

■S1 群 (情報環境とメディア) - 5 編 (通信・放送・インターネットの融合)

2 章 インターネット放送

(執筆者: 甲藤二郎) [2009 年 5 月 受領]

【本章の構成】

本章では以下について解説する.

2-1 VoD サービス

2-2 ダウンロードサービス

■S1 群-5 編-2 章

2-1 VoD サービス

(執筆者：甲藤二郎) [2009年5月 受領]

インターネット放送とは、インターネット上でマルチメディアコンテンツを配信、再生する技術である。インターネット放送は、マルチメディアデータを転送しながら再生する技術と、マルチメディアデータを転送してから再生する技術に大別される。本稿では前者を VoD (Video on Demand) サービス、後者をダウンロードサービスと呼び、特に前者について詳細な説明を行う。

2-1-1 VoD サービスのプロトコル階層

インターネット上の VoD サービスでは、TCP/IP プロトコルの上でマルチメディアコンテンツが配信される。図 1・1 は、インターネット上の VoD サービスで使用される典型的なプロトコルスタックを示したものである。基本的な構成要素として、マルチメディアコンテンツを配信・再生する階層、VoD サービスのセッション制御を提供する階層、著作権 (DRM: Digital Right Management) を管理する階層、Web ブラウザ経由でポータルサイトとして顧客に番組情報を提供し、セッションにナビゲートする階層がある。

Video	Audio	RTSP, 独自	DRM	portal
RTP (or MPEG-2 TS, FLV)				
UDP (or TCP, HTTP/TCP)		TCP (or HTTP/TCP, UDP)		HTTP/TCP
IP				
Network (Ethernet, Wireless LAN, ...)				

図 1・1 VoD サービスのプロトコル階層

2-1-2 マルチメディアコンテンツの配信・再生

図 1・1 左部の Video, Audio を最上位とする階層が、インターネット上のマルチメディアコンテンツの配信・再生を行う階層である。この場合、マルチメディアコンテンツを転送しながら再生 (ストリーミング) する必要がある、通常のデータ転送では必要のない同期再生機能が求められる。これに用いられるのが、タイムスタンプを運ぶ RTP (Realtime Transport Protocol)¹⁾ と MPEG-2 TS (Transport Stream)、及び、同期再生情報を含んだ FLV (Flash Video) などのファイルフォーマットである。

(1) マルチメディアコンテンツ

ビデオコンテンツは、MPEG-2 Video や H.264/AVC などの、映像圧縮に関する国際標準方式を用いて圧縮されることが多い。まず、MPEG-2 Video は、広くデジタル放送や DVD で使用されている国際標準方式である。また、H.264/AVC は、2003 年に標準化が完了した最新の国際標

準方式である。既にワンセグや一部の衛星放送などに使用されるとともに、今後の MPEG-2 Video の置き換えが期待されている。また、インターネット上に広く普及している FLV でも、以前は H.263 ベースの圧縮方式が使用されていたが、2007 年末からは、H.264/AVC ベースの圧縮方式が使用されている。一方、音声・音楽コンテンツは、MPEG-1 Layer III (MP3), MPEG-2 AAC, MPEG-4 AAC などの国際標準方式を用いて圧縮されることが多い。

インターネット上のコンテンツ配信という点で、以前は CIF (352×288) や QCIF (176×144) の小サイズで、更にはフレームレートを落とした低解像度コンテンツが主流だったが、昨今は、ブロードバンド化の進展とともに、SDTV (720×480) 解像度はもちろんのこと、前章の IPTV にも関連付けられるように、HDTV (1920×1080) クラスの高解像度コンテンツも配信対象になりつつある。

(2) RTP, MPEG-2 TS, ファイルフォーマット

インターネットは従来の電話網や放送網と比べると品質の安定しないネットワークであることから、伝送遅延の揺らぎ（ジッタ）やパケットの廃棄が発生する。そこで、通常のデータ転送を補充する層が必要になる。その代表例が RTP であり、インターネット上のマルチメディアコンテンツのストリーミング配信のための各種機能を提供している。

