In some cases, when discussing a specific desired functionality, this document refers to SIP and complementary technologies such as protocols for floor control, camera control, etc. for achieving other functionalities. It is important to emphasize that while much application level functionality exists, a basic tool is needed to address the specific needs of the particular application. A basic tool should be effective and convenient for general use in its area of expertise, so that it would be appropriate to build on the results of emerging applications.

In some cases, when discussing a specific desired functionality, this document refers to SIP and SIP/SDP/RTP/RTCP suites of protocols to enhance SIP in order to add new required functions or to provide the same functionality in a more efficient way.

In other cases, when discussing a specific desired functionality, this document refers to SIP and SIP/SDP/RTP/RTCP suites of protocols to enhance SIP in order to add new required functions or to provide the same functionality in a more efficient way.
used in. This information is presented to highlight some possible obstacles and interoperability problems that need to be considered on the way towards the desired networks convergence.

2. Multimedia Application Requirements

2.1 General

Each device has a set of its capabilities in terms of its CPU processing power, additional HW characteristics and the algorithms it supports. During a session's lifetime, its characteristics may be changing dynamically both as a result of network conditions and as part of a broader application (such as interconferencing).

All the requirements, presented in this chapter, are required for providing basic reliable services: meaning establishing a session of a certain quality, based on capabilities agreement during the session establishment and sustained throughout the duration of the session.

The problem can be described as a lack of expressiveness in three following areas:

- Capabilities Specification
- Resource Reservation
- Media Stream Control

The desired ultimate goal is to:

- Express the capabilities (i.e. supported media, CODEC algorithms, bandwidth, etc.) without a need in configuration
- Signal the total resources, required for a specific session (probably in terms of capabilities)
- Within a session, explicitly open and close media streams, modify the parameters of a certain stream within the boundaries of previously announced capabilities and reserved resources

2.2 Capabilities Specification

2.2.1 Bit Rate

SDP [2] has a concept of "Application-Specific Maximum Bandwidth" that can be applied to "m" (media) line and specified in kilobits per second.

The required functionality is to express CODEC capabilities for both ranges of rates and a discrete number of rates. The "discrete number of rates" requirement has a number of purposes:

- Efficient resources allocation
- Multiple applications where the rates from different sources should be matched
- Interworking with terminals using multiplexing schemes, such as H.320 [12] and H.324 [13], in which only a discrete number of bit rates are available.

The defined capabilities may be used both for resource reservation and the actual control of a media stream.

2.2.2 Advanced CODEC schemes support

Video CODEC algorithms (among them H.263 [9]) have a concept of CIF (Common Intermediate Format Definition) with its derivatives: QCIF, 4CIF, etc. This definition of resolution implies the number of pixels and the format of the composed picture.

The challenge of providing quality video sessions over imperfect networks results in inventing of new coding algorithms with numerous operational modes fitting various environments. A famous example is the latest H.263 Recommendation ("H.263+") [10] that specifies a coded representation that can be used for compressing the moving picture component of audio-visual services at low bit rates and has numerous number of standard options described in its Annexes.

All of the mentioned characteristics (bit rate, resolution, H.263+ options) are examples of CODECs capabilities. Many CODEC implementations are capable of changing the mode of their operation in real time, as a result of changing conditions, and signaling the new mode within the RTP payload header.

RTP has a definition of a specific RTP Payload header for each CODEC scheme carrying both the configuration and the dynamically changing segmentation information describing the format of the transmitted RTP packets.

In order to use various options dynamically during a session lifetime, without possibility of changing a call, a "capability announcement" mechanism should exist, empowering modes of operation, supported by both of the sides.

In the future, SDPng work [22] may define a Language, expressive enough to describe different modes of
operation. SIP may be expanded to carry some of the functions. SIP extensions may be needed to support the "capabilities announcement" phase from the actual opening of media streams.

2.2.3 Lip Synchronization

One of the basic requirements for a multimedia session is the synchronization between audio and video streams when presenting them to the user.

2.2.3.1 Skew

The different timing of two media streams usually derives from an unequal processing time required for the encoding of the streams by the originating device. This difference is referred to as a skew. There is defined as a maximum time that the two media streams are delayed from each other as delivered to the transport network. Skew is usually measured in milliseconds.

Using RTP ([1]) timestamping services, during an active multimedia session the receiving device is capable of computing the skew and adjusting its buffers accordingly to provide the end user video and audio display in a synchronized manner.

Additional useful functionality, helping to deal with the lip synchronization problem, is providing the receiving device with a skew metric before the actual media streams are transmitted. Knowing the maximum skew value in advance allows the receiving device, based on the device's display capabilities, to adjust its buffers in order to avoid a playback that is out of sync.

