

モバイルマルチメディア信号処理技術特集

マルチメディア配信技術

本稿では、音楽や映像などのマルチメディア配信技術として、マルチメディアトランスポート技術や配信制御技術、マルチメディアコンテンツのメタデータを記述する MPEG・7 について触れる。マルチメディア配信は次世代移動網における主要なアプリケーションとして期待されており、これらの諸技術がいずれも重要となる。

よしむら 吉村 たけし 健 かわはら 河原 としろう 敏朗 せきぐち 関口 しゅんいち 俊一 えとう 栄藤 みのる 稔

1. まえがき

携帯電話によるインターネットアクセスがiモードにより実現し、インターネット上のさまざまなコンテンツを携帯電話を介して見るようになってきている。現時点ではE-mailの送受信やWebブラウジングが主な利用形態であるが、次世代移動網により高伝送レートの無線アクセスが可能になれば、これらに加えて音楽やビデオなどのマルチメディアコンテンツへのアクセスも一つの主要なアプリケーションになると思われる。

本稿では、このようなマルチメディア配信を可能とする諸技術について紹介および解説する。2章では、インターネット上で所望の品質を満たしながら、マルチメディアコンテンツの伝送を可能とするトランスポート技術について説明する。3章では、マルチメディアセッションを制御する諸技術を紹介し、4章ではマルチメディアコンテンツ記述用メタデータの国際標準である MPEG (Moving Picture Experts Group)・7 を解説するとともに、そのマルチメディア配信への応用について述べる。最後に5章において本稿のまとめを行う。

2. マルチメディアトランスポート技術

2.1 リアルタイムトランスポート

E-mail や Web ブラウジングでは、ある程度の伝送遅延は許容できるもののパケットロスのない信頼性の高いパケット伝送が要求される。このようなアプリケーションに対し、再送により信頼性の高いパケット伝送を提供する TCP (Transmission Control Protocol) が用いられている。

一方、マルチメディアコンテンツの伝送には、ある程度のパケットロスは許容できるものの伝送遅延の少ないパケット伝送が要求される。そのため、このようなアプリケーションに対しては TCP ではなく UDP (User Datagram Protocol) と RTP (Real-time Transport Protocol) [1] が一般に用いられる。UDP は TCP のように再送による信頼性は提供せず、パケットを配送するのみの非常に単純なプロトコルである。

それに対し、RTP は音声や映像などのリアルタイム性を有するメディアの伝送を目的として IETF (Internet Engineering Task Force) の RFC (Request for Comment) 1889 で規定されているプロトコルである。RTP の主な機能として以下の機能が挙げられる。

(1) ペイロードタイプの指定

RTP ペイロードの符号化に用いられている符号化アルゴリズムを指定する。

(2) シーケンスナンバーの付与

インターネットでは順番どおりのパケット伝送を保証しないため、受信者が送信順番どおりに並び替えることができるようにしている。

(3) タイムスタンプの付与

インターネット上で伝送遅延ゆらぎ (ジッタ) が生じてても、受信者が適切なタイミングでマルチメディアデータを再生することを可能とする。

RTP はそれ自身では単独では動作しないプロトコルのフレームワークであり、ペイロードタイプとペイロードフォーマットのマッピングを決めるプロファイル文書と、アプリケーションごとにペイロードフォーマットを記述する文書によって完全に規定される。現在、次世代移動網における音声・映像符号化である適応マルチレート (AMR: Adaptive Multi Rate) や MPEG-4 の RTP ペイロードフォーマットも IETF の AVT (Audio Video Transport) WG (Working Group) で規格化に向けて作業中である。

また、RTP のセッション情報を伝達するために RTCP (RTP Control Protocol) が RTP とともに規定されている。RTP セッション参加者は定期的に RTCP パ

