

ストリーミング技術

櫻井 智明 ((株)キールネットワークス)

山本 文治 ((株)アイアイジェイメディアコミュニケーションズ)

1999年12月15日

Internet Week 99 パシフィコ横浜

(社)日本ネットワークインフォメーションセンター編

この著作物は、Internet Week 99 における櫻井 智明氏と山本 文治氏の講演をもとに当センターが編集を行った文書です。この文書の著作権は、櫻井 智明氏、山本 文治氏および当センターに帰属しており、当センターの同意なく、この著作物を私的利用の範囲を超えて複製・使用することを禁止します。

©1999 Tomoaki Sakurai, Bunji Yamamoto,
Japan Network Information Center

目次

1	概要	1
2	ストリーミングメディアとは	1
3	ストリーミングシステムでの伝送技術	5
4	映像や音声の品質と圧縮技術	8
5	事例紹介	12
6	大規模配信に対する検討	16
7	著作権に対する検討	17
8	まとめ	17

1 概要

この講演では、プロトコルや音声・画像圧縮技術も含めながら、ストリーミングシステムのしくみを説明していきます。また、ストリーミングシステムを使った事例や、利用のためのノウハウも紹介していきます。

なお、ここでは特定製品で利用できる技術のみではなく、広範囲に渡って利用されている技術を取り上げていきます。

2 ストリーミングメディアとは

現在、インターネットを利用するメディアとしては電子メールと WWW が代表的です。では、この講演で取り上げるストリーミングメディアは、インターネットの中でどのような位置付けなのでしょう。

既存メディアとインターネットを利用したメディアを対比してみると、表 1 のようになります。

表 1：既存メディアとインターネットを利用したメディアの対比

既存メディア	インターネットを利用したメディア
口コミ	電子メール
新聞	WWW
ラジオ、テレビ	ストリーミングメディア

「電子メール」は、電話、ポケットベル、メーリングリスト等も含めてコミュニティを形成するものであることから、既存の「口コミ」と同様のメディアであると考えられます。また、「WWW」はあらかじめ記録された情報を検索し閲覧する形のもので、「新聞」と同様の、情報を一方向に流すメディアと考えられます。

これに対して、「ラジオ」や「テレビ」では表現形態が多様化し、情報を大量消費します。そして、「ストリーミングメディア」もインターネットを使ってラジオやテレビと同様の多彩な表現形態を利用したいという要求から生み出されたものと考えられます。

2.1 圧縮技術の必要性

実際にコンテンツを発信、受信するためには、そのコンテンツが持つ情報量を考慮しなければなりません。たとえば、いくつかのマルチメディア情報では、表 2 に示す帯域が要求されます。

表 2：各マルチメディア情報の要求帯域

分類	種類	要求帯域
オーディオ	電話音声	64kbps
	CD(圧縮なし)	1.4Mbps
動画	MPEG-1	1.5Mbps
	MPEG-2	4 ~ 9Mbps
	NTSC	220Mbps

インターネットで実際に利用できる回線帯域は、1994 年頃に ISP が登場して、個人向けのダイヤルアップ IP サービスが開始された頃は、14.4/28.8kbps のモデムや 19.2kbps の TA によるものが個人利用の限界でした。1995 年には RealAudio 1.0 や StreamWorks 1.0 が提供されてストリーミングメディアが利用できるようになりました。現在では、33.6/56kbps のモデムや ISDN を使ったダイヤルアップルータが普及しています。MP による ISDN 回線の 2B 接続 (128kbps) や SOHO 向けの常時接続サービスが提供されるようになってきています。

回線ごとに利用可能な帯域を、表 3 に示します。

表 3：回線ごとの帯域

回線	利用可能な帯域
アナログ電話	300bps ~ 56kbps
ISDN	64kbps、128kbps
T1、フレームリレー	~ 1.54Mbps
LAN (Ethernet)	10Mbps
Fast Ethernet	100Mbps
光ファイバ	155Mbps、622Mbps
ギガビットネットワーク	1Gbps

以上から、現在では、十分な容量のネットワークがあれば、マルチメディア情報をそのまま伝達できるようになったことがわかります。

ただし、インターネットでは、ユーザのアクセス回線だけでなく、バックボーンでもマルチメディア情報をそのまま伝達できるほどの容量を十分に確保することが難しいという問題があります。このため、情報を圧縮して伝送データ量を削減する必要があります。

