Network Working Group
INTERNET-DRAFT
Expires: April 2003

Magnus Westerlund
Ericsson

A Transport Independent Bandwidth Modifier for the Session Description Protocol (SDP).

<draft-westerlund-mmusic-sdp-bwparam-01.txt>

Status of this memo

This document is an individual submission to the IETF. Comments should be directed to the authors.

Abstract

The existing Session Description Protocol (SDP) bandwidth modifiers and their values include the bandwidth needed also for the transport and IP layers. When using SDP in protocols like SAP, SIP and RTSP and the involved hosts reside in networks running different IP versions, the interpretation of what type of lower layers that is included is not clear. This documents defines a bandwidth modifier that does not include transport overhead, instead additional packet rate attributes are defined. The transport independent bitrate value together with the packet rate can then be used to calculate the real bitrate over the actually used transport.

TABLE OF CONTENTS

1. Definitions

1.1. Glossary

RTSP - Real-Time Streaming Protocol

SDP - Session Description Protocol
SAP - Session Announcement Protocol
SIP - Session Initiation Protocol

TIAS - Transport Independent Application Specific maximum, a bandwidth modifier.

1.2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [7].

2. Changes

The following changes has been done to this version compared to draft-westerlund-mmusic-sdp-bwparam-00.txt:

- TIAS definition has been changed to not any longer include RTP headers in a RTP context.
- The TIAS value is now defined in bits per seconds instead of kilobits per second.
- Rules for how to use TIAS together with "maxprate" and "avgprate" has been defined
- The old prate attribute has been renamed "maxprate" and a new attribute "avgprate" has been defined.
- Use cases for TIAS has been defined
- Clarifications in the algorithm to derive transport dependent bitrates.
- The RTCP related problem is also included in the problem description and a solution proposed.

3. Introduction

Today the Session Description Protocol (SDP) [1] is used in several types of applications. The original application is session information and configuration for multicast sessions announced with Session Announcement Protocol (SAP) [2]. SDP is also a vital component in media negotiation for the Session Initiation Protocol (SIP) [3] by using the offer answer model [4]. The Real-Time Streaming Protocol (RTSP) [5] also makes use of SDP to declare what media and codec(s) a multi-media presentation consist of to the client.

3.1. The Bandwidth Attribute

In SDP there exist a bandwidth attribute, which has a modifier used to specify what type of bitrate the value refers to. The attribute has the following form:

b=<modifier>:<value>

Today there are four modifiers defined which are used for different purposes.

3.1.1. Conference Total

The Conference Total is indicated by giving the modifier "CT". The meaning of Conference total is to give a maximum bandwidth that a conference session will use. Its purpose is so that it is possible to decide if this session can co-exist with any other sessions. Defined in RFC 2327 [1].

3.1.2. Application Specific Maximum

The Application Specific maximum bandwidth is indicated by the modifier "AS". The interpretation of this attribute is depending on the application's notion of maximum bandwidth. For an RTP application this attribute is the RTP session bandwidth as defined in RFC 1889 [6]. The session bandwidth includes the bandwidth that the RTP data traffic will result in, including the lower layers down to IP layer. So the bandwidth is in most cases calculated over RTP payload, RTP header, UDP and IP. Defined in RFC 2327 [1].

3.1.3. RTCP Report bandwidth

Today there is a draft [9], currently in the RFC editors queue to become a Proposed Standard, which defines two new bandwidth modifiers. These modifiers "RS" and "RR", define the amount of bandwidth that is assigned for RTCP reports by active data senders respectively RTCP reports by receivers only.

3.2. IPv6 and IPv4

Today there are two IP versions 4 [15] and 6 [14] used in parallel on the Internet. This creates problems and there exist a number of possible transition mechanisms.

```
------------------            ----------------------
| IPv4 domain    |            | IPv6 Domain        |
|                | ---------| |                    |
| ----------     |-| NAT-PT |-|      ----------    |
| |Server A|     | ---------| |      |Client B|    |
| ----------     |            |      ----------    |
------------------            ----------------------
```

Figure 1. Translation between IPv6 and IPv4 addresses.

- To achieve connectivity between an IPv4 only host and an IPv6 only host translation is necessary. This translator can be for example Network Address Translation - Protocol Translation (NAT-PT) [13], see Figure 1. However to get connectivity for large number of protocols, Application Level Gateway (ALG) functionality is also required at the node. To be able to locate hosts through the translation node DNS ALG must be supported.