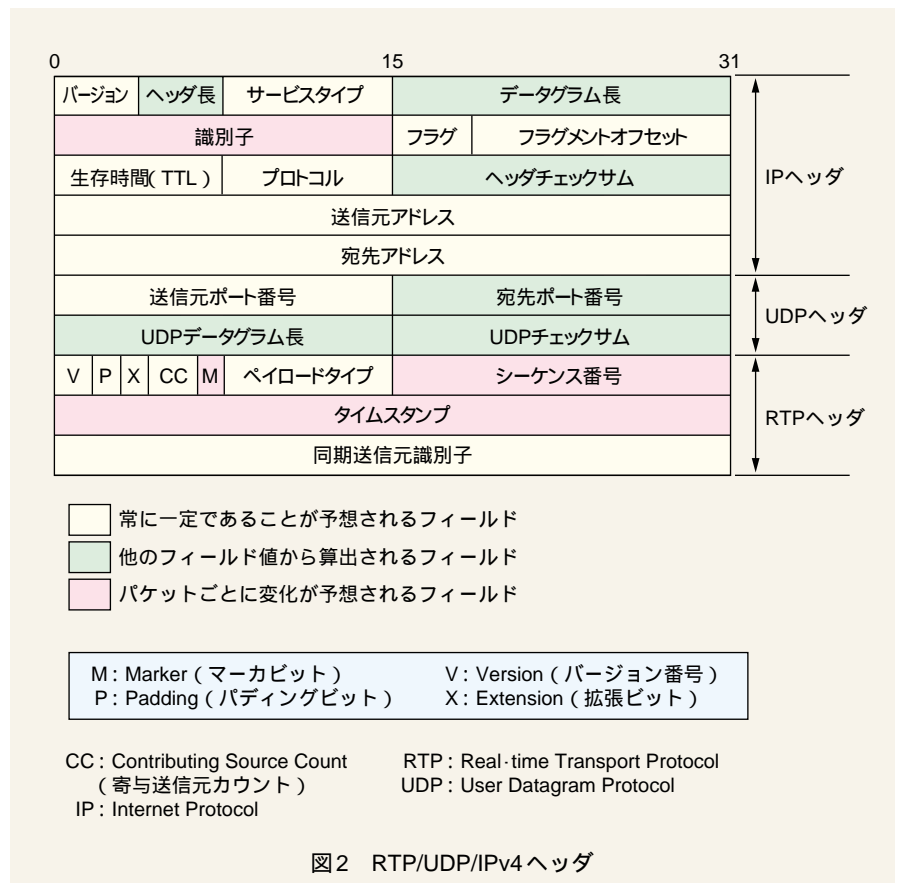
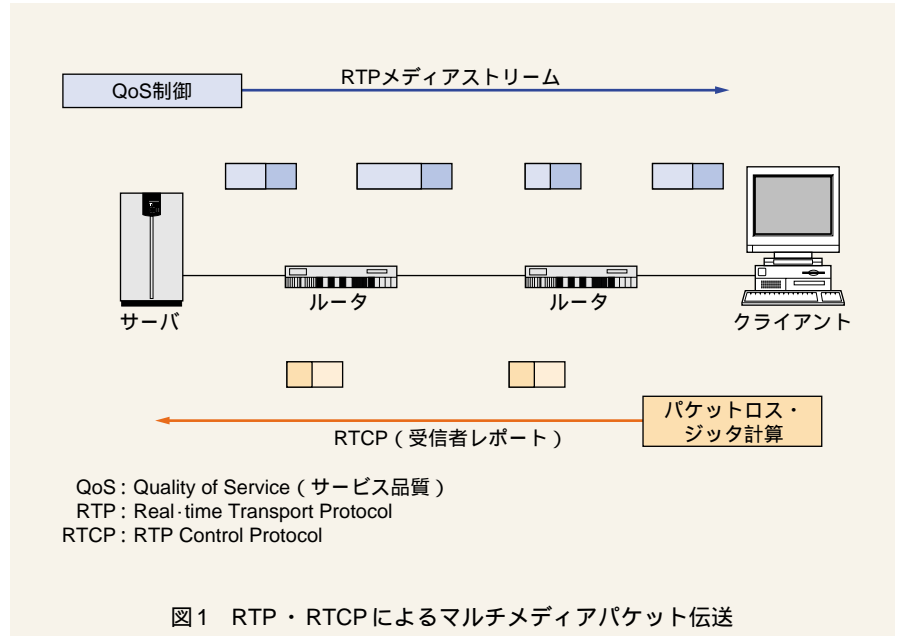
ケットを送信し、各参加者固有の情報
を伝達する。加えて、RTCPの受信者
レポートによりパケットロス数やジッ
タといった受信品質情報を伝達でき
るため、これらの情報から通信状況
を把握し、送信レート制御などのサ
ービス品質（QoS：Quality of Service）
制御に活用することができる（図1）。
IETFのAVT WGでは、RTCPの情報を基
にRTPパケットの再送や音声/映像符
号化パラメータの変更も可能とする
ようにRTCPの拡張について検討が
始まっている。

2.2 ロバストヘッダ圧縮

RTPパケットは、UDPおよびIP
（Internet Protocol）ヘッダも含めると、
その合計のヘッダ長が40バイトにも
達する。音声パケットの場合、1パケ
ットあたりのペイロード長は20～30
バイト程度であり、RTP/UDP/IPヘッ
ダのオーバーヘッドの占める割合が非常
に大きい。特に、低速なモデム回線や
移動通信回線を介して通信を行う際
には、このヘッダのオーバーヘッドのた
め効率的な通信を行うことができない。
そのため、ネットワーク資源の有効利
用の観点からRTP/UDP/IP CRTP（ヘ
ッダ圧縮）が注目を集めている。

現在、モデムなどの低速リンクを想
定したものととして、IETFのRFC2508
[2]で規定されているRTP/UDP/IP
CRTPがある。CRTPは、図2に示すよ
うにRTP/UDP/IPヘッダ情報の大部分
が一定もしくはその差分値が一定であ
ることを利用し、1つ前のヘッダ情報
をもとに差分をとることで効率的なヘ
ッダ圧縮を可能としている。CRTPに
より40バイトのRTP/UDP/IPヘッダ
を2～4バイトにまで圧縮することが
できる。

しかし、CRTPはパケットロスなど
の伝送路誤りを想定しておらず、その
まま移動通信環境へ適用することは困
難である。すなわち、CRTPでは1つ
前の非圧縮ヘッダ情報をもとにヘッダ
圧縮を行っているため、1つのパケッ