2.2 ストリーミングシステムの動作

図1に、実際のストリーミングシステムの処理の流れを示します。

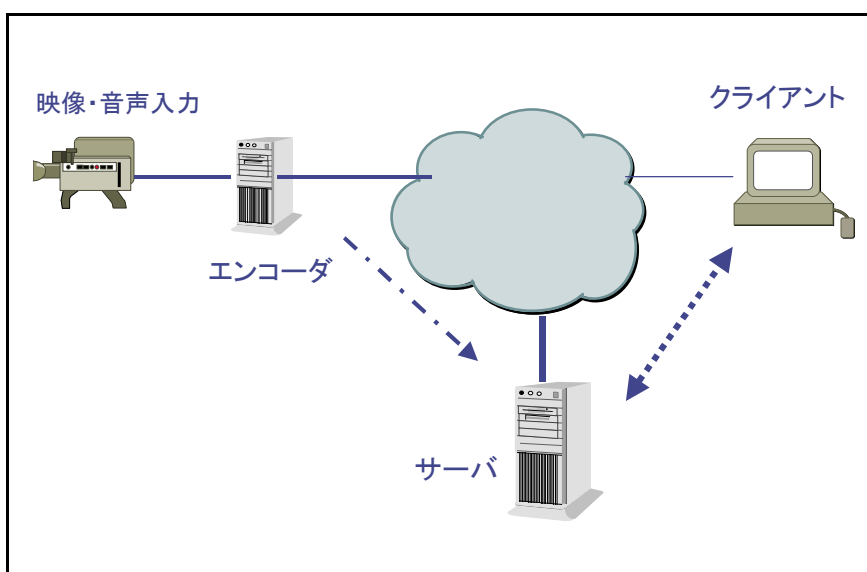


図1：ストリーミングシステムの処理順序

ストリーミングシステムは、その方式から次の2種類に大別されます。

- ライブ
- オンデマンド

ライブでの処理は、次のようなものとなります。

1. 映像や音声をアナログデータとして収録する。
2. エンコーダによって、アナログデータをデジタルデータにエンコードする。
3. エンコードしたデジタルデータを圧縮した後、パケット化する。
4. パケット化したデータをサーバに中継する。
5. サーバは各クライアントの状態に応じたデータ量のパケットを送る。
6. クライアントは受け取ったデータをバッファに蓄え、データをデコードして再生する。

これに対してオンデマンドでは、エンコードしたデジタルデータを圧縮してファイルとした後に、そのファイル全体をサーバに蓄えます。また、オンデマンドでは、エンコーダから流れてくるパケットの代わりに、サーバに蓄積されているファイルをパケットにしてクライアントに送ります。

このようなストリーミングシステムでは、エンコーダによるデジタル化や圧縮の処理、インターネット上での配信状況やクライアントでのバッファリング処理等の各状態によって、データ再生までの時間が変動します。

インターネットユーザの回線はモデムから専用線、LAN までさまざまな環境が考えられます。これまではそれぞれの帯域に合わせてファイルを別々に作成したり、エンコーダを複数台用意したりしてきました。最近では 1 台のエンコーダや 1 つのファイルで複数の帯域をサポートするマルチエンコードが利用できるようになってきています。

2.3 情報の再生方法

サーバに蓄えられた情報を再生するためには、次の 3 種類の方法があります。

- ダウンロード

クライアントにメディアファイルを http や ftp 等によってコピーし終えた後に再生を開始します。

- 疑似ストリーミング

ダウンロードと同様の方法でファイルを転送しますが、転送が完了するのを待たずに再生を開始します。

- ストリーミング

連続的にデータを転送し、再生を実行します。

ダウンロードによる再生では、再生開始までの時間は対象となるファイルサイズと回線容量に依存します。また、クライアント側にファイルが残ってしまうため、加工や流用等による著作権の問題が発生することが懸念されます。クライアントにファイルが残ってしまうという問題は、ストリーミングシステムを利用することで解決されます。

また、疑似ストリーミングによる再生では、フロー制御ができないため安定した再生が困難です。実際には、疑似ストリーミングでは 128kbps の 5 分間程度のファイルを再生することも困難だと言われ、数分間のファイルしか再生できません。また、疑似ストリーミングではバッファリング技術も利用することができず、ライブに適用することもできません。

