ETSI TR 126 924 V18.0.0 (2024-05)

Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Multimedia telephony over IP Multimedia Subsystem (IMS);
Study on improved end-to-end Quality of Service (QoS) handling for Multimedia Telephony Service for IMS (MTSI)
(3GPP TR 26.924 version 18.0.0 Release 18)

Reference
RTR/TSGS-0426924vi00

Keywords
5G,LTE,UMTS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under https://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

Multimedia Telephony for IMS (MTSI) is a standardized service for conversational telephony, TS 22.173 [2]. The media handling and interaction are specified in TS 26.114 [3]. MTSI has been specified such that the user experience of multimedia telephony is equivalent to or better than corresponding circuit-switched telephony services while still having efficient resource usage. Multimedia telephony also exploits the richer capabilities of IMS where media components can be used symmetrically or asymmetrically in different directions.

1 Scope

TS 26.114 define media handling and interaction for the Multimedia Telephony Service for IMS (MTSI) including mechanisms for the negotiation of bandwidth using the SDP bandwidth modifiers. The present study investigates potential improvements for the end-to-end QoS handling with the purpose to improve the network resource allocation for variable bit-rate codecs, rate-adaptive codecs and asymmetric sessions (i.e. different bitrates for different directions). The study will focus on SDP extensions and the interaction with the policy control.

The present document:

- 1 – Identifies high-level use cases
- 2 – Evaluates for these use cases the current limitations and the expected benefits
- 3 – Establishes recommended high-level functional requirements and related recommended technical requirements
- 4 – Discusses potential solutions
- 5 – Studies impact of potential solutions on networks and terminals

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 22.173: "IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1".

[3] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[4] 3GPP TS 23.203: "Policy and charging control architecture".

[5] 3GPP TS 29.212: "Policy and Charging Control (PCC); Reference points".

[6] 3GPP TS 29.213: "Policy and charging control signalling flows and Quality of Service (QoS) parameter mapping".

[7] 3GPP TS 29.214: "Policy and charging control over Rx reference point".

[8] IETF RFC 4566 (2006): "SDP: Session Description Protocol", M. Handley, V. Jacobson and C. Perkins.

[9] IETF RFC 3264 (2002): "An Offer/Answer Model with the Session Description Protocol (SDP)", J. Rosenberg and H. Schulzrinne.

[10] IETF RFC 3890 (2004): "A Transport Independent Bandwidth Modifier for the Session Description Protocol (SDP)"., M. Westerlund.

[11] 3GPP TS 23.401: "Technical Specification Group Services and System Aspects; GPRS enhancements for E-UTRAN access".

[12] 3GPP TS 24.229: "IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3".

[13] IETF RFC 5939: "Session Description Protocol (SDP) Capability Negotiation".

[14] IETF RFC 7006: "Miscellaneous Capabilities Negotiation in the Session Description Protocol (SDP)".

[15] IETF RFC 6871: "Session Description Protocol (SDP) Media Capabilities Negotiation".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [1] and the following apply.
A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply.
An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

AF	Application Function
ARP	Allocation and Retention Priority
AVP	Attribute-Value Pair
EPC	Evolved Packet Core
GW	Gateway
IBCF	Interconnection Border Control Function
LTE	Long Term Evolution
MBR	Maximum Bitrate
MTSI	Multimedia Telephony Service for IMS
P-CSCF	Proxy Call Server Control Function
PCC	Policy and Charging Control
PCEF	Policy and Charging Enforcement Function
PCRF	Policy Charging and Rules Function
PDN-GW	Packet Data Network Gateway (PGW)
RTCP	RTP Control Protocol
SDP	Session Description Protocol
SGW	Serving Gateway
TrGW	Transition Gateway

4 Overview

Clause 5 provides a high-level description of the network elements that are involved in the session setup and resource reservation. The rest of the present document is organized as follows:

Clause 6 describes the use cases analysed in this study. A gap analysis is performed.

Clause 7 describes the recommended requirements that can be derived from the gap analysis.

Clause 8 describes and evaluates potential solutions.