トロスにより送信側と受信側で同期は
ずれが生じ、以降のパケットを正しく
復元することができなくなる。そのた
め、1つのパケットロスが受信側での
パースト的なパケット廃棄につなが
り、特に移動通信環境のように誤りの
多い環境では、頻繁なパケットロスが

伝送効率を大きく低下させることにな
る。

このような状況を受けて、IETFの
ROHC（Robust Header Compression）
WGでの活動など、パケットロス耐性
のあるRTPヘッダ圧縮方法の検討が進
められている。ROHC WGで検討され

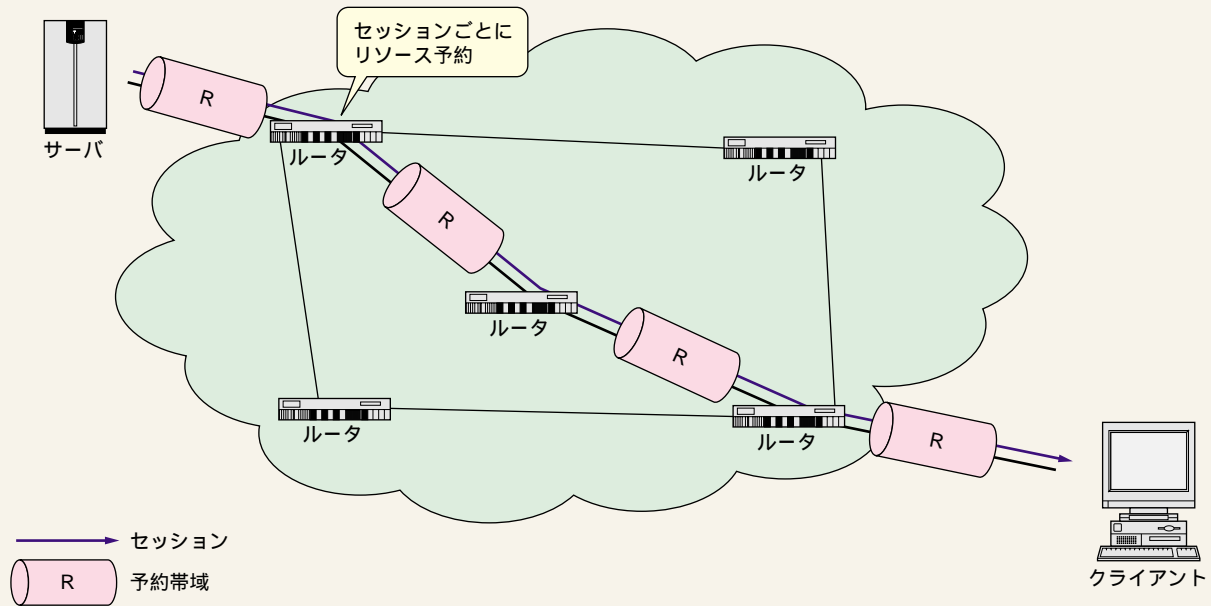


図3 IntServによるQoS制御

ているヘッダ圧縮方法は、以下の特徴により、パケットロス耐性を高めている。

- ① ある特定のフィールド値にイレギュラーな変更があった際に、該当フィールドの差分値そのものではなく、フィールド値の下位数ビットを複数回伝達することにより、1つ前のパケット以外からも復元を可能とする。
- ② 圧縮前のRTP/UDP/IPヘッダから算出されたチェックサムを圧縮ヘッダに付加し、受信側で当該チェックサムを用いて正しく復元できたかどうかをチェックする。
- ③ 完全なパケットロス耐性を要求する場合には、受信側から送信側へACK (Acknowledgment) を返送し、送信側ではACK対象のパケットから復元を可能とするだけの情報を圧縮ヘッダに付加し、送信する。

これらに加え、送受信局間で同期はずれが生じた際でも同期回復後に逆向きにヘッダ復元をすることでパケットロス数を低減する手法をドコモからROHC WGへ提案し、本提案も含めて

規格化される予定である。

2.3 QoS制御

現状のインターネットはベストエフォート型のネットワークであり、スループットや遅延、パケットロスなどQoSの制御は何もしていない。しかし、スループットや遅延要求の厳しいマルチメディアコンテンツの伝送にはなんらかのQoS制御が必要となる。そのため、QoSを制御するネットワークアーキテクチャが数多く検討されており、中でもIntServ (Integrated Service) とDiffServ (Differentiated Service) が注目を集めている。

IntServはRSVP (Resource Reservation Protocol) [3]というシグナリングプロトコルを用いてセッションごとにバッファや帯域などのネットワーク資源を予約し、この予約資源により所望のQoSを満足するネットワークモデルである(図3)。IntServによりEnd-to-Endの遅延を保証するGuaranteed Serviceなどの提供が可能となるものの、RSVPのシグナリングトラフィックが多いことや各ルータですべてのセッション状態を管理する必要があるなど

の問題点が指摘されている。

一方DiffServ [4]は、IntServのように各セッション単位でQoS制御をするのではなく、あらかじめ限定された数のサービスクラスごとにQoS制御をおこなうため、スケーラブルなネットワークモデルとなっている。図4にDiffServにおけるQoS制御を示す。DiffServでは、DiffServを提供するネットワークドメイン (Diffservドメイン) の端に位置するエッジルータにおいて、送信元ホストとのサービスレベル契約 (SLA: Service Level Agreement) や各パケットのヘッダ情報などから当該パケットのサービスクラスを決定し、IPv4ヘッダのTOS (Type of Service) フィールドまたはIPv6ヘッダのTraffic Classフィールド (これらをDSフィールドという) にサービスクラスをマークする。DSドメイン内では同一のサービスクラスのセッションは集約され、サービスクラスごとに異なる転送処理 (PHB: Per Hop Behavior) が適用される。