インターネット上の伝送遅延の揺らぎを許容しながらマルチメディアコンテンツの同期再生を実現するためには、まず、受信端末で十分な個数の受信パケットをバッファリングし、揺らぎを吸収した上で、各パケットの再生タイミングに従って復号、再生を行う必要がある。そのためには、各パケットに復号タイミングを明示するタイムスタンプを付与するか、事前に復号時間間隔を固定し、その情報を通知する必要がある。

v=2	P	X	CSRC count	M	packettype	sequence number
time stamp						
synchronization source (SSRC) identifier						
contributing source (CSRC) identifiers						
(payload format extension)						
multimedia data						

図 1・2 RTP ヘッダの構成

図 1・2 は、各パケットに付与される RTP ヘッダの構成を表している。このなかで、同期再生に使用されるのがタイムスタンプ (time stamp) であり、4 バイトを使って、マルチメディアコンテンツの同期再生タイミングを指定している。この情報に従って、ビデオ、オーディオそれぞれの同期再生が可能になる (メディア内同期)。ただし、RTP では、メディアごとに時間軸の異なるタイムスタンプの付与を認めており、ビデオとオーディオの同期 (メディア間同期) を実現するためには、更に RTCP-SR (Sender Report) パケットが必要になる。ここで RTCP (RTP Control Protocol) とは、RTP と併用され、同期再生情報に加え、通信状況を表す各種の統計情報を交換するためのプロトコルである。

図 1・3 は、RTCP-SR パケットの構成を表している。図中の SR は送信者としての情報、RR

(Receiver Report) は受信者としての情報を示す。そして、メディアパケットごとに付与される RTP ヘッダとは異なり、RTCP パケットは送受信端末間で定期的に変換される。SR に含まれる RTP タイムスタンプは、名前の通り、RTP ヘッダのタイムスタンプと時間軸を共有している。SR には、更に 8 バイトの NTP (Network Time Protocol) タイムスタンプ情報が付与されている。これがビデオ、オーディオに共通の時間軸であり、この時間軸上にメディアごとの RTP タイムスタンプをマッピングすることにより、メディア間同期が実現される。

v=2	P	RC	PT=SR=200	length	
SSRC of sender					SR
NTP time stamp (most significant word)					
NTP time stamp (least significant word)					
RTP time stamp					
sender's packet count					
sender's octet count					
SSRC_n					RR
fraction lost		cumulative number of packets lost			
extended highest sequence number received					
interarrival jitter					
last SR					
delay since last SR (DLSR)					

図 1・3 RTCP-SR パケットの構成

RTP/RTCP は、このほか、RTP ヘッダのシーケンスナンバを用いたパケット廃棄の検出や、RTCP パケットに記載された各種の統計情報を活用した輻輳制御を可能にする。ただし、本稿で扱う VoD サービスは、後述するように、トランスポート層のプロトコルとして TCP (Transmission Control Protocol) を用いるものも多く、その場合は、パケット伝送の信頼性が保証され、かつ TCP 自身の輻輳制御を利用するため、それらの技術に関する詳細は省略する。

アクトビラのように、RTP の代わりに、デジタル放送や DVD に使用されている MPEG-2 TS を用いるものもある。この MPEG-2 TS は、RTP と同様のタイムスタンプを運ぶ機能と、ビデオデータとオーディオデータを多重化する機能を持つ。タイムスタンプを用いる点では RTP と同じであるが、タイムスタンプがメディア共通である点、複数メディアの多重化を実現する点が RTP とは異なる (RTP では、多重化は下位の IP に任せる)。また、MPEG-2 TS がデジタル放送の転送フォーマットであることから、デジタル放送対応のテレビジョンとの相性が良い点が特徴として挙げられる。

また、YouTube やニコニコ動画で使用されている FLV のように、ファイルフォーマットで配信する場合もある。ファイルフォーマットとは、先頭に各種の制御情報を埋め込み、ビデオデータとオーディオデータを記録したファイル形式のことであり、一種の多重化フォーマットと考えることもできる。ファイルフォーマットの中にはフレームレート (復号時間間隔) が記載されており、セッション開始時に受信端末にこの情報を通知し、パケット受信中はこの時間間