Clause 9 provides the conclusion and recommendations for further standardization efforts.

5 Current QoS reservation mechanisms during session setup

5.1 System description

IMS uses local resource reservation where each IMS network allocates resources only for its own access.

A simplified description of the functions that are used for the bandwidth negotiation and the bearer setup in EPC and LTE is shown in the figure below for the case of two IMS networks A and B. The two networks may include other elements (like IBCF and TrGWs as border elements between networks) that may modify the SDP offer/answer according to interconnection agreement. In case border elements are present the SDP examples in the present document may be modified or seen as generated by such network entities instead of UEs.

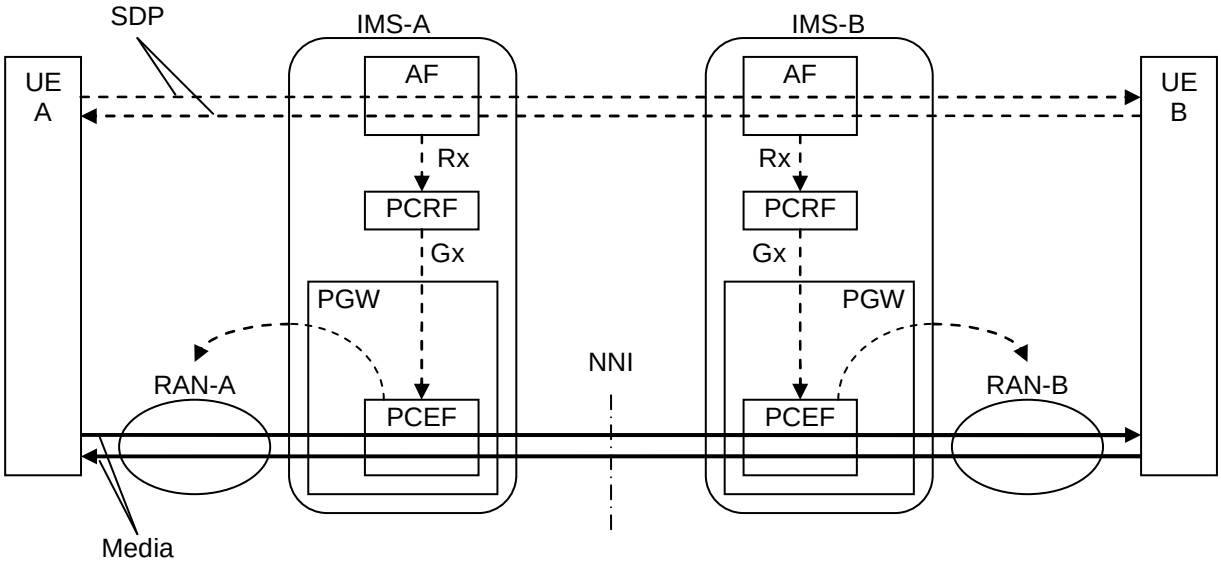

Figure 5.1-1: High-level description of the functions that are involved in the bandwidth negotiation and resource reservation in an IMS network when EPC is used

The resource allocation and bearer setup in EPC/LTE follows the Policy and Charging Control (PCC) procedure in TS 23.203 [4], TS 29.212 [5], TS 29.213 [6] and TS 29.214 [7], and can on a high level be described as follows (a more detailed description is found in TS 29.213 [6] clause 6):

1. The P-CSCF, acting as an Application Function (AF), analyses the SDP offer and the SDP answer and determines the session information and the media information that will be allocated. The AF instructs the Policy Charging and Rules Function (PCRF) to allocate resources for the Service Data Flow. Both the SDP session information and the media information are included in the Rx service information.

   - Before sending the service information to the PCRF, the AF maps the m-lines, c-lines, b-lines and the direction attributes from the SDP to the corresponding Attribute-Value Pairs (AVP) in the service information. The AF provides the media-related attribute lines in the SDP in transparent container AVPs as part of the media information. The AF may also provide an application identifier.