これによりサービスクラス単位で異なるQoS制御を実現することができる。現在、IETFのDiffServ WGにおい

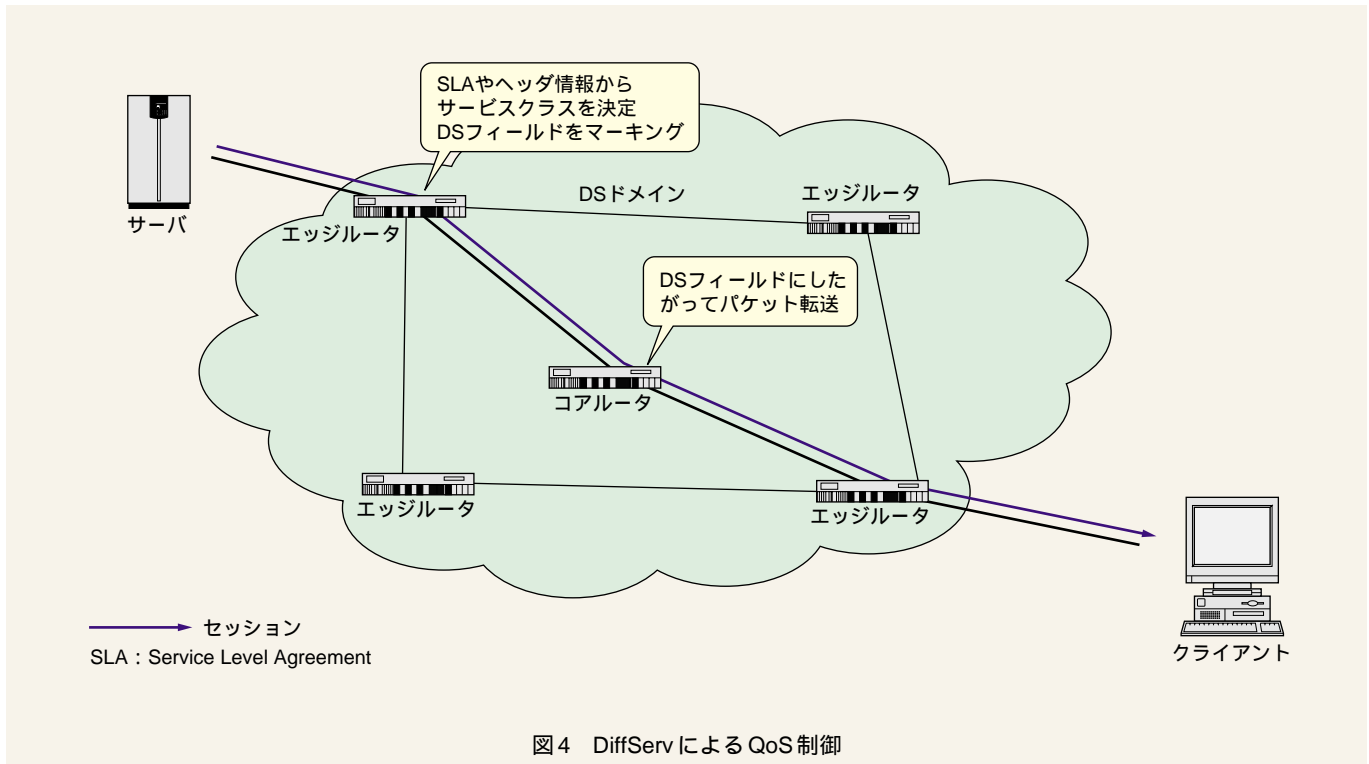


図4 DiffServによるQoS制御

表1 RTSPメソッド一覧

メソッド	説明
SETUP	リソース確保, セッション開始
PLAY	再生開始
PAUSE	一時停止
TEARDOWN	リソース解放, セッション終了
RECORD	記録開始
DESCRIBE	デスクリプション取得
ANNOUNCE	デスクリプション更新 (セッション中のメディア追加など)
OPTIONS	使用可能なオプション問合せ
REDIRECTION	他サーバ接続指示 (負荷分散, コンテンツ移動など)
GET_PARAMETER	パラメータの取得
SET_PARAMETER	パラメータの設定

するクライアントから制御するプロトコルである。RTSPクライアントはメソッドを記述したリクエストメッセージをサーバへ送信し、サーバはレスポンスメッセージを返送するとともにメソッドに対応する処理を行う。RFC2326で規定されているメソッドを表1に示す。これらのメソッドに加えて、再生時間や再生速度などをメッセージヘッダで記述することにより、さらに詳細な指示をすることができる。