隔を保ちながら再生することにより、同期再生を実現することができる。また、AVI(Audio Video Interleave)などの古いファイルフォーマットはそのままではストリーミング配信に適さなかったが、FLVやMPEG-4ファイルフォーマットなどは、当初からストリーミング応用を想定しており、そのままVoDサービスに適用することができる。

(3) トランスポートプロトコル

マルチメディアコンテンツを配信するトランスポートプロトコルとしては、もともとはUDP(User Datagram Protocol)が想定されていた。その理由は、通常のデータ転送で広く使用されているTCPでは、パケットロスを検出すると再送を行うために伝送遅延が大きくなり、かつ、コネクション志向であるために、IPマルチキャストのような1対多の放送応用には適さないと考えられていたからである(IPマルチキャストはUDPを使用する)。

しかし、TCPが放送応用に適さないことは変わらないが、ブロードバンド化の進展とともに、再送遅延の問題が深刻ではなくなってきた。セッション開始から再生開始までの遅延時間をプレイアウト遅延と呼ぶが、これはすなわちジッタ吸収のためのバッファリング時間であり、ネットワークの広帯域化とともに小さくできる。また、メモリやディスクの大容量化の影響も大きく、10分程度の映像をローカルに貯め込んでしまうことも容易であり、セッション開始から一気にTCPでダウンロードしてしまい、以降はローカルに再生することができる。このため、ライブ放送を除き、多くのVoDサービスがTCPもサポートするようになっていく。

更に、近年はセキュリティ対策のために、ファイヤウォールを用いて、Webや電子メールなど、特定のアプリケーションにしか外部との通信を許さない場合が多い。この場合、Webが使用するポート番号(80番)は多くのケースで利用可能であり、VoDサービスもまた、(TCP上で)HTTP(Hyper-Text Transfer Protocol)を用いてストリーミング配信を行う場合が増えている。これはHTTPストリーミングと呼ばれ、以前はUDPとTCPのポート番号探索後、すべてが閉じている場合の最終手段として利用されることが多かった。これが最近ではデフォルトとしてHTTPストリーミングを想定し、単独のビューアを開かずに、Webブラウザ上にコンテンツ表示するものが増えている。具体例としては、YouTube、ニコニコ動画、GyaOなどが挙げられる。

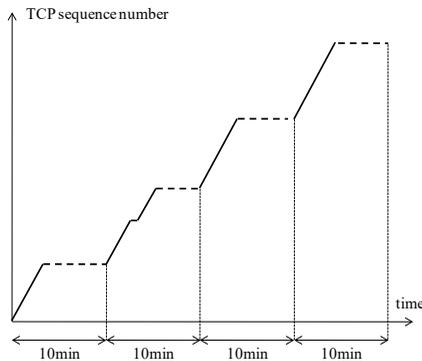


図 1・4 TCP シーケンスナンバーの挙動

図 1・4 は、HTTP ストリーミングを用いた VoD サービスの例として、GyaO における TCP のシーケンスナンバーの振る舞いを示したものである。GyaO の場合、約 10 分ごとに CM が入って

は、ストリーミング再生が再開される。一方、図 1・4 を見ると、セッション開始、もしくは再開後にマルチメディアコンテンツのダウンロードが始まり、数分でダウンロードを終了しては、以降は PC 内にデータを取り込んだまま再生を続けていることが分かる。仮にコンテンツのビットレートを 1 Mbit/s とした場合、これを 10 分間格納するための記憶領域は 600 Mbit、すなわち 75 MB (バイト) であり、昨今のメモリサイズであれば余裕を持って格納できるデータ量である。一方、YouTube でも同様の実験を行ったが (各コンテンツに 10 分の制限がある)、傾向は図 1・4 と同様であり、数分でダウンロードを終了し、以降は再生を継続する。