最後に示したストリーミングでは、クライアントとサーバ間での伝送を制御することで、効率的な処理が実施されています。

ストリーミングによる再生では、バッファリング技術が活用されています。バッファリング技術は、不安定で低速な伝送回線に対応するためのもので、受信したデータをそのまますぐに再生せずに、いったんキューイングします。また、バッファリング技術はデータ落ちにも対応し、データの入力量が出力量を上回ったときにもバッファに蓄積するようにしています。このようなバッファリングのための時間は遅延時間となっていますが、インターネットを利用したストリーミングではバッファリングは必須技術となっています。

3 ストリーミングシステムでの伝送技術

ストリーミングシステムでは、必ずしもすべてのパケットを受け渡す必要はありません。ストリーミングシステムは、限られた資源の中で運用されるため、信頼性よりも遅延を少なくすることが優先されています。

ストリーミングシステムでは UDP や、その上位レイヤーのプロトコルである RTP (Real-time Transport Protocol) が利用されることが多いです。TCPによる帯域制御のオーバーヘッドを回避するため、サーバが直接データの流量を制御します。TCP はクライアントの状況をサーバに伝えるための管理用セッションに使用されます。ただし、ファイアウォールを通過するために TCP や HTTP が伝送に使用されることもあります。

さらに、ストリーミングシステムでは次のエラー補正技術も利用されています。

- データ再送
伝送されるパケットに一連の番号等のタグを付け、クライアントが受け取れなかったパケットをサーバに再送要求するようにしています。
- FEC (Forward Error Correction)
伝送するデータに冗長性を持たせ、受け取れないパケットをあらかじめ補完できるようにしています。ただし、FEC では、伝送しなければならぬデータ量も増加します。

前述のように、ストリーミングシステムでは UDP とともに次のプロトコルが利用されています。

- RTP (Real-time Transport Protocol、RFC1889)
映像データや音声データを配送することを念頭に置いた、リアルタイムデータ転送のためのプロトコルです。MPEG 等、さまざまなデータを転送するためにペイロードの方法も別の RFC 等で定義されています。QoS や信頼性に関する機能は提供されていません。
- RTSP (RealTime Streaming Protocol、RFC2326)
データ制御のためのプロトコルで、クライアントとサーバ間のネゴシエーションに用いられます。多くのストリーミングシステムで採用されています。実際のデータ配送には他のプロトコルが用いられます。

RTP や RTSP による実際の通信は、図 2 に示すような順序で実行されます。

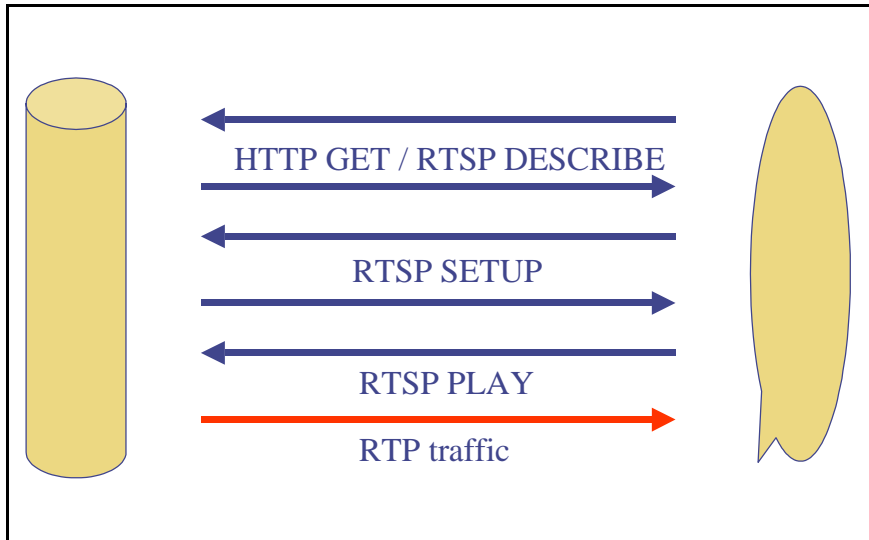


図 2 : ストリーミングシステムのプロトコル

1. HTTP の GET コマンドか RTSP の DESCRIBE コマンドによって、クライアントがサーバにデータを要求し、サーバがデータの位置を返します。
2. RTSP の SETUP コマンドを使って、ポート番号の決定等、転送のための準備を実施します。
3. クライアントが RTSP の PLAY コマンドを送ることで、RTP を使ってサーバからデータが転送されます。

