An Evaluation of Session Initiation Protocol (SIP) for use in Streaming Media Applications

draft-whitehead-mmusic-sip-for-streaming-media-02

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on April 25, 2007.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

This draft summarizes a set of use-cases and their associated requirements that suggest a convergence between the Session Initiation Protocol (SIP) [2] and the Real Time Streaming Protocol (RTSP and RTSP v2) [3] and [4] that may be beneficial for streaming
media applications. This benefit is especially apparent in the context of converged/blended media services.

Table of Contents

1. Introduction ......................................................... 3
2. Terminology .............................................................. 4
3. Use Case Scenarios .................................................. 4
   3.1. Characteristics .................................................. 4
   3.2. Use Case Descriptions ......................................... 5
       3.2.1. Video Surveillance .................................... 5
       3.2.2. Blended services/videoconferencing ..................... 5
       3.2.3. Sharing a video with another person over a multi-media call ................................................. 6
       3.2.4. Allow access to personal/private video content ...... 6
       3.2.5. VOD services that requires resource or QOS-guarantees ...................................................... 6
       3.2.6. Settlement across provider boundaries ................. 6
       3.2.7. Intelligent selection of media encoding ............... 7
4. Required capabilities/Derived Requirements ............................ 7
   4.1. Scalability ....................................................... 7
   4.2. Signaling Latency ............................................... 7
   4.3. User identification, authentication and authorization ....... 7
   4.4. Accounting, charging, and settlements ........................ 8
   4.5. Server/client Location Discovery .............................. 8
   4.6. NAT and Firewall Traversal ................................... 8
   4.7. Session-based transport policy control ....................... 8
   4.8. Extensible with respect to application control signaling .................................................. 8
   4.9. Support media negotiation ....................................... 8
   4.10. Allow proxies ..................................................... 8
   4.11. Support media negotiation ..................................... 9
   4.13. Keeping DRM rights during a mobility session .............. 9
5. Exploring the Solution Space ......................................... 9
   5.1. SIP Features ..................................................... 9
   5.2. RTSP Features ................................................... 10
   5.3. Other Solutions ................................................ 10
   5.4. Analysis .......................................................... 10
6. Recommendations ...................................................... 11
7. IANA Considerations .................................................. 11
8. Security Considerations ............................................. 11
   Appendix A. Acknowledgements ...................................... 12
   Appendix B. Change History .......................................... 12
9. References ...................................................................... 12
   Authors’ Addresses .................................................... 13
   Intellectual Property and Copyright Statements .................... 14
1. Introduction

IP-based networks are continually improving in terms of bandwidth capacity and transport quality of service. At the same time, broadband services are continually expanding globally -- both in terms of reach and value-added. These developments are leading to an increase in the number and variety of deployment scenarios for streaming media applications. Many of these scenarios impose challenging new requirements on the signaling protocols used for these applications in terms of flexibility, scalability and network independence.

Historically, RTSP [3] and [4] has been the protocol of choice for streaming media applications and has covered both session control and media control. An obvious approach to address these new requirements then is to extend RTSP. This strategy appears to be able to address some of the new requirements, but not others. In particular extending RTSP to meet some of these new requirements would involve introducing protocol mechanisms that already exist elsewhere, namely in SIP and its associated extensions.

An alternative approach is to consider the possibility of using SIP for some of the functions needed by streaming media applications. While historically SIP has been used for communication services, the protocol itself is flexible enough (by design) to signal a wide variety of media streams. Moreover, driven in large part by the requirements associated with IP-based communication services, SIP has been extended over the years to address many of the same requirements currently facing next-generation media streaming applications. Rather than reinvent or duplicate protocol mechanisms in RTSP that already exist in SIP, a reasonable strategy may be to find a way to use SIP instead or in conjunction with RTSP.

This document presents some of the use cases that suggest a convergence between SIP and RTSP. These will also be eventually used to derive requirements on the service signaling protocol. Then, high-level strategies will be defined.

The goal of the draft is not propose all SIP or SIP/RTSP solutions but to considers possible ways in which RTSP is not sufficient for some streaming media applications and where SIP could fill the gap. The purpose of the document is to give a list of use-cases (section 3), a list of derived capabilities (section 4)*, a rationale for considering SIP in conjunction with RTSP (section 5), and recommendations for future work (section 6). The technical solution options may be investigated in a future draft based on the current document.
2. Terminology

In this document, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in BCP 14, [1] and indicate requirement levels for compliant implementations.