- IPv6 nodes belonging to different domains running IPv6, but lacking IPv6 connectivity between them solves this by tunneling over the IPv4 net, see Figure 2. Basically the IPv6 packets are put as payload in IPv4 packets between the tunneling end-points at the edge of each IPv6 domain.

```
---------------  ---------------  ---------------
| IPv6 domain |  | IPv4 domain |  | IPv6 Domain |
|             |  |-------------|  |             |
| ----------  |--||Tunnel     ||--| ----------  |
| |Server A|  |  |-------------|  | |Client B|  |
| ----------  |  |             |  | ----------  |
---------------  ---------------  --------------|
```

Figure 2. Tunneling through a IPv4 domain

IPv4 has minimal header size of 20 bytes. While the fixed part of the IPv6 header is 40 bytes.

The difference in header size results in that the bandwidth used for the IP layer is different. How big the difference is, depends on the packet rate.

4. The Bandwidth Signaling Problems

When an application wants to use SDP to signal the bandwidth required for this application some problems becomes evident depending on the transport layers.

4.1. What IP version is used

If one signals the bandwidth in SDP, with for example "b=AS:" for an RTP based application, one cannot know if the overhead is calculated for IPv4 or IPv6. An indication to which protocol has been used when calculating the bandwidth values is given by the "c=" connection data line. This line contains either a multicast group address or a unicast address of the data source or sink. The "c=" lines address type may be assumed to be of the same type as the one used in the bandwidth calculation. There seems to exist no specification pointing this out.

In cases of SDP transported by RTSP this is even less clear. The normal usage for a unicast on-demand streaming session is to set the connection data address to a null address. This null address does have an address type, which could be used as an indication. However this is not clarified anywhere.

Figure 1, illustrates a connection scenario between a streaming server A and a client B over a translator here designated as a NAT-PT. When B receives the SDP from A over RTSP it will be very difficult for B to know what the bandwidth values in the SDP represent. The following possibilities exist:

1. The SDP is unchanged and "c=" null address is of type IPv4. The bandwidth value represents the bandwidth needed in an IPv4 network.
2. The SDP has been changed by the ALG. The "c=" address is changed to IPv6 type. The bandwidth value is unchanged.
3. The SDP is changed and both "c=" address type and bandwidth value is changed. Unfortunately, this can seldom be done, see 4.2.

In case 1 the client can understand that the server is located in an IPv4 network and that it uses IPv4 overhead when calculating the bandwidth value. The client almost never convert the bandwidth value, see section 4.2.

In case 2 the client does not know that the server is in an IPv4 network and that the bandwidth value is not calculated with IPv6 overhead. In cases where a client reserve bandwidth for this flow, too little will be reserved, potentially resulting in bad Quality of Service (QoS).

In case 3 everything works correctly. However this case will be very rare. If one tries to convert the bandwidth value without further information about the packet rate significant errors will be introduced into the value.

4.2. Converting bandwidth values

If one would like to convert a bandwidth value calculated using IPv4 overhead to IPv6 overhead the packet rate is required. The new bandwidth value for IPv6 is normally "IPv4 bandwidth" + "packet rate" * 20 bytes. Where 20 bytes is the usual difference between IPv6 and IPv4 headers. The overhead difference may be other in cases when IPv4 options [15] or IPv6 extension headers [14] are used.

As converting requires the packet rate of the stream, this is not possible in the general case. Many codecs has many possible packet rates. Therefore some extra information in the SDP will be required. The "a=ptime:" parameter may be a possible candidate. However this parameter is normally only used for audio codecs. Also its definition [1] is that it is only a recommendation, which the sender may disregard from. A better parameter is needed.

4.3. Header Compression

Another mechanism that alters the actual overhead over links is header compression. Header compression uses the fact that most network protocol headers have either static or predictable values in their fields within a packet stream. Compression is normally only done on per hop basis, i.e. on a single link. The normal reason for doing header compression is that the link has fairly limited bandwidth and significant gain in throughput is achieved.

There exist a couple of different header compression standard. For compressing IP headers only, there exist RFC 2507 [10]. For compressing packets with IP/UDP/RTP headers CRTP [11] was created at the same time. More recently the Robust Header Compression (ROHC) working group has been developing a framework and profiles [12] for compressing certain combinations of protocols like IP/UDP, IP/UDP/RTP.

When using header compression the actual used overhead will be less deterministic but in most cases an average overhead can be determined for a certain application. If a network node knows that some type of header compression is employed this can taken into consideration. To be able to do this with any accuracy the application and packet rate is needed.