図5にRTSPのシーケンス例を示す。クライアントはDESCRIBEメソッドを含むリクエストメッセージを送信すると、サーバは対象となるメディアのセッション記述をSDP (Session Description Protocol) [6]などを用いて返送する。そして、クライアントがSETUPメッセージを送信することにより、サーバのリソースが確保され、同時にセッションが開始する。続いてPLAYメッセージによりメディアデータの伝送が開始され、PAUSEメッセージによりメディアデータ伝送が中断される。最後に、TEARDOWNメッセージによりサーバのリソースが解放され、セッションが終了する。

て、低遅延・低ジッタ・低パケットロスでのパケット伝送を提供するEF (Expedite Forwarding) クラスや、4つの異なる優先度クラスと各優先度クラスごとに3つの廃棄クラスを有するAF (Assured Forwarding) クラスが規定されている。

3. マルチメディア配信制御技術

3.1 セッション制御 音楽や映像などマルチメディアコン

テンツの視聴の場合、ファイルのダウンロードとは異なり、途中からの再生や一時停止、さらには早送りやスロー再生などユーザからのさまざまなリクエストをリアルタイムに処理することが望まれる。

そのようなマルチメディアセッションの制御を可能とするプロトコルとしてRTSP (Real Time Streaming Protocol) [5]がIETFのRFC2326で規定されている。

RTSPは、サーバからのマルチメディアコンテンツの伝送を、遠隔に位置

も注目されている。SDPは3.1節で述べたRTSPの中で送受信間でのメディア再生能力の交換に用いられるほか、SMILの次期バージョンの中でメディア記述言語の一部として統合される動きもある。

3.3 ファイルフォーマット

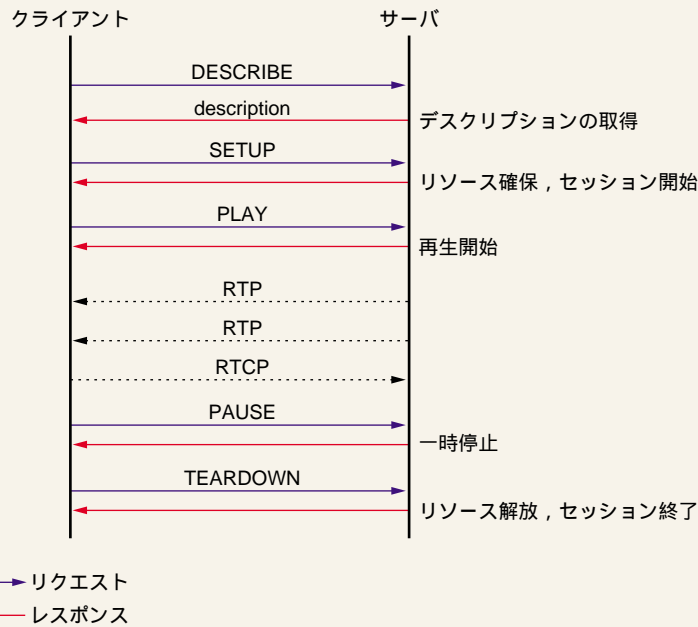
マルチメディア配信では、マルチメディアコンテンツをサーバにファイルとして蓄積しておき、それをユーザの要求に応じて伝送するという手順となる。したがって、サーバに蓄積するファイルのフォーマットも、伝送プロトコルと同様に規定される必要がある。マルチメディアコンテンツのケット伝送のプロトコルは、2章で述べたように伝送目的に特化したプロトコルを用い、ビデオ、オーディオなどの情報が同期のための情報が付加されて独立にケット化されて伝送されるのが一般的である。このため、サーバに蓄積するファイルフォーマットには

- ① マルチメディアコンテンツの同期情報保持
- ② 伝送プロトコルの規定する形式へのケット化の容易性
- ③ マルチメディアコンテンツの多重化（単独のファイルでの複数メディアコンテンツの保持）

が必要となる。

ISO/MPEGでは、このような要求を満たすファイルフォーマットの標準として、MPEG-4ファイルフォーマット（MP4 File Format）をMPEG-4システムバージョン2で規定している。本フォーマットは、上記要求条件のうち特に②）を重視した構成となっており、メディアデータは‘mdat’と呼ばれる領域にフリーフォーマットで格納、メディアデータ間の時間間隔、メディアデータサイズ、サンプル数、およびファイル先頭からのオフセットなどの情報は、‘moov’と呼ばれるヘッダ領域に格納される。

この他に同様の目的をもったデファクトの規格としては、ASF（Advanced



RTP : Real-time Transport Protocol

図5 RTSPシーケンス例

3.2 メディア記述

マルチメディアの配信はインターネット上のHTML（Hyper Text Markup Language）コンテンツをきっかけとして普及してきた。今後は、オーディオビジュアルを含むあらゆるメディアを統合してプレゼンテーションを行うためのメディア記述言語の重要性が高まっている。インターネットにおけるメディア記述言語は主としてW3C（World Wide Web Consortium）にて策定されており、XML（Extensible Markup Language）[7]をベースとして柔軟かつ拡張性の高い言語仕様へ進化しつつある。HTMLはXMLベースのXHTMLへ置き換えられ、HTMLにはもともと備わっていなかったメディアプレゼンテーション機能を他の言語から補完して利用できるようになる。マルチメディアという観点から重要なのはSMIL（Synchronized Multimedia Integration Language）[8]であり、コンテンツのメディア間同期の規定やコンテンツ視聴環境に応じて選択可能なコンテンツ選択肢記述などの機能を備える。