アプリケーションの実装やクライアントの要求等により、実際のデータ転送には http や tcp、udp が使われることもあります。

4 映像や音声の品質と圧縮技術

映像品質ではなめらかな動きが要求されます。既存メディアにおける1秒間のフレーム数は、表4のとおりです。

表4：既存メディアのFPS (Frame Per Sec)

メディア	1秒間のフレーム数 (fps)
8mm フィルム	16
映画	24
NTSC	30 (インターレース)
PAL/SECAM	25 (インターレース)

このような既存メディアの状況から、ストリーミングシステムでも16fps程度の映像であれば動画として認識されるものと思えます。

ストリーミングシステムでは利用できる帯域があらかじめ決まってしまうため、映像の解像度と動きの間でトレードオフが生じます。たとえば、今回の講演のようなセミナー内容をライブで伝えるような場合、講演者自身や表示されているスライドの内容にほとんど動きは生じません。その一方、会話の内容がきちんと聞きとれる音声帯域を確保し、適切にスライド内容を読みとれる解像度を確保する必要があります。

これに対してスポーツ中継等では、解像度よりも動きが重視されることがあります。解像度を犠牲にし、動きを優先しなければならないこともあるのです。

4.1 映像圧縮

たとえば、縦 320 ドット、横 240 ドット、画素深度 24 ビット、コマ数 15fps の映像を無圧縮で伝送するためには、それぞれの値を乗算した 27.648Mbps の帯域が必要となります。実際にはこのような広帯域による伝送は困難であるため、図 3 に示すように H.261、MPEG1、MPEG2、MPEG4 といった圧縮技術が次々と開発されてきました。

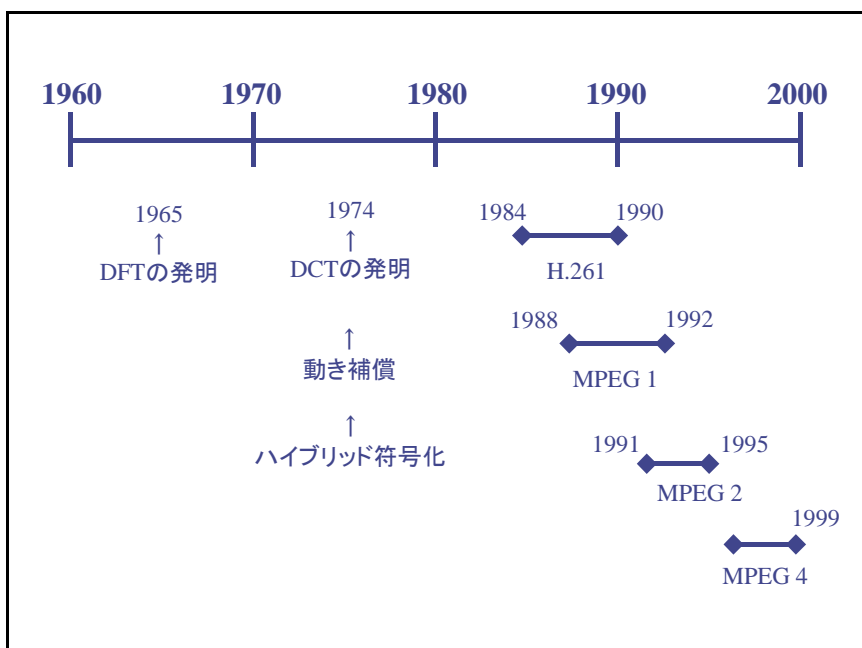


図 3 : 画像通信研究の歴史

たとえば、前のフレームと比べて動きがあった部分のみを送るようにすることで、伝送するデータ量を減らすことができます。また、実際には黒色である部分を黒に近い色とすることで、視聴者には黒色と思わせながらデータ量を減らすこともできます。これらの圧縮技術では、時間軸に沿った画面の相関関係と、人間の視覚特性を利用した符号の偏在化によって、圧縮率を高めています。また、これらの圧縮技術では画質と圧縮率の間にトレードオフが存在し、復元性を 100 パーセント要求しない非可逆な符号化技術も使われています。