3. Use Case Scenarios

The scope of scenarios for this document includes applications with the following characteristics: content-on-demand, streaming media, unicast-media streams, live or recorded content, ubiquitous access (any-device, any-access).

While of interest, non-streaming media applications, such as downloaded media services, are outside the scope of this document.

3.1. Characteristics

For the purposes of this document, the term ‘controlled streaming media application’ represents a class of applications with the following characteristics:

- Multiple servers that can be a source of content but showing up as a single muxed stream at the client.
- One or more clients can receive the content.
- The media stream(s) needs to be delivered isochronously, in the most common case: the client intends to begin rendering the media before delivery is complete.
- Less common but equally valid, the server does not have resources to buffer content until the client is ready to receive it, e.g., a live feed.
- A session exists between source (e.g. server or peer) and destination (e.g. client or peer).
- The session is established, managed, and terminated through the use of a signaling protocol, in which control messages are exchanged (either directly or indirectly) between the source and the destination referred to as 'session signaling'.
- The application supports media stream control.

The client(s), or a proxy element acting on behalf of the client(s), has the ability to manipulate the media stream (or other aspect of the application) via signaling. This is referred to as application-signaling (or media control signaling).
3.2. Use Case Descriptions

As IP-based broadband data services have continued to develop and expand, opportunities for streaming media applications have also proliferated and expanded beyond the traditional framework. This section describes several streaming media application use case scenarios. These scenarios illustrate the variety of conditions and environments in which streaming media applications need to operate.

Use cases are used with the purpose of clarifying the ‘streaming media application’ and to explore the application space. The objectives are to:
- Clarify the frame / scope the discussion.
- Illustrate some of usage scenarios.
- Identify some of the key attributes that characterize these use cases.

3.2.1. Video Surveillance

This is the first of a number of use cases that relate to "conversational video". In this use case a user wants:
- Either to switch from the unidirectional audio-video monitoring session into a 2-way (bi-directional) conversation when wanting to interact with an unknown visitor. In this case the session is "upgraded" from one-way to two-way.
- Or to switch from the unidirectional "live" received content to a unidirectional "recorded" content with a rewind "play trick command".

Instead of RTSP, SIP can be used to setup the communication, which provides the ability to switch from the one-way monitoring communication to the bidirectional communication. RTSP features related to trick plays can be used to remotely interact on the stream. For example the remote viewer should be able to place a rewind command to the previously recorded content. This is one use case for the convergence of both protocols.

3.2.2. Blended services/videoconferencing

In this use case, the user wants to switch between a "bidirectional" live conversation to a "unidirectional" recorded content. This use case is also for multiple services (streaming/communications/info) to use common signaling infrastructure. In this case SIP is again used for authentication billing and location. RTSP allows the remote viewer/user to place a "trick mode" rewind into a command videoconferencing context (to see what happened before joining in).
3.2.3. Sharing a video with another person over a multi-media call

This is another "conversational video" use case. A user already in an audio-video conversation (using a SIP based protocol) wants to provide local audio-video content to the called party. If the remote user wants to watch the live content (you should see what I am witnessing now) or request the content (can I look at the game you recorded yesterday?), a lot of the communication setup, generally established by RTSP, can be avoided with SIP the relationship (including authentication and billing) between the two parties is already established. This use case does not mean that the bidirectional "conversational" are switched to unidirectional "recorded" streams, but that "recorded" streams are to be added to the "conversational" streams.

3.2.4. Allow access to personal/private video content

In this scenario a user wants to remotely access personal content stored on a variety of media devices (watch at pre-recorded show from a mobile device at work). While the streaming of the content and the trick plays will use established RTSP functionality the use of SIP locator services, strong authentication and authorization as well as presence make this solution more feasible. For example, if the content is stored on a device located on a private network behind a FW/NAT, SIP via its name/location registration mechanism can be used to locate and connect to the device. In addition for this use-case, the access to a video content should make "session mobility" possible. In other words, when viewing a video content, it should be possible for the session of a user to seamlessly switch from one terminal (e.g., mobile phone) to another one (e.g., television).

3.2.5. VOD services that requires resource or QOS-guarantees

Consider a Video on Demand (stored video) service provided as a unicast session to an end user device from a server. The user requests a VOD movie. The VOD server determines the video that user wants to watch, and then contacts the appropriate network element (NE) and requests to reserve resources for the user and confirm back to the server. The problem is here that until this point, the NE can not fully estimate and negotiate the media resources needed for the whole duration of the VOD. SIP has pre-conditions that could be used; RTSP has no such functionality.