すでに実用化されているSMIL1.0勧告に続いて、次バージョン（SMIL2.0）ではXHTMLとの統合やモバイル用サブセットなどの検討が進められている。

一方、ISO（International Organization for Standardization）/MPEGでも、MPEG-4において、メディアオブジェクトのプレゼンテーション記述としてBIFS（Binary Format for Scene）が規格化されている[9]。これは、VRML（Virtual Reality Modeling Language）をベースとしており、シーン空間中におけるオブジェクトの時空間配置をバイナリ表現する。ユーザインタラクションを取り入れたマルチメディアプレゼンテーションが特徴である。BIFSをXML表記するとともにSMILのコンポーネントなどと統合したテキスト記述フォーマットの検討も現在MPEGで進められている。4章に述べるMPEG-7もXML表記が基本であり、上記のようなXMLベースのメディア記述言語と統合して用いられる可能性がある。

このほか、メディアの送受信能力のネゴシエーション情報としてSDP[6]

Streaming Format, Microsoft社), QuickTime (Apple社) などがある。

4. MPEG-7

4.1 概要

ISO/MPEGでは、MPEG-4標準化に引き続き、MPEG-7 [10]の標準化活動を行っている。MPEG-7では、従来のMPEGシリーズのようなメディアの符号化方式から離れ、マルチメディアコンテンツへのアクセス(検索、フィルタリング、ブラウジング)を効率化する目的に使用されるメタデータフォーマットの標準化を行っている。主なマイルストーンとしては、2001年3月にFCD(最終委員会草案)を発行し技術仕様を凍結する。

標準化完了は2001年9月を予定している。MPEG-7で標準化されたメタデータをコンテンツに関連付けることにより、さまざまなプラットフォームに分散配置されるコンテンツへのアクセスを標準的な手順に従って実行可能となり、コンテンツの利用価値を高めることが可能となる。

具体的に標準化される項目を図6に示す。MPEG-7では、メタデータをXML文書としてインスタンス化することを基本としており、各メタデータのフォーマット(またはシンタックス)を定義する言語としてDDL(Description Definition Language)を標準の一部として規格化する。DDLはW3Cで標準化作業中のXMLスキーマをベースとする。

また、メタデータの種別としてD(Descriptor)、DS(Description Scheme)を標準化する。Dは主として信号レベルで捉えられるメディアの特徴(feature)を記述するものであり、DSは種々のDまたはDSを組み合わせることによってメディアの持つ時空間的構造(structure)を記述しようとするものである。Dの事例としては画像の色ヒストグラム、DSの事例としてはビデオプログラムの構造記述(シーン、ショットなどの定義)などが挙げられる。一般に、Dは画像検索や類似コンテンツの分類などへの利用が想定されている。DSは長時間のコンテンツに対して必要な部分へのアクセス

(頭出し)を効率化したり、ユーザの好みや視聴環境(端末性能やネットワークQoSなど)に応じてコンテンツの見せ方をカスタマイズするなどの用途が考えられている。

また、DやDSを実際に運用する場合にシステムとして必要な要素はSystemsパートとして規格化される。たとえば、D/DSの伝送効率を向上するためのバイナリ表現、アクセスユニットなどの単位でD/DSを伝送するための規定などがこれに該当する。

4.2 主な技術要素

D/DSの記述対象範囲は非常に多岐に渡る。詳細なD/DSの定義は文献[10]などに譲るが、MPEG-7の実用化にあたって技術的なポイントとなるのはD/DS記述の獲得技術と、実際のアプリケーションにおけるD/DSの利用技術である。これらはいずれも標準化対象外であるため、実装に依存する。D/DSの獲得については膨大な情報量のコンテンツに対してメタデータを付与するためにメディア特徴や構造などの自動抽出処理が必要となる。たと