- H.261

動画圧縮技術の 1 つである H.261 は、テレビ会議やテレビ電話を主な用途として 1990 年に策定されました。H.261 では、CIF (Common Intermedia Format) と DCT (Discrete Cosine Transform) を組み合わせたハイブリッド符号化アルゴリズムが利用され、144 × 176 ドットで 1 ~ 15fps の映像を、伝送レート 64kbps で再生できます。

- MPEG1

1992年にビデオCD向けとして策定されたMPEG1では、通常のビデオ映像が100分の1に圧縮され、320×240ドットで30fpsの画像を、伝送レート1.15Mbpsで再生できます。

- MPEG2

MPEG1に対してMPEG2は、DVDやデジタル放送のために1994年に策定されました。MPEG2では、15～20分の1の圧縮が可能で、720×480ドットの画像を、伝送レート4Mbpsで再生できます。

- H.263

1996年には、H.261を改良してヘッダ情報を減らしたH.263がテレビ電話向けとして開発され、96×128、144×176、288×352ドットで1～15fpsの各画像を、伝送レート10kbps～2Mbpsで再生できるようになりました。

- MPEG4

64kbps～2Mbpsといった低ビットレートまでの符号化も考慮したMPEG4は、移動体通信やインターネットに適用するために1999年に策定されました。

動画像に対する圧縮技術は、データ落ちに弱いという問題があります。圧縮のためにデータの差分を利用していることが原因です。また、パン、ズーム、切り替え、追いかかけ等の画面全体の変化に追従することも困難です。

4.2 音声圧縮

音声圧縮にも、映像圧縮と同様の考えが適用できます。このとき、次の3種類がオーディオデータのパラメータとなります。

- サンプリングレート
- サンプリングビット
- モノラル/ステレオ

また、音質は、次の2つの項目によって評価されます。

- 周波数特性
- 了解度

図 4 に示すように人間の可聴帯域は 20KHz 程度で、男性の話し声の中心は 1KHz、女性の話し声の中心は 2KHz となっています。このような周波数特性に対して、電話は 5KHz を上限とした周波数特性によって機能を果たしています。これと同様に、音声圧縮でも利用するコーデックでの特性を考慮する必要があります。

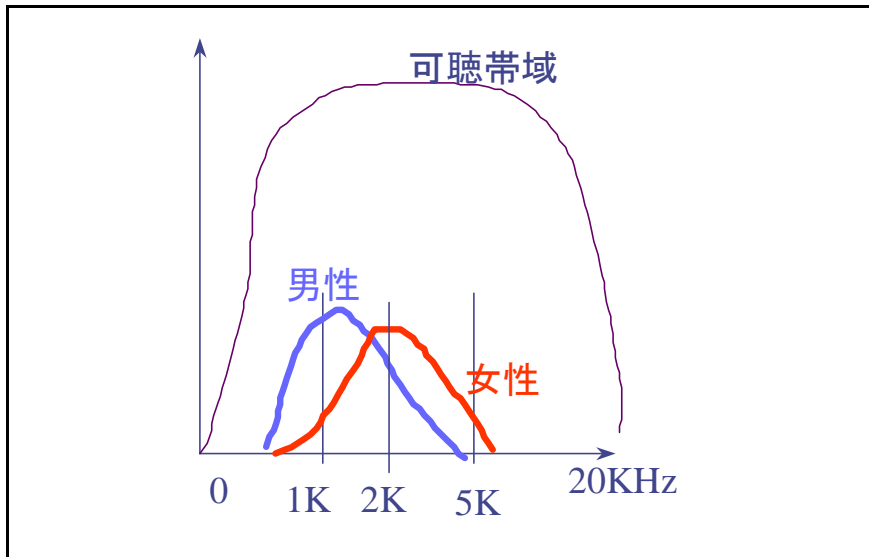


図 4 : 周波数特性

また、音声では、データの忠実度や識別度に合わせて圧縮技術を使い分けるようにします。たとえば、音楽を圧縮するときには音の特性の忠実度を確保するようにします。これに対して、セミナー等で会話の識別度を確保するときには、人の喉の構造を解析しシミュレートする方法等も利用されています。

次に、現在利用できる代表的な音声圧縮方法を示します。

- MD (ATRAC)
- MPEG Audio Layer-3
- G.721 ADPCM
- G.728 LD-CELP
- G.723.1 ACELP/MP-MQL
- G.729 CS-ACELP
- TwinVQ
- Dolby AC-3 base
- MS Audio