   - The media information includes the media properties, for example maximum UL/DL bitrates for media, UL/DL bitrates for RTCP, codec information, etc.

2. The PCRF converts the requested session information into a set of QoS parameters for the Service Data Flow.

   - The PCRF can also take other information into account when determining the QoS parameters, for example operator policies and subscription information.

- The PCRF may use media-level SDP attribute lines in the service information, the identity of the application and/or operator specific policies to override bandwidth information directly included in the service information. Codec specific algorithms that the PCRF can apply to derive bandwidth information are not standardized, but can be based on the QoS examples in Annex E of TS 26.114 [3].

- The common QoS parameters for all Service Data Flows are: QoS Class Identifier (QCI), and Allocation and Retention Priority (ARP). The additional parameters for Service Data Flows characterized by a GBR QCI are: Maximum Bit Rate (MBR) and Guaranteed Bit Rate (GBR), for uplink and downlink respectively. The additional parameters for Service Data Flows characterized by a non-GBR bearers are: Maximum Bit Rate (MBR), for uplink and downlink, respectively.

3. The PCRF requests the Policy and Charging Enforcement Function (PCEF) to assign the set of QoS parameters to the Service Data Flow through a PCC rule. Additionally the PCRF may request the PCEF to assign for all non-GBR bearers an APN Aggregated MBR (APN AMBR). The PCEF applies the PCC rule(s) to a Service Data Flow by mapping the associated traffic to an existing bearer or by establishing and mapping the traffic to a dedicated bearer (or bearers) between the UE and the PCEF. This includes sending a bearer setup request or a bearer modification request to the RAN to set up or modify Radio Bearer(s) in accordance with the QoS parameters. The PCEF is located in the Packet Data Network Gateway (PDN-GW), a.k.a. PGW.

    - When reserving radio resources, the RAN may also take into account the possible bandwidth savings caused by speech pauses and Robust Header Compression (ROHC). This is not handled by the PCC specifications.

4. The PGW monitors the RTP media traffic and the RTCP traffic, and enforces compliance to APN-MBR rates for non-GBR bearers by dropping packets that exceed the limit when needed. For GBR bearers the enforcement of the downlink MBR is in PGW and uplink MBR in the RAN. The PGW also enforces the downlink and uplink MBR per Service Data Flow.

If a non-GBR bearer was requested and if RAN can set up/modify a Radio Bearer with the requested QoS parameter then the RAN does not reserve dedicated resources for the bearer, i.e. the available bandwidth of the radio bearer can be below the MBR and vary over time without any notification from the RAN to the PGW.

If a GBR bearer was requested and if RAN can set up/modify a Radio Bearer with the requested QoS parameters then it is expected that RAN reserves dedicated resources for the bearer based on the requested GBR. If MBR for a bearer is greater than GBR, the available bandwidth for the bearer, for bitrates greater than GBR and up to the requested MBR, is not guaranteed, i.e. the available bandwidth of the radio bearer can be between the GBR and MBR and vary over time without any notification from the RAN to the PGW.

If the RAN cannot set up/modify a Radio Bearer then the bearer setup/modification request will be rejected. The PCRF may inform the P-CSCF that resources to be associated to the Service Data Flow could not be allocated. The P-CSCF takes action on the SIP session.

A QoS aware terminal will detect if the available bearer resources (as indicated via GBR and MBR values) for a Service Data Flow are lower than the total bandwidth for media and RTCP as indicated in SDP. The terminal will then start a new SDP offer/answer to update the media according to the available resources, see also TS 26.114 [3] clause 6.2.7.

5.2 Simple SDP negotiation and bearer setup

The description below provides more details on the handling session setup negotiation and the corresponding bearer allocation for the simple voice-only case.

It is in this case assumed that RAN will set up a MBR=GBR bearer.