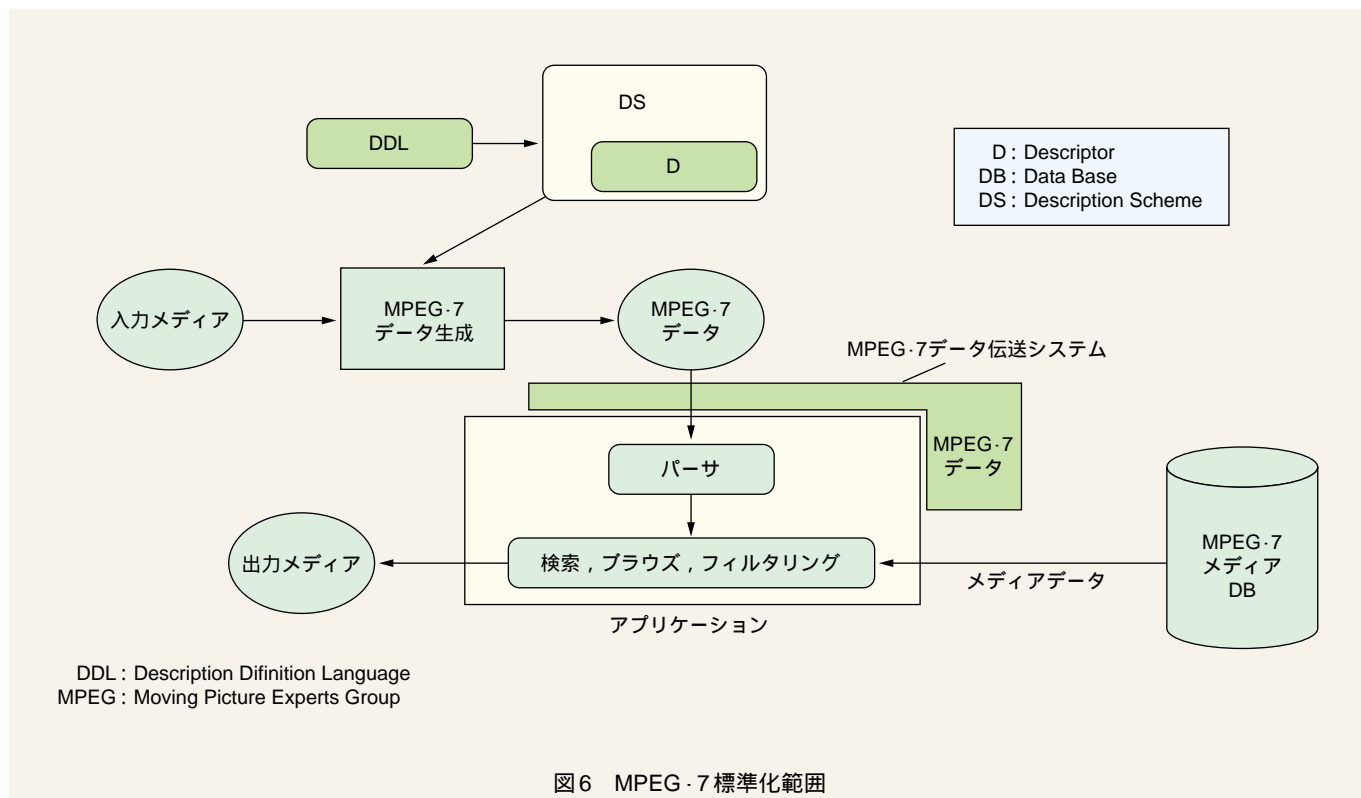
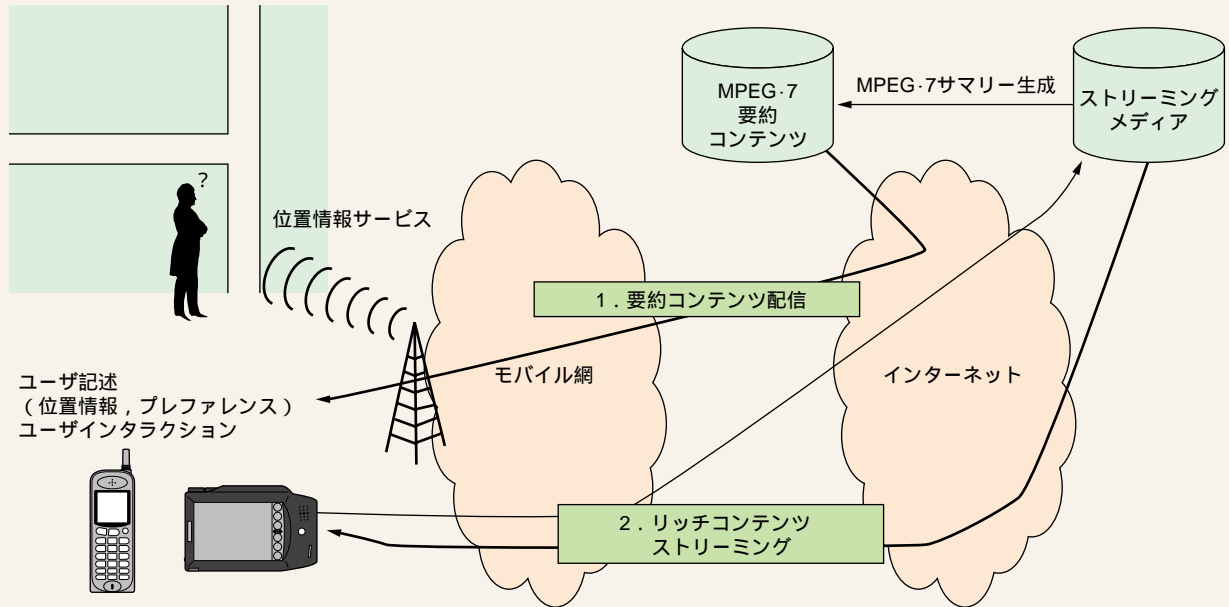


図6 MPEG-7標準化範囲



MPEG : Moving Picture Experts Group

図7 MPEG-7のモバイルマルチメディア配信応用事例

ば映像の構造を抽出するために映像中のキーフレームを自動検出する[11]などの処理が必要である。また、得られたD/DSを有効に活用するアプリケーション・システムの開発も重要となる。画像検索システムにおいては、システムの要求条件に合致する検索結果を導くためのD/DSマッチング処理アルゴリズムが必要となる。また、1ソース・マルチユース環境の実現においてもMPEG-7は重要な役割を果たす。付加されたMPEG-7記述をヒントにして、コンテンツをさまざまな異種インフラへ自動的に適応化して配送する技術などが重要となる。