4.4. RTCP problems

When RTCP is used between host in IPv4 and IPv6 networks over an NAT-PT similar problems exist. The RTCP traffic going from the IPv4 domain will result in a higher RTCP bitrate than intended in the IPv6 domain due to the larger headers. This may result in up to 25% increase in required bandwidth for the RTCP traffic. The largest increase will be for small RTCP packets when the number of IPv4 hosts are much larger than the number of IPv6 hosts. Fortunately as RTCP has a limited bandwidth compared to RTP it will only result in a maximum of 1.75% increase of the total session bandwidth when RTCP bandwidth is 5% of RTP bandwidth. The increase in bandwidth will in most cases be less.

At the same time this results in unfairness in the reporting between an IPv4 and IPv6 node. The IPv6 node may report in the worst case with 25% longer intervals.

4.5. Future development

Today there is work in IETF to design a new datagram transport protocol, which will be suitable to use for real-time media. This protocol is called the Datagram Congestion Control Protocol (DCCP). This protocol will most probably have another header size than UDP, which is mostly used today. This results in even further numbers of possible transport combinations to calculate overhead for.

5. Problem Scope

The problems described in chapter 4 does effect all the protocols that uses SDP to signal bandwidth parameters which contains transport level information.

In the MMUSIC WG there is work on a replacement of SDP called SDP-NG. That work is RECOMMENDED to consider the problems outlined in this draft when designing solutions for specifying bandwidth in SDP-NG.

6. Requirements

A solution to the problems outlined in this draft should meet the following requirements:

- The bandwidth value SHALL be given in a way so that it can be calculated for all possible combinations of transport overhead.

7. A Solution

7.1. Introduction

This chapter describes a solution for the problems outlined in this document for the Application Specific (AS) bandwidth modifier.

The CT is a session level modifier and cannot easily be dealt with. To address the problems with different overhead the CT value is RECOMMENDED to be calculated using reasonable worst case overhead.

The RR and RS modifiers will hopefully be possible to clarify before the publishing of their specification.

7.2. The TIAS bandwidth modifier

7.2.1. Usage

A new bandwidth modifier is defined to at least be used for the following purposes:

- Resource reservation. A single bitrate can be enough to use for resource reservation. Some characteristics can be derived from the stream, codec type, etc. In some cases more information is needed, then another SDP parameter will be required.

- Maximum media codec rate. With the definition below of "TIAS" the given bitrate will mostly be from the media codec. Therefore it gives a good indication on the maximum codec bitrate required to be supported by the decoder.

- Communication bitrate required for the stream. The "TIAS" value together with "maxprate" can be used to determine the maximum communication bitrate the stream will require. By adding all maximum bitrates from the streams in a session together, a receiver can determine if its communication resources are sufficient to handle the stream. For example a modem user can determine if the session fits his modems capabilities and the established connection.

- Determine the RTP session bandwidth and derive the RTCP bandwidth. The derived transport dependent attribute will be the RTP session bandwidth in case of RTP based transport. The TIAS value can also be used to determine the RTCP bandwidth to use when using implicit allocation. RTP [6] defines that if not explicitly stated, additional bandwidth shall be used by RTCP equal to 5% of the RTP session bandwidth. The RTCP bandwidth can be explicitly allocated by using what is defined in [9].

7.2.2. Definition

A new media level bandwidth modifier is defined:

b=TIAS:<bandwidth-value>

The Transport Independent Application Specific Maximum (TIAS) bandwidth modifier has an integer bitrate value in bits per second. A fractional bandwidth value SHALL always be rounded up to the next integer. The bandwidth value is the maximum needed by the application without counting IP or transport layers. For RTP based applications, TIAS can be used to derive the RTP "session bandwidth" as defined in section 6.2 of [6]. However in the context of RTP transport the TIAS value is defined as:

> Only the RTP payload with its media SHALL be used in the calculation of the bitrate, i.e. the lower layers (IP/UDP) and RTP headers are excluded. This will allow one to use the same value for both RTP based media transport as non-RTP transport and stored media.

Note 1: The usage of bps is not in accordance with RFC 2327 [1]. This change has no implications on the parser, only the interpreter of the value must be aware. The change is done to allow for better resolution and has also been used for the RR and RS bandwidth modifiers, see [9].

Note 2: RTCP bandwidth is not included in the bandwidth value. In applications using RTCP, the bandwidth used by RTCP is either 5% of the RTP session bandwidth including lower layers or as specified by the RR and RS modifiers [9]. To assign an equal amount of RTCP bandwidth to user of either IPv4 or IPv6, a special rule is defined below in chapter 7.2.4 used to derive RTCP bandwidth.