The SDP negotiation between the UEs uses the following SDP offer and SDP answer examples:

Table 5.2-1: Example SDP offer/answer for the session setup for a simple narrow-band voice-only VoLTE call (IPv6)

SDP offer
m=audio 49152 RTP/AVP 97 98 b=AS:38 b=RS:0 b=RR:2000 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:38 b=RS:0 b=RR:2000 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240

This means that both clients want to receive 38 kbps RTP media (including 24 kbps for IPv6/UDP/RTP overhead). They also agree on using RS+RR = 2 kbps for RTCP for the RTP session which means 1 kbps per UE since both UEs will be active senders and the RTCP bandwidth is then divided equally between the UEs. This means that each client is prepared to receive 39 kbps (38 kbps for media and 1 kbps for RTCP).

The PCRF will however allocate RS+RR = 2 kbps for the RTCP bandwidth, both for uplink and downlink. The radio bearers will therefore be set up with (resulting in no additional resource allocation to allow for redundancy transmission):

- MBR-UL = GBR-UL = 40 kbps (38 kbps for media and 2 kbps for RTCP)
- MBR-DL = GBR-DL = 40 kbps

5.3 Other system aspects

The following is not considered in this study:

- RTCP bandwidth allocation, since this would either scale the bandwidths with a fixed percentage, e.g. adding 5 %, or would add a fixed offset, e.g. 2 kbps.
- ROHC usage, since ROHC is only used between the UE and the eNodeB and the usage is not known on the application layer and therefore does not change the bandwidth values expressed with the b=AS bandwidth modifiers.

5.4 Relationship between QoS parameters and rate adaptation

5.4.1 Introduction

This section describes the relationship between the QoS parameters and the rate adaptation, especially how the MTSI client in terminal would use the MBR and GBR parameters. This description assumes that the MTSI client in terminal knows the QoS parameters. It is also assumed that the rate adaptation for uplink and downlink are handled separately and that they are aligned for the uplink QoS parameters and downlink QoS parameters, respectively.

NOTE: MBR and GBR are here described as matching directly the bandwidth parameter b=AS. In reality, the value needs to be translated when converting between the QoS parameters and the SDP parameters because: the SDP parameters, e.g. b=AS, does include the IP, UDP and RTP overhead but not the RTCP bandwidth; the QoS parameters does include both the media bandwidth and the RTCP bandwidth but the IP, UDP and RTP overhead may be significantly reduced if header compression is used.

These sections show only the bitrate adaptation. Other forms of adaptation, e.g. frame aggregation and redundancy, are not included.

5.4.2 MBR=GBR bearer

When an MBR=GBR bearer is used, it is expected that the client will use bitrates below GBR whenever adaptation is triggered.

For ECN-triggered adaptation, TS 26.114 clause 10 describes a mainly "binary" adaptation, which can briefly be described as follows (assuming that nothing else triggers the adaptation):

- When congestion is indicated, it is expected that the adaptation logic adapt the bitrate down to or at least towards the ECN_min_rate. The down-switch can be configured to either go directly to ECN_min_rate or reduce the bitrate in steps. In either case, it is expected that the down-switch reduce the bitrate relatively quickly, as long as the congestion persist. The ECN_min_rate parameter can be configured. Otherwise, TS 26.114 defines a default procedure for determine the value.

- When no congestion is indicated, it is expected that the adaptation logic adapts the bitrate up to or at least towards the highest allowed in the session, which corresponds to MBR. The up-switch can also be configured but it is expected that the bitrate is increased in steps with some waiting period in-between each increment. This means that the up-switch should be relatively slow.

Bitrates between ECN_min_rate and MBR can occur also for ECN-triggered adaptation but should normally only happen during transition periods between congestion and no congestion.

ECN-triggered adaptation should normally not result in bitrates below ECN_min_rate, but this can happen when adaptation is triggered by other means, for example high packet loss rate, long delays or large jitter, if not limited by the codec configuration at session setup.

In this case, the GBR is not used in the adaptation.

5.4.3 MBR>GBR bearer

When an MBR>GBR bearer is used, it is expected that the adaptation tries to maintain a bitrate in-between these values in most cases.

The ECN-triggered adaptation for MBR>GBR bearers is virtually identical to what is described above for MBR=GBR bearers. The only difference is that the ECN_min_rate should be aligned with the GBR.