Using the "reverse" RTCP control packets, as in [2] and commands defined for certain CODECs and carried by SDP/RTP/RTCP definitions, it is possible to express "resources reservation requirements or CODECs multi-rate requirements for a high quality call and pleasant end user experience. Allowing a receiving device to adjust its buffers accordingly increasing probability to provide the end user video and audio display in a synchronized manner.

Currently, in SIP/SDP/RTP/RTCP systems the media streams are transmitted. Knowing the maximum skew value in advance allows the receiving device, based on the device's display capabilities, to adjust its buffers in order to avoid a playback that is out of sync.

2.3 Resource Reservation

As end users may support more than one CODEC scheme for a specific media type. The initial reason for choosing a specific CODEC is to match the CODEC scheme, supported by the participants, when negotiating the session. Therefore, it is crucial that all peers explicitly choose a certain combination of audio and video CODECs without exceeding CPU processing power limitations or certain available network bandwidth.

Using the same system with a defined (CPU processing power and a support for services) CODEC scheme, you would have to configure your system in order to properly deliver the media stream to the receiving application. As one time you might want to receive a classical concert program, at another time you might want to be able to participate in a medical surgery session. In this example, the application residing on your computer represents its CODEC capabilities, CPU constraints and, possibly, the local network limitations. The other side, once it receives the "registration" of its supported CODEC capabilities, CPU constraints and, possibly, the local network limitations, allows it to make the appropriate requirements for providing a service with a certain quality. One of the possible ways is to express "resource reservation ability" by grouping the "capabilities", discussed above. It is an open question, if SIP, as a peer call control protocol, requires actual "resources reservation commands" being imposed on the other side.

2.4 Media Stream control

This section presents requirements to effectively control a particular media stream unilaterally. These requirements do not refer to manually driven commands such as floor control or camera control.

It is interesting to mention that despite the fact that users commands are not in the scope of a call control protocol (such as SIP), a cumulative among them and the media streams should be taken into consideration during the design of such protocols. For example, a particular camera (managed by some application protocol) should be synchronized with its associated video stream, managed by SIP.

Video requires broader control than voice. Some examples of video specific control are listed towards the end of this section. It is believed that a convenient means for media streams control is required for both applications. Some applications obviously complicates the problem.

Currently, in SIP/SDP/RTCP/RTSP systems the media control commands are divided between SIP/SDP conventions and commands defined for certain CODECs and carried by "reverse" RTCP control packets, as in [4].

2.4.1 An ability to reference a specific media stream

The basic requirement is the ability to reference a particular media stream within a session.

This functionality does not currently exist within the SIP. It is needed to control a specific thing within a single media stream the whole times of media sessions (e.g. a camera lens has to be repositioned. Moreover, the remaining side has to perform matching against the old information for each one of the m lines in order to recognize the change.

This functionality should be fixed for the "next" SIP version, without affecting the results of existing work.

2.4.2 Effective addressing of media streams in multimedia sessions

In most cases, the change of media stream parameters is of no importance except for example [1]. Therefore, both, of the sides may issue a request for a conditioning change or command "cumulatively", generating a so-called "race-condition".

Currently, SIP solves this problem by introducing the "after-after" headers fields in the INVITE message. One of

Internet Draft SIP for Video February 22, 2001

Page 6
the disadvantages of this approach is that it locks the

whole session, when the collision exists in a certain
media stream only.

### 2.4.3 Explicit start and stop of data transmitting in a
certain direction

Explicit start and stop of data transmitting in a
certain direction is the first basic "control command"
out of set of media controls. Its necessity becomes
obvious during the design of PSTN interworking. Early
establishment of a media path in one direction only,
strict billing regulations are just a few of the
examples.

Today SIP addresses this functionality by "putting
media streams on Hold" by setting "c" destination
addresses to value "0.0.0.0".

### 2.4.4 Bandwidth changes

Applications involving video are particularly prone to
frequent bandwidth changes causing packets lost, error
conditions, etc.

The first cause for frequent changes is the network
changing conditions. Future IP based wireless networks
will become a real test bed for SIP services.

Similar changing conditions would frequently be caused
by the "multimedia nature" of the video services. Some
examples are presented below.

Today, in many integrated services, "multimedia
communications" includes a data service (such as T.120
[16]) bundled together with voice and video sessions.
Opening and closure of the data stream may significantly
change the desired parameters of the media streams.

The same effect may exist in multimedia applications,
where instead of using G723 derivatives, video streams
are provided to the users in separate windows. This mode
of operation may be advantageous to the user, who has a
better control over the session and can use network
resources in a more effective way.

Additional example of an "application condition" in a
multi-conferencing service, where adding of a participant

### 2.4.5 Video/CODEC Specific Commands

Various video specific techniques have been used in
today's networks in order to cope with the conditions
mentioned above with minimum service degradation and as
seamlessly as possible to the users. Below are some of
the examples:

H.261 and H.263 video CODECs have a notion of
picture's building blocks: "full picture", GOB and
MacroBlock (MB). The decoder would have as ability to
recognize synchronization degradation and explicitly
request from an encoder for a "full picture", a whole GOB
or a whole MacroBlock.