現在では CPU の高速化に伴い、適用する帯域に合わせて複数のコーデックが採用されるようになってきています。複数のコーデックを利用できるようになっているため、実際の音源に合わせてコーデックを選択すべきでしょう。

また、初期のストリーミングシステムでは映像と音声を同期させることができませんでした。現在ではサウンドボードのクロック利用等の方法によって映像と音声を同期できるようになっています。では、映像と音声の間では、どの程度の遅延までが許されるのでしょうか。通常は、300 ミリ秒以内であれば不自然とは感じられないため、この値以内であれば同期していると考えられます。

5 事例紹介

他の複数のメディアと組み合わせたストリーミングメディアの利用が広く試みられてきています。SMIL (Synchronized Multimedia Integration Language)、HTML + TIME 等と組み合わせることで、Web、テキスト、静止画等とストリーミングメディアを統合させているものがあります。チャット等の他のインターネットアプリケーションと連携させることで、情報を流すだけでなく対話性を実現しようとする試みもあります。

この講演では、次の事例を紹介します。

- Webrama (5.1 を参照)
- オペラ LIFE (5.2 を参照)
- School Of Internet (5.3 を参照)
- LIVE! ECLIPSE 99 (5.4 を参照)

5.1 Webrama

Webrama とは、Web (ウェブ) と Drama (ドラマ) を組み合わせた造語です。Webrama では、画像とテキストを同期させたり、音声に合わせてその内容を表示させたりという処理が、SMILによって実現されています。たとえば、ユーザは特定の章をクリックすることで、既に鑑賞した内容の続きから観ることができます。また、ドラマの途中でユーザの選択によって内容を変更することで、ストーリーに多様性を持たせています。

このような同期によって、これまでよりも短い時間で多言語に対応できるようになるため、映画業界等での利用が検討されています。

5.2 オペラ LIFE

オペラ LIFE では、ストリーミングアプリケーションとして QuickTime 4.0 と WMT 4.0 を、クライアントアプリケーションとして Director 7.02 をそれぞれ利用し、各製品を同期させることでインターネット上の観客に独自のサービスを提供しました。このようなアプリケーションの組み合わせによって、ユーザは、受け取ったストリーミングデータを単に観るだけでなく、あらかじめ用意された独自の演出に従って、会場とは異なる視点で鑑賞することができます。また、オペラ LIFE では、HBG (Hyper Broad Gathering) という方法を使い、クライアントからのフィードバックを演奏者に返すようにもしました。

さらに、オペラ LIFE では、会場での演出にもインターネットによるストリーミングを利用しました。この演出では、インターネットを使い、東京での演奏をニューヨークとフランクフルトのダンサーに送り、その演奏に合わせて演技している様子を東京に送り返すようにしました。当然、このような長距離での伝送では、東京から送り出した演奏がニューヨークやフランクフルトに到着するまでに遅延が発生します。さらに、その演奏に合わせてダンサーの演技が東京に返されたときにも遅延が発生します。ここでは、このような遅延を重ねた映像を地球規模での演出の効果として利用しました。

5.3 School Of Internet

1997年9月に作業グループが設立された WIDE プロジェクトの SOI (School Of Internet) では、1999年9月までに約 2,800 名の受講者が登録し、4 つの大学の 14 コースと 45 種類の特別講義が約 320 時間分蓄積されています。SOI では HTML による資料と映像データや音声データを同期させたり、講義の様態とホワイトボードの内容を同時に表示させたりする際にストリーミング技術が利用されています。

ただし、SOI では、講演者がコンピュータ上で作成した資料を OHP にしてから表示し、その内容をカメラで撮影した後にデジタル処理するといった手間がかかっています。現在ではこのような中間でのアナログ処理を省いて、デジタルデータを直接利用したいという要求に対して、対応製品が提供され始めています。

5.4 LIVE! ECLIPSE 99

LIVE! ECLIPSE 99 は、1999 年 8 月 11 日に欧州から中近東にかけて観測された皆既日食をインターネットで中継したプロジェクトです。プロジェクトの計画当初には 9 カ所の観測地点からのストリームを複数のエンコーダでデジタル化し、マルチキャストで同一グループに配信したり、ユニキャストで配信しユーザに任意に切り替えさせたりする方式等が検討されました。