3.2.6. Settlement across provider boundaries

If a commercial VOD service is being offered by one party (e.g. a service provider) but receives carriage (transport) by one or more other parties (network providers) a mechanism is needed to allow
service signaling to exchange transaction identifiers for the purpose of charge correlation and settlements. SIP (and its associated extensions) supports these capabilities. RTSP at present does not.

### 3.2.7. Intelligent selection of media encoding

A user orders content to be delivered to its current device the content could exist in different format (e.g. standard definition or high definition) or encoding (MPEG2 or MPEG4 for example). Media negotiations may need to be informed by network transport capabilities. This is based on knowledge of access-network type.

### 4. Required capabilities/Derived Requirements

This section lists key requirements derived from the application use cases described in the previous section. The requirements are described in terms of the capabilities provided by a prospective solution.

#### 4.1. Scalability

Any solution must be able to accommodate:
- Millions of clients and servers.
- An individual server may need to support thousands of parallel sessions.
- An individual client may need to support a number of simultaneous sessions.

#### 4.2. Signaling Latency

Because the use cases refer to live content the latency budget is important for the user experience. Thus, the solution must support the following requirements
- The session negotiation should complete in a few seconds at most (TBC: need confirmation).
- The media control operations should complete in less than a second (TBC: need confirmation).

#### 4.3. User identification, authentication and authorization

In many targeted personal video streaming solution (including peer to peer) there is still a need for identifying source and destination, authenticate users and authorize access.
4.4. Accounting, charging, and settlements

In the case of commercial applications billing aspects need to be addressed. Billing aspects must also provide mechanisms to support charging correlation and settlements between one or more parties that collaborate to deliver the service.

4.5. Server/client Location Discovery

Location and discovery of the end point of a session is essential for personalization and targeted services.

4.6. NAT and Firewall Traversal

The solution must provide a way to traverse NATs and Firewalls. For example, when switching from a remote monitoring communication to a conversational communication, the NATs bindings should not be renegotiated.

4.7. Session-based transport policy control

The signaling mechanism should provide a means to insure that sufficient network resources are available to deliver the service at the desired quality of experience. In the event that sufficient resources are unavailable, the signaling mechanism should provide a means for denying the service request.

4.8. Extensible with respect to application control signaling

(Support many different application types)

...words to be added...

4.9. Support media negotiation

The solution must provide a way to negotiate all media sessions (e.g.: conversational and streaming) as a whole, as described in the RFC3264 with the offer/answer so that both parties can estimate the media resources involved in the session.

4.10. Allow proxies

Any solution to be deployed in a large network and provide adequate scalability should support proxies. This also enables aggregation points for enhanced services.
4.11. Support media negotiation

The signaling protocol should support the following capabilities with respect to negotiating media flows:
- Support for negotiating per-flow/media QoS and bandwidth requirements.
- Ability to add, delete, and modify media flows to a session.
- Ability to support both uni-directional and bi-directional flows in a single session.

4.12. Support auto-configuration/installation

The solution should enable per user and per device installation and configuration. This may require device discovery and user authentication.

4.13. Keeping DRM rights during a mobility session

The signaling protocol should support the following capabilities with respect to maintaining DRM during session mobility:
- Support for negotiating per session and per device DRM.
- Ability to broker between different DRM models should the content require it.

5. Exploring the Solution Space

The set of use cases outlined above present a compelling argument for considering SIP for establishing "conversational video" communications or for those sessions that potentially become conversationsational at some point. The use cases also show the need to keep established media controls. This section reviews protocol features and analyses the use cases vis-a-vis the features.

5.1. SIP Features

SIP supports a rich set of capabilities that are useful in the context of streaming media applications. In particular, SIP has the following properties that can be leveraged in streaming sessions:
- Acts as a rendezvous protocol (with many capabilities).
- Carries the session description protocol.
- Supports invitation to unicast or multicast sessions.
- Supports SDP with the Offer/Answer Model.
- Works with NAT via ICE for SIP.
- Supports unidirectional and bidirectional communications e.g., can switch to 2-way or mix with streaming services.
o Supports a set of P-headers useful in the context of commercial service settings, for example:
  * Charging-ids: useful for 3rd party content providers.
  * Access-network headers: useful for inferring proper content encoding.