2-1-3 セッション制御、その他

(1) RTSP と SDP

図 1-1 中央部の RTSP (Real-Time Streaming Protocol) ²⁾ の階層は、VoDセッションの管理を司るプロトコルスタックである。RTSP は、当初からインターネット上のストリーミングを想定して作成された制御プロトコルであり、MPEG で検討された DSM-CC (Digital Storage Media Command and Control) を参考に、マルチメディアコンテンツの再生、停止、早送り、スキップ、記録などの VCR (Video Cassette Recorder) コマンドを定義している。また、メッセージの体系は HTTP をベースとしており、HTTP に類似した手順でサーバ・クライアント間のセッション制御を実行する。

表 1・1 RTSP のメソッド一覧

メソッド	方向	要求条件	内容
DESCRIBE	C->S	推奨	セッション情報の取得 (SDP 等)
ANNOUNCE	C->S, S->C	オプション	C->S: クライアントからのセッション情報の通知 S->C: セッション情報の更新
GET_PARAMETER	C->S, S->C	オプション	セッションパラメータの取得
OPTIONS	C->S, S->C	必須	オプション機能のチェック
PAUSE	C->S	推奨	メディア転送の中断
PLAY	C->S	必須	メディア転送の開始、再開
RECORD	C->S	オプション	メディア情報の記録
REDIRECT	S->C	オプション	リダイレクション
SETUP	C->S	必須	セッションの初期化
SET_PARAMETER	C->S, S->C	オプション	セッションパラメータの設定
TEARDOWN	C->S	必須	セッションの終了

表 1・1 は、RTSP が定義しているリクエストメッセージのメソッド一覧を示している。この表において、S はサーバ、C はクライアントを表し、多くの使用例においてクライアントからサーバに向けてリクエストメッセージが送信され、サーバから HTTP と同様のレスポンスメッセージ (200 番台: 成功, 500 番台: サーバエラーなど) が返される。典型的な使用例として、

クライアントは、セッション開始前に

- DESCRIBE メソッドによってサーバから SDP³⁾ (Session Description Protocol, 後述) を取得し、
- SETUP メソッドによってセッションの開始 (資源の確保) を要求し、
- PLAY メソッドによってサーバにストリーミング配信を開始させる。

また、セッション中は、

- PAUSE メソッドと PLAY メソッドを用いてストリーミングの停止、再開を指示し、
- 最終的に TEARDOWN メソッドを用いてセッションを終了する。

サーバからクライアントに通知される SDP は、VoD セッションを開始するために必要な情報を集めたテキストデータである。この SDP には、以下のような情報が含まれる。

- ストリームタイプ (ビデオ, オーディオ, 制御など)
- 使用する圧縮アルゴリズムと転送レート
- 使用するポート番号と転送プロトコル (RTP など)

復号できない圧縮アルゴリズムが通知されたとき、クライアントは必要に応じてプラグインをダウンロードしようとする。また、使用できないポート番号などが通知されたとき、クライアントは別の通信手段を要求することができる。

RTSP の各種メッセージは、通常は 554 番のポート番号を使って、TCP で転送される。これは一般的に、制御メッセージは即時性よりも信頼性が重要なためである。一方、80 番のポート番号を使用する HTTP ストリーミングの場合、多くのケースですべてがブラウザ上のインタフェースとして実装されており、この場合は RTSP と同等のメッセージを独自に定義し、マウスによるボタン操作を介して、各種の VCR コマンドを実行することになる。

(2) DRM サーバとポータル

図 1・1 右部にある DRM サーバの階層は、マルチメディアコンテンツの著作権を管理する。本件については 1 章、8 章に詳細な説明があり、本章では詳細は割愛する。一方、ポータルサーバは、通常の Web ページとして顧客を引き付け、VoD サービスの番組一覧を提供し、顧客をセッション (番組) にナビゲートする。デジタル放送には EPG (Electronic Program Guide) があり、定期的に配信・更新されるが、インターネット放送にも SAP (Session Announcement Protocol) と呼ばれる IP マルチキャストを用いた番組配信方式があった。しかし、様々なポータルサイトがアクセス数を競っている現状では、Web ページ経由の番組案内の方が、使い勝手も情報量も勝っているものと考えられる。