実際には、帯域幅の減少や操作の煩雑さから各方式は採用されず、図 5 に示す構成によって、単一ソースからのストリーミングに見せかけることになりました。

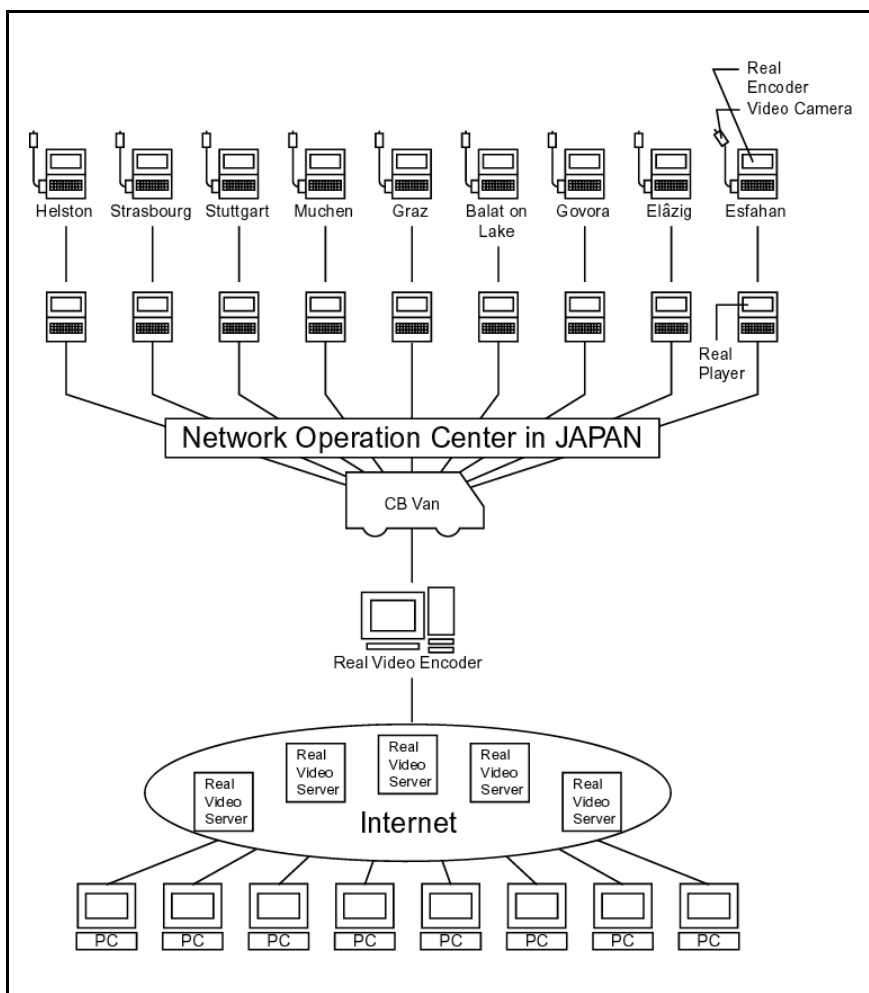


図 5 : LIVE! ECLIPSE 1999 のシステム構成

このシステムでは、各観測地点でエンコーダを動作させ、データを東京大学内の9台のマシンに送ります。そして、受け取ったデータからスキャンコンバータによって映像信号を取り出し、ビデオ編集機器を使って画像を編集し直した後に、再度エンコードして配信用サーバに送るようにしています。

このうち各観測地点での映像伝送システムには、映像品質を重視してエンコーダのパラメータをチューニングする必要もあったため、RealSystem G2 を採用しました。また、ユーザへの配送には、安定性を重視して RealVideo 5.0 を採用しました。

観測地点の1つであったイランのイスファハン工科大学に実際に行ってみると、東京大学までの RTT が 1000 ミリ秒を超えるほどの遠距離であり、IRC による連絡のみが唯一の連絡手段となりました。また、イスファハン工科大学のリンクが 64kbps であったため、中継当日には他のポートをすべて利用禁止としたり、イランからカナダまでの衛星回線がボトルネックとならないよう、イスファハン工科大学からのストリームを制限したりする等の処置を実施する必要がありました。

6 大規模配信に対する検討

最近では、xDSL、CATV、衛星、無線等の広帯域なアクセス回線が登場してきています。また、大規模なインターネット中継も広まってきています。このため、大規模な配信を前提とした負荷分散を検討する必要性が出てきています。