When other triggers than ECN are used for the adaptation, for example high packet loss rate, long delays or large jitter, the MBR and GBR parameters can be used to describe the normal operating range. Adaptation to bitrates below GBR can be done but should happen relatively rarely. The local access should ensure that adaptation to bitrates below GBR should not be needed. However, the performance in the remote access or the aggregated end-to-end performance may be such that adaptation below the local GBR is needed.

Similar to above, the adaptation to bitrates below GBR assumes that the codec configuration at session setup allows for using bitrates below GBR.

5.4.4 Different QoS settings in different networks

MTSI is a communication service that almost always involves two users. The users are expected to almost always be geographically separated and will therefore use different accesses, often even different operators. Different operators can be expected to use equipment from different vendors and/or different releases. This means that different networks may use different QoS parameter settings. It may even happen that one operator use MBR=GBR bearers while the other operator use MBR>GBR bearers. The devices may even use different access technologies. For example, one user may use LTE while the other uses WiFi. Therefore, if the GBR values are different in the different networks then it can happen that the bitrate is higher than GBR in one network while still being lower than GBR in the other network, or vice versa.

6 Use cases

6.1 General description

This study describes various use cases ranging from relatively simple use cases to more complex use cases. The simple use cases include only one or a few fixed-rate codecs while the more complex use cases include rate-adaptive codecs. The simple use cases are included for the purpose of discussing one issue at a time, even though these use cases may not be the most realistic for real deployments since it is not realistic to assume that all codecs will be allowed in all types of access networks. For example, it is unlikely that the PCM codec will be allowed in LTE or HSPA RAN due to the relatively high bitrate, and there can be an operator policy that removes this codec. However, the main issue in this study is not what exact bitrates are required for certain codecs but rather what happens if several codecs (or configurations) are being offered with different bitrates.

The present document uses the terminology MBR and GBR instead of Max-Requested-Bandwidth-UL/-DL used by the AF, Max_DR_UL/_DL and Gua_DR_UL/_DL used by the PCRF, and Maximum Authorized Bandwidth UL/DL and Guaranteed Authorized Data Rate UL/DL used by the PCEF, respectively. This is done to improve the readability.

6.2 Use case A: Single fixed-rate speech codec

6.2.1 General description

Alice and Bob are setting up a voice-only session. Both UEs support only the A-law PCM (64 kbps) codec. Both UEs use 20 ms frame lengths and encapsulate only 1 frame in each packet. Alice sends the SDP offer and Bob sends the SDP answer as shown below.

Table 6.2.1-1: SDP offer/answer for single codec

SDP offer
m=audio 46000 RTP/AVP 8 b=AS:88 a=rtpmap:8 PCMA/8000/1 a=ptime:20 a=maxptime:20
SDP answer
m=audio 46002 RTP/AVP 8 b=AS:88 a=rtpmap:8 PCMA/8000/1 a=ptime:20 a=maxptime:20

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive 88 kbps (64 kbps for the PCM encoding of the media + 24 kbps for IPv6/UDP/RTP overhead).
- UE-A will send 88 kbps.
- UE-B (Bob) wants to receive 88 kbps.
- UE-B will send 88 kbps.
- Rate adaptation is not possible.
- It is not possible to adapt the packetization because:
  - The bandwidth is limited to 88 kbps. Given that the codec needs 64 kbps this means that the IP/UDP/RTP overhead can be no more than 24 kbps, which corresponds to max 50 packets per second. This means that the packetization is at least 20 ms.
  - However, the 'maxptime' parameter limits the packetization to max 20 ms per packet.

- Hence the only option is to use exactly 20 ms per packet.

The Application Functions uses mapping rules to derive the session information from the SDP offer and the SDP answer:

- In IMS-A:
  - UE-A max send rate is 88 kbps.
  - UE-A min send rate is 88 kbps.
  - UE-A max receive rate is 88 kbps.
  - UE-A min receive rate is 88 kbps.
- In IMS-B:
  - UE-B max send rate is 88 kbps.
  - UE-B min send rate is 88 kbps.
  - UE-B max receive rate is 88 kbps
  - UE-B min receive rate is 88 kbps

The AF sends these parameters together with the remaining media-related information to the PCRF.