7.2.3. Usage Rules

To allow for backwards compatibility towards users of SDP that does not implement "TIAS", it is RECOMMENDED to also include the "AS" modifier when using "TIAS". The presence of any value, even with problems is better than none. However an SDP user understanding TIAS when present SHOULD ignore the "AS" value and use TIAS instead.

When using TIAS for an RTP transported stream the "maxprate" attribute SHOULD be included. The "avgprate" MAY also be included. If the "avgprate" is included the "maxprate" MUST be included.

7.2.4. Deriving RTCP bandwidth

To fairly assign RTCP bandwidth to host when both IPv4 and IPv6 hosts are communicating a special method for RTCP bandwidth is here defined. This consist of both a special method for how to calculate the bandwidth for RTCP corresponding to the normal 5% and what to do when calculating the deterministic RTCP interval.

The below defined methods SHALL be used for RTP based transport when done over IP and the bandwidth of the session is signaled with "TIAS". The actual transport layer, UDP etc., does not matter but will taken into consideration in the calculation. It assumes that all communicating host uses the same transport protocols with exception for IP version.

To determine the RTCP bandwidth the transport dependent bitrate is calculated, in accordance with chapter 7.4, using the actual transport layers used with the exception that the IP version used MUST be IPv6. No extension headers SHALL be taken into consideration. So an IPv4 host sending with RTP/UDP will for the RTCP calculation determine a transport dependent bitrate based on RTP/UDP/IPv6 headers. The RTCP bitrate is then equal to 5% of the calculated value.

When determining the RTCP sender interval the following changes shall be made in the calculation. The RTCP packet size used to update the RTCP variable "avg_rtcp_size" (Section 6.3.3 in [6]) SHALL be include the size of an IPv6 header and not IPv4 if used. This shall be done independent of the reason why the "avg_rtcp_size" is updated. When initializing the "avg_rtcp_size" the size of an IPv6 SHALL be used.

By calculating in the RTCP sending rules that the RTCP sender always uses IPv6 there will be fairness between IPv4 and IPv6 hosts.

7.3. Packet Rate parameters

To be able to calculate the bandwidth value including the actually used lower layers two packet rate parameters is also defined.

The maximum packet rate parameter is defined as:

a=maxprate:<packet-rate>

The <packet-rate> is a floating-point value for the streams maximum packet rate. If the number of packets is variable the given value SHALL be the maximum the application, can produce in case of live stream, or for on-demand streams have produced. The value is calculated by taking the largest value a sliding window 1 second long produces when moved over the entire media stream. In cases that this can't be calculated, i.e. for example a live stream, a estimated value of the maximum rate the codec can produce for the given configuration and content SHALL be used.

An average packet rate is defined as:

a=avgprate:<packet-rate>

The <packet-rate> is a floating point value of the average packet rate for the stream. The average value SHOULD be calculated as an average value over the whole stream. If not possible to calculate the value supplied SHALL be an estimate of the most likely packet rate to be used.

The "maxprate" and "avgprate" attribute are media level SDP attributes.

7.4. Converting to Transport Dependent values

When converting the transport independent bitrate value (bw-value) into a transport dependent value including the lower layers the following MUST be done:

1. Determine which lower layers that will be used and calculate the sum of the sizes of the headers in bits (h-size). In cases of variable header sizes the average size SHOULD be used.
2. Retrieve the packet rate to use from the SDP (prate = maxprate or avgprate).
3. Calculate the transport overhead by multiplying the header sizes with the packet rate (t-over = h-size * prate).
4. Round the transport overhead up to nearest integer in bits(t-over = CEIL(t-over)).
5. Add the transport overhead to the transport independent bitrate value (bitrate = bw-value + t-over)

When the above calculation is performed using the "maxprate" the bitrate value should be the absolute maximum the media stream will use over the transport used in the calculations.

When calculated using the "avgprate" the value is derived is the maximum value that will be needed for the average packetization.

7.5. ABNF definitions

This chapter defines in ABNF from RFC 2234 [8] the bandwidth modifier and the packet rate attribute.

