を結んで双方向遠隔講義を行った。全14回の講義を3大学が交代で行い、すべての講義においてそれぞれの大学から他の2つの大学へのDV伝送にIPv6マルチキャストが用いられた (DV over IPv6 Multicast)。

多地点双方向遠隔授業に利用したネットワークのトポロジーを図-6に示す。この時、マルチキャストルーティングプロトコルとしてはPIM-SMを用いた。図-6には、

各サイトを送信者とした時のマルチキャストツリーもそれぞれ示されている。UT Router1をランデブーポイントと設定したためセッション開始直後はUT Router1を中心とした1つの共有木が形成されたが、PIM-SMでは共有木は最短木に移行するため、図-6に示すような最短木が3つ形成された。

次に2001年3月2日に行われた東京大学大学院情報理工学系研究科の設立シンポジウムにおいて、米国イリノイ大学のThomas Defanti教授の講演を、米国イリノイ州シカゴにあるイリノイ大学EVL (Electronic Visualization Laboratory) からユニキャスト中継した。シンポジウムの行われた東京大学弥生講堂の映像もEVLへユニキャスト中継され、情報理工学研究科の青山友紀教授とThomas Defanti教授との間で議論が行われた。

日米間遠隔講演に利用したネットワークのトポロジーを図-7に示す。DVの伝送にはIPv6を用いた (DV over IPv6)。この時の会場の様子を図-8に示す。

おわりに

本稿ではDV over IP技術を中心としたリアルタイムマルチメディア通信システムについて、その基盤技術を解説した。さらに、本システムの構成、本システムにおける映像系統、音声系統の詳細を解説するとともに、本システムの運用例を示した。

参考文献

- 1) DVTS Project: <http://www.sfc.wide.ad.jp/DVTS/>
- 2) IETF IPNG WG: <http://www.ietf.org/html.charters/ipngwg-charter.html>
- 3) Deering, S.: Multicast Routing in Internetworks and Extended LANs, Proc. ACM SIGCOMM'88 (Aug. 1988).
- 4) Synchronized Multimedia Integration Language (SMIL) 1.0 Specification, W3C Recommendation 15-June-1998 (<http://www.w3.org/TR/REC-smil/>).
- 5) KAME Project: <http://www.kame.net/>
- 6) WIDE 大学 School of Internet: <http://www.soi.wide.ad.jp/>

(平成13年6月5日受付)



図-8 日米遠隔講演の様子