当然のことですが、回線の帯域によってアクセスを処理できる能力は決まってしまう。たとえば、1.5Mbps の回線では 20kbps のストリームを同時に 75 個までしか流せませんし、45Mbps の回線では 45kbps のストリームを同時に 1,000 個までしか流せません。このため、大規模な配信のためには負荷分散が必要となります。また、中継時の Web サーバの高負荷によって、中継点へのポイントを持つ Web サーバ自体にアクセスできないという問題も表面化し始めています。このような問題に対処する方法として、図 6 に示すような中継サーバキャッシュの利用があります。

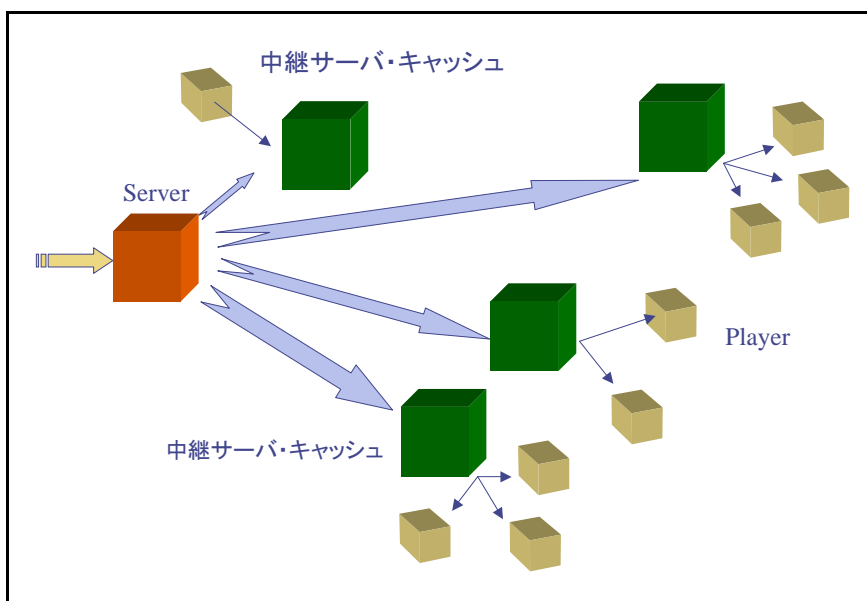


図 6：中継装置の設置

アクティブサーバとも呼ばれる中継サーバキャッシュは、Web サーバと連動して機能し、その更新頻度を Web サーバ側から制御できます。

7 著作権に対する検討

今後は、著作権を配慮した配信がより必要となります。既に示したストリーミング配信とダウンロード配信では、著作物の扱いが異なっています。

ここで、参考のために、それぞれの定義を著作権の見地から再度まとめます。

- ストリーミング配信

配信者が受信者の受信装置に著作物を複製する意図を持たずに行う方式による配信です。

- ダウンロード配信

受信者が受信装置に著作物を複製する方法による配信です。

また、公衆送信では「同時送信」と「インタラクティブ送信」の区分があります。ラジオ、テレビ等の無線放送や有線放送は、常に提供され続ける内容を、ユーザがチャンネル等を切り替えて受け取ることができる「同時送信」となっています。これに対して、インターネット放送が含まれる自動公衆送信は、送信される場所にアクセスし、クリック等の動作によって受信したい意図を示した後にコンテンツを受け取るため、「インタラクティブ送信」となっています。

現在の同時送信とインタラクティブ送信という区分によって、同時送信の権利しか得ていない既存の放送事業者や有線放送事業者は、その内容をそのままインターネット送信に利用できないという事態になっています。この制限は、1996年にWIPO(World Intellectual Property Organization : 世界知的所有権機関)によって検討されたデジタルデータの扱いを日本独自に解釈したものであり、米国をはじめ、各国ごとに異なったものとなっています。

さらに、音楽配信については、試聴した後に気に入った曲のみを購入してダウンロードする方法が一般化しています。このような状況に対して、現在SDMI(Secure Digital Music Initiative)という団体を中心に、各曲のコピー回数や添付する関連情報が検討されています。

8 まとめ

単にストリーミングメディアで情報を発信するのではなく、社会的な特性や技術的な特性を理解した上でストリーミングメディアを利用するとよいでしょう。また、常に最新の技術を自分の目や耳で確かめていくことも大切でしょう。