The bandwidth modifier:

```
TIAS-bandwidth = "b" "=" "TIAS" ":" bandwidth-value

bandwidth-value = 1*DIGIT
```

The maximum packet rate attribute:

```
max-p-rate-def = "a" "=" "maxprate" ":" packet-rate CRLF
```

The average packet rate attribute:

```
avg-p-rate-def = "a" "=" "avgprate" ":" packet-rate CRLF

packet-rate = 1*DIGIT ["." 1*DIGIT]
```

8. Protocol Interaction

8.1. RTSP

The "TIAS", "maxprate" and "avgprate" can today be used with RTSP. To be able to calculate the transport dependent bandwidth, some of the transport header parameters will be required. There should be no problems for a client to calculate the required bandwidth(s) prior to a RTSP SETUP. The reason is that a client supports a limited number of transport setups when the SDP "m=" line is taken into account. The "m=" line will signal to the client the desired transport profiles.

8.2. SIP

The usage of "TIAS" together with "maxprate" should not be different from the handling of the "AS" modifier currently in use. The needed transport parameters will available in the transport field in the "m=" line. The address class can be determined from the "c=" field and the clients connectivity.

8.3. SAP

In the case of SAP all available information to calculate the transport dependent bitrate should be present in the SDP. The "c=" information gives the used address family for the multicast. The transport layer, e.g. RTP/UDP, for each media is evident in the media line ("m=") and its transport field.

9. Security Consideration

The bandwidth value that is supplied by the parameters defined here can, if not protected, be altered. By altering the bandwidth value one can fool a receiver to reserve either more or less bandwidth than actually needed. Reserving too much may result in unwanted expenses on behalf of user and also blocking of resources that other parties could have used. If to little bandwidth is reserved the receiving users quality MAY be effected. Trusting a to large TIAS value may also result in that the receiver will turn down the session due to insufficient communication and decoding resources.

Due to these security risks it is RECOMMENDED that the SDP is authenticated so no tampering can be performed. It is also RECOMMENDED that any receiver of the SDP performs an analysis of the received bandwidth values so that they are reasonable and is what can be expected for the application. For example an AMR encoded voice stream claiming to use 1000 kbps is not reasonable.

10. IANA Consideration

This document register two new SDP media level attribute "maxprate" and "avgprate", see section 7.3.

A new bandwidth modifier "TIAS" is also registered in accordance with the rules requiring a standard tracks RFC. The modifier is defined in section 7.2.

11. Acknowledgments

The author would like to thank Gonzalo Camarillo and Hesham Soliman for their work reviewing this document.

The author would also like to thank all persons on the MMUSIC working group's mailing list that has commented on this specification.

12. Author's Addresses

Magnus Westerlund Tel: +46 8 4048287
Ericsson Research Email: Magnus.Westerlund@ericsson.com
Ericsson AB
Torshamnsgatan 23
SE-164 80 Stockholm, SWEDEN

13. References

13.1. Normative references

[1] M. Handley, V. Jacobson, "Session Description Protocol (SDP)", IETF RFC 2327, April 1998.
[2] M. Handley et al., "Session Announcement Protocol", IETF RFC 2974, October 2000.
[3] J. Rosenberg, et. al., "SIP: Session Initiation Protocol", IETF RFC 3261, June 2002.
[4] J. Rosenberg, H. Schulzrine, "An Offer/Answer Model with Session Description Protocol (SDP)", IETF RFC 3164, June 2002.
[5] H. Schulzrinne, et. al., "Real Time Streaming Protocol (RTSP)", IETF RFC 2326, April 1998.
[6] H. Schulzrinne, et. al., "RTP: A Transport Protocol for Real-Time Applications", IETF RFC 1889, January 1996.
[7] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.
[8] D. Crocker and P. Overell, "Augmented BNF for syntax specifications: ABNF," RFC 2234, Internet Engineering Task Force, Nov. 1997.

13.2. Informative References

[9] S. Casner, "SDP Bandwidth Modifiers for RTCP Bandwidth", IETF WG draft, draft-ietf-avt-rtcp-bw-05.txt, November 2001, Work in progress
[10] M. Degermark, B. Nordgren, S. Pink, "IP Header Compression", IETF RFC 2507, February 1999.
[11] S. Casner, V. Jacobson, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links", IETF RFC 2508, February 1999.
[12] C. Bormann, et. al., "RObust Header Compression (ROHC): Framework and four profiles", IETF RFC 3095, July 2001.
[13] Tsirtsis, G. and Srisuresh, P., "Network Address Translation - Protocol Translation (NAT-PT)", RFC 2766, Internet Engineering Task Force, February 2000.
[14] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, Internet Engineering Task Force, December 1998.
[15] J. Postel, "internet protocol", RFC 791, Internet Engineering Task Force, September 1981.