5.2. RTSP Features

RTSP is widely used for controlling the playback of media flows such as the delivery of webcam monitoring content, the delivery of Video on Demand and the delivery of IPTV. RTSP as defined in [3] has the following properties that are important to build on for any streaming solution:
- Acts as a lightweight rendezvous protocol.
- Supports trick plays and media control (pause/rewind/forward/...).
- Carries and interprets the session description protocol.
- Supports invitation to unicast or multicast sessions.
- Is a recognized standard for streaming applications.

5.3. Other Solutions

Other protocols have been proposed to control media flows and resources recently the Media Resource Control Protocol (MRCP) [5]. MRCP provides the requests, responses, and events needed to control the media processing resources. Hence is ideal for voice sessions and eventually multimedia. But MRCP relies on the Real Time RTSP to establish and maintain the session. Hence the solution proposed by MRCP could eventually become complementary to a SIP/RTSP solution.

... more inputs required - to be completed ...

5.4. Analysis

There are obvious overlaps between SIP, MRCP and RTSP and there are standalone solutions that use one or the other. The use cases as well as requirements of the previous sections do highlight the need for some form of collaboration between SIP and RTSP in particular. This is also reflected by the convergence in industry between essentially the telecommunications (where signaling SIP is dominant) and entertainment (where RTSP streaming is the favored solution).

The use-cases clearly show that most of the time the choice is not SIP only or RTSP only or another protocol only but that the advantages of more integration between protocols lead to more robust solutions and a richer solution space. The use cases also show that that the session setup negotiations are usually independent of the media controls. Hence in many of the use cases SIP could be used to replace the RTSP SETUP and DESCRIBE methods, leaving RTSP to be used
for media control.

There has already been work [6] on defining the integration. This opens the way for new service development especially for entertainment. Open issues remain however. For example there is currently no consensus on whether the SETUP and DESCRIBE methods should be kept or not. If the SETUP and DESCRIBE methods are not used, then the number of message round-trips is less important and the integration with an offer/answer mechanism is feasible. This implementation is described in [6]. If the SETUP and DESCRIBE are still used, then no RTSP header field parameters need to be conveyed within the offer and answer and the integration with legacy RTSP servers may be easier and not require either further development or the use of gateways.

It is also important to note that other Standards Developing Organizations (SDOs) like the Alliance for Telecommunications Industry Solutions ATIS, the Digital Video Broadcast (DVB), the European Telecommunications Standards Institute (ETSI) TISPAN Next Generation Networks (NGN) CableLabs to name a few, have standardized streaming with RTSP and adopted SIP for offer/answer services. RTSP has been widely adopted, has been proven successful and provides a good "running code and rough consensus" reason to keep it for media controls within SIP sessions. This does not mean that this will not change in the future but gives an impetus for the current work to investigate the SIP/RTSP integration further.

6. Recommendations

We propose that further work be initiated to define further how to signal streaming media sessions using SIP based on the use cases defined in this document and the solution space identified. We propose to reuse RTSP as a control stream negotiated by SIP/SDP as to keep compatibility with existing streaming solutions both in the entertainment and entreprise spaces. This works has been already initialized in [6].

7. IANA Considerations

The RTSP 'encoding format' and the new media attributes may need to be registered.

8. Security Considerations

No rogue 3rd party should be allowed to get access to the SIP
identity and use it to setup an un-authorized RTSP session.

**Appendix A. Acknowledgements**

The authors would like to acknowledge those who provided valuable inputs for this document namely Darren Loher, C. Steck, Osher Hmelnizky, Jonathan Rosenberg, David Ress, Ravishankar Shiroor, Martti Mela and Xupei Li.

Thank you also to JK Muthukumarasamy, Jim Baratz and Sam Ganesan for many emails and personal discussions.

**Appendix B. Change History**

v01
- Removed sections on particular solutions
- Refined the use cases sections and the scenarios
- Added recommendations based on discussions at and after IETF 65.

v02
- Added discussion on solution space following suggestions at and after IETF 66
- Added reference to companion SDP draft
- Added requirements when needed
- Fixed typos and wording as appropriate

9. References


Authors’ Addresses

Steven Whitehead
Verizon Laboratories Inc.
40 Sylvan Road
Waltham, MA  02451
USA

Email: xavier.marjou@francetelecom.com

Marie-Jose Montpetit
Motorola Connected Home Solutions
55 Hayden Avenue, Suite 3000
Lexington, MA  02421
USA

Email: mmontpetit@motorola.com

Xavier Marjou
France Telecom
Rue Pierre Marzin
Lannion  22300
France

Email: xavier.marjou@orange-ft.com