4.3 マルチメディア配信への応用

MPEG-7のモバイル活用としては、端末のメディア表示・再生能力、ネットワーク資源の制約の中でマルチメディアコンテンツを効率良く配信するサービスへの応用が考えられる。図7に具体的な活用事例を示す。本システムでは、MPEG-7メタデータを用いてマルチメディアコンテンツの要約を生成し、要約コンテンツだけをプッシュ配

信することによってユーザにコンテンツの概要を伝達する。今後は移動携帯端末のサービスとして位置情報の活用が期待されており、本システムにおいても端末の位置情報に応じて適切なコンテンツをプッシュするなどの応用が考えられる。ユーザはあらかじめ端末に入力済みのユーザプレファレンスと要約コンテンツが提供するリッチコンテンツのMPEG-7記述とに基づいて、ユーザの好みにカスタマイズされたコンテンツの部分視聴が行えるようになる。このような段階的なマルチメディアコンテンツの提供手段が与えられれば、ユーザもコンテンツプロバイダも必要な情報を必要に応じて提供・収集することができ、モバイルネットワークにおけるマルチメディア配信の促進が期待される。

5. あとがき

本稿ではマルチメディア配信に関わるトランスポート技術や制御技術、そしてMPEG-7を紹介し、標準化動向についても簡単に触れた。マルチメデ

ィアトランスポート技術として、リアルタイムメディアの伝送を目的としたRTPやヘッダのオーバーヘッドを削減するロバストヘッダ圧縮、さらにはDiffServなどのQoS制御技術について解説した。また、マルチメディア配信制御技術として、遠隔からのセッション制御を可能にするRTSPや、メディアプレゼンテーション機能を提供するSMIL、サーバに蓄積するためのファイルフォーマットなどを紹介した。最後に、メタデータフォーマットを規定するMPEG-7の概要と標準化動向について説明するとともに、マルチメディア配信への応用について述べた。

モバイル環境においてマルチメディア配信を実現するためには、いずれも重要な技術であり、今後の動向が注目される。

文献

- [1] H.Schulzrinne, S.Casner, R.Frederick, and V.Jacobson: RTP: A Transport Protocol for Real-Time Applications, RFC1889, Jan.1996.
- [2] S.Casner and V.Jacobson: Compressing IP/UDP/RTP headers for low-speed serial links, RFC2508, Feb.1999.

- [3] R.Braden, L.Zhang, S.Berson, S.Herzog, and S.Jamin: Resource ReSerVation Protocol (RSVP), RFC2205, Sep.1997.
- [4] S.Blake, D.Black, M.Carlson, E.Davies, Z.Wang, and W.Weiss: An Architecture for Differentiated Services, RFC2475, Dec.1998.
- [5] H.Schulzrinne, A.Rao, and R.Lanphier: Real Time Streaming Protocol, RFC2326, Apr.1998.
- [6] M.Handley and V.Jacobson: SDP: Session Description Protocol, RFC2327, Apr.1998.
- [7] Extensible Markup Language (XML) 1.0, W3C Recommendation, 6 Oct.2000 (<http://www.w3.org/TR/2000/REC-xml-20001006>).
- [8] Synchronized Multimedia Integration Language (SMIL) 1.0 Specification, W3C Recommendation 15 June 1998 (<http://www.w3.org/TR/REC-smil/>).
- [9] Information technology - Coding of audio-visual objects - Part 1: Systems, ISO/IEC 14491-1.
- [10] Introduction to MPEG-7, ISO/IEC JTC1/SC29/WG11/N3545, July, 2000.
- [11] R.Brunelli et.al, " A Survey on the Automatic Indexing of Video Data ", Journal of Visual Communication and Image Representation 10, 78-112, 1999.

用語一覧

ACK : Acknowledgment	PHB : Pre Hop Behavior (転送処理)
AF : Assured Forwarding	QoS : Quality of Service (サービス品質)
AMR : Adaptive Multi Rate (適応マルチレート)	RFC : Request for Comment
ASF : Advanced Streaming Format	ROHC : Robust Header Compression
AVT : Audio Video Transport	RSVP : Resource Reservation Protocol (帯域保証)
BIFS : Binary Format of Science	RTCP : RTP Control Protocol
CC : Contributing Source Count (寄与送信元カウント)	RTP : Real-time Transport Protocol
CRTP : ヘッダ圧縮	RTSP : Real-Time Streaming Protocol
DDL : Description Definition Language	SDP : Session Description Protocol
DiffServ : Differentiated Service (優先度処理)	SLA : Service Level Agreement
EF : Expedite Forwarding	SMIL : Synchronized Multimedia Integration Language
FCD : 最終委員会草案	TCP : Transmission Control Protocol
HTML : Hyper Text Markup Language	TOS : Type of Service
IETF : Internet Engineering Task Force	UDP : User Datagram Protocol
IntServ : Integrated Service	VRML : Virtual Reality Modeling Language
IP : Internet Protocol	W3C : World Wide Web Consortium
ISO : International Organization for Standardization (国際標準化機構)	WG : Working Group
MPEG : Moving Picture Experts Group	XML : Extensible Markup Language