(3) 全体の手順

図 1・5 は、ポータルサーバへのアクセスから VoD サービスの開始までの手順の例を示したものである (括弧内の数字はポート番号を示す)。ユーザは最初にポータルサーバにアクセスし、視聴したいコンテンツを検索、探索する。所望のコンテンツが見つかると、必要に応じてユーザ認証を行い、コンテンツの DRM を確認し、VoD サーバにアクセスする。ここまでの手順は、通常 HTTP か HTTPS を用いて行う。VoD サーバへのアクセス後は、RTSP を用いてセッションを開始するか、Web ブラウザ上のインタフェースを介してセッションを開始するか、のいずれかとなる。マルチメディアコンテンツの配信もまた、RTP (あるいは MPEG-2 TS) を使用するか、HTTP ストリーミングを行うか、のいずれかになる。

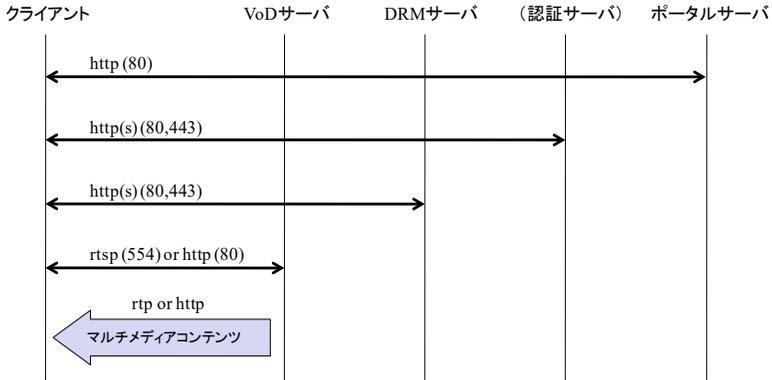


図 1・5 VoD サービスの開始手順の例

2-1-4 VoD サービスの具体例

VoD サービスの典型例としては、古くから競争を続けているマイクロソフトの Windows Media Technologies, アップルの iTunes/QuickTime, リアルネットワークスの Real Systems, アドビの Flash Video などが挙げられる。Flash Video はもっぱら HTTP ストリーミングを行うが、その他の三手法は RTSP (もしくは独自の制御プロトコル) と HTTP ストリーミングを使い分けられる実装になっている。

HTTP ストリーミングに特化したポータルとしては、YouTube, ニコニコ動画, Yahoo 動画, GyaO などが挙げられる。YouTube とニコニコ動画は FLV を使用し、Yahoo 動画と GyaO は Windows Media を使用している。一方、先の例は主に PC をプラットフォームとするものであるのに対し、アクティブラは、ネットワーク接続したデジタルテレビをプラットフォームとしている。このため、MPEG-2 TS を多重化フォーマットとして採用するなど、独特の規格の採用を行っている。

■参考文献

- 1) H. Schulzrinne, et al. : “RTP: A Transport Protocol for Real-Time Applications,” IETF RFC 3550, Jul. 2003.
- 2) H. Schulzrinne, et al. : “Real Time Streaming Protocol,” IETF RFC 2326, Apr. 1998.
- 3) M. Handley, et al. : “SDP: Session Description Protocol,” IETF RFC 4566, Jul. 2006.

■S1 群-5 編-2 章

2-2 ダウンロードサービス

(執筆者：甲藤二郎) [2009年5月 受領]

VoD サービスがマルチメディアコンテンツを転送しながら復号・再生する技術を基盤とするのに対し、ダウンロードサービスはいったんダウンロードを完了してから復号・再生するものである。ネットワーク的には、技術的な難易度は高くない。その一方で、RSS を活用してニュースやブログにおけるマルチメディアコンテンツの公開、発信を容易にするなど、マルチメディアコンテンツの Web 2.0 的な使い方を普及・促進させるための活動が多数見られている。

ダウンロードサービスの代表例としてはポッドキャストが挙げられる。これは、コンテンツの入手側から見れば、Web 上に新たに発信されたマルチメディアコンテンツを、RSS の入手を介して自動的に・定期的にダウンロードしたり、ポッドキャストの URL に直接アクセスしてダウンロードしたりすることができる。また、コンテンツの発信側から見れば、RSS を更新・修正・公開することで、多数のユーザに対してコンテンツの情報を一斉に配信することができる。