In SIP/SDP/RTP/RTCP systems, the only analogous
functionality is defined in RFC-2032 "RTP Payload Format
for H.261 Video Streams" [4] that defines a "Full INTRA-
frame Request" (FIR) to be carried in RTCP "reverse"
control packet. This technique is definitely an exception
to a normal RTCP design and therefore does not work in
cases where RTCP is used as an alternate transport
mechanism defined by SDPng [3].

A simple example of a video specific command is a
request to "freeze a picture" in this case originated
from the encoder towards the decoder. In case the encoder
is aware of oncoming massive changes in the transmitted
picture, it would request the decoding side to stop
presenting the changes, until a new stable image is
encoded and transmitted.

Another inherent example of a video specific command
is a request to change the tradeoff between temporal and
spatial resolutions, i.e. the tradeoff between the rate
of the samples and the resolution of the picture. This
request would be originated by the decoder towards the
encoder (if the encoder has the capability to dynamically
change this tradeoff).

### 2.4.6 Transmission of media stream commands

Today the media stream commands are transmitted by
STCP. "FIR" command defined for H.261 and described above

may result in a change of transmitted stream parameters
or in a reconsideration of resource capabilities.

### 2.4.6 Video/CODEC Specific Commands

Various video specific techniques have been used in
today's networks in order to cope with the conditions
mentioned above with minimum service degradation and as
seamlessly as possible to the users. Below are some of
the examples:

H.261 and H.263 video CODECs have a notion of
picture's building blocks: "full picture", GOB and
MacroBlock (MB). The decoder would have as ability to
recognize synchronization degradation and explicitly
request from an encoder for a "full picture", a whole GOB
or a whole MacroBlock.

In SIP/SDP/RTP/RTCP systems, the only analogous
functionality is defined in RFC-2032 "RTP Payload Format
for H.261 Video Streams" [4] that defines a "Full INTRA-
frame Request" (FIR) to be carried in RTCP "reverse"
control packet. This technique is definitely an exception
to a normal RTCP design and therefore does not work in
cases where RTCP is used as an alternate transport
mechanism defined by SDPng [3].

A simple example of a video specific command is a
request to "freeze a picture" in this case originated
from the encoder towards the decoder. In case the encoder
is aware of oncoming massive changes in the transmitted
picture, it would request the decoding side to stop
presenting the changes, until a new stable image is
encoded and transmitted.

Another inherent example of a video specific command
is a request to change the tradeoff between temporal and
spatial resolutions, i.e. the tradeoff between the rate
of the samples and the resolution of the picture. This
request would be originated by the decoder towards the
encoder (if the encoder has the capability to dynamically
change this tradeoff).

### 2.4.6 Transmission of media stream commands

Today the media stream commands are transmitted by
STCP. "FIR" command defined for H.261 and described above

may result in a change of transmitted stream parameters
or in a reconsideration of resource capabilities.

### 3. Interoperability with existing video systems

Today several protocols are defined to support
multimedia system both for Circuit Switched and Packet
networks. Part of them (such as H.323[11] and H.324M[12])
are deployed, others (such as H.320M) are intended to be
used in future networks.
It is important to be aware of the architecture of these systems and the interesting challenges they introduce. Below is the list of these specifications.

3.1 H.320

ITU-T Recommendation H.320 [12] "Narrow-band visual telephone systems and terminal equipment" is defined for use over ISDN networks. H.320 is especially successful in Europe. H.320 specifications define an umbrella system using the following protocols: H.221, H.245, H.263, H.231, H.120.

3.2 H.324 and H.324M


3.3 H.323


4. Conclusion

This draft is a first attempt to present and summarize issues needed for video services support in SIP/SDP/RTP/RTCP systems. Fast of the solutions may be defined in a short period of time. More advanced features or complicated problems will be resolved in the future by extending the SIP/SDP/RTP/RTCP

It is important to be aware of the current limitations or open issues in the standard, based on the impossibility of the requirements. SIP allows for extensions adding functionality in a standard interoperable manner.

An alternative possible approach might be the definition of conventions for certain SIP based multimedia systems.

5. Security Considerations

This document does not introduce new security requirements to existing SIP/SDP/RTP/RTCP systems.

6. References


7. Acknowledgements
Sasha Ruditsky, Yair Miranda, Ilia Jorsal, Iianan Shlad and Danny Levin participated in earlier discussions on this topic.

8. Authors' Addresses
Orit Levin
RADVision Inc.,
575 Corporate Drive Suite 420
Mahwah, NJ 07430
Phone: +1 201 529 4300
Email: orit@radvision.com

Full Copyright Statement
Copyright (c) The Internet Society (2000). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.