The PCRFs then uses the session information, and possibly also the remaining media-related information to determine the following QoS parameters for each local radio network:

Table 6.2.1-2: QoS parameters determined by the PCRF

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	MBR-UL_A	88 kbps	MBR-DL_B	88 kbps
	GBR-UL_A	88 kbps	GBR-DL_B	88 kbps
B->A	MBR-DL_A	88 kbps	MBR-UL_B	88 kbps
	GBR-DL_A	88 kbps	GBR-UL_B	88 kbps

In this case, the two RANs only have the option to allocate MBR=GBR bearers since the codec does not support rate adaptation, since the 'maxptime' parameter prevents encapsulating more frames in the packet and since the bandwidths offered with b=AS prevent higher packet rates than 50 packets per second.

6.2.2 Gap analysis

Comparing the UEs send rates with the QoS parameters gives:

Table 6.2.2-1: Comparison between media bitrate and QoS parameters

Direction	Media rate	QoS parameters A	QoS parameters B	Gap
A->B	88 kbps	MBR-UL_A=88 kbps GBR-UL_A=88 kbps	MBR-DL_B=88 kbps GBR-DL_B=88 kbps	None, bearers optimally allocated
B->A	88 kbps	MBR-DL_A=88 kbps GBR-DL_A=88 kbps	MBR-UL_B=88 kbps GBR-UL_B=88 kbps	None, bearers optimally allocated

In this case, no issues are found.

6.3 Use case B: Several fixed-rate speech codecs

6.3.1 General description

6.3.1.1 Overview

Alice and Bob are setting up a voice-only session. Both UEs support 3 different codecs, A-law PCM (64 kbps), μ-law PCM (64 kbps) and G.729 (8 kbps). Alice sends the SDP offer as shown below which includes all three codecs. UE-B accepts only one codec. If UE-B accepts to use either A-law PCM or μ-law PCM then this gives the same session and bearer setup as shown above for Use case A. This case is therefore not considered any further below. If UE-B accepts to use the G.729 codec then UE-B sends the SDP answer as shown below.

It should be noted that the G.729 codec has a 10 ms frame length.

Table 6.3.1.1-1: First SDP offer and SDP answer for use case B.

SDP offer
m=audio 46000 RTP/AVP 8 0 18 b=AS:88 a=rtpmap:0 PCMU/8000/1 a=rtpmap:8 PCMA/8000/1 a=rtpmap:18 G729/8000/1 a=ptime:20 a=maxptime:80
SDP answer
m=audio 46002 RTP/AVP 18 b=AS:32 a=rtpmap:18 G729/8000/1 a=ptime:20 a=maxptime:80

This means that Alice is limited to send max 32 kbps and Bob is limited to send max 88 kbps.

6.3.1.2 SDP impacts on media handling

For the media handling in the UEs, the SDP offer/answer negotiation means for the A->B direction:

- UE-A can send max 32 kbps because of the limitation to 32 kbps in the SDP answer.
- The media encoding needs 8 kbps which means that the IPv6/UDP/RTP overhead is limited to max 24 kbps, which is achieved for 50 packets per second (two 10 ms frames per packet = 20 ms per packet).
  - This should be regarded as the normal packetization scheme.
- The IPv6/UDP/RTP overhead can be reduced if the packet rate is reduced. This gives a few possible variants, for example (but not limited to):
  - Send 4 non-redundant frames (= 40 ms) in each packet. This gives a packet rate of 25 packets per second and an IPv6/UDP/RTP overhead of 12 kbps. This leaves 12 kbps for redundancy transmission (max 150 % redundancy). However, the 'maxptime' parameter prevents using this amount of redundancy. With 'maxptime' set to 80 it is only possible to include 40 ms of redundant frames in each packet (8 kbps), which gives a redundancy level of 100 %. As such, the total bandwidth becomes 28 kbps.
  - Send 6 non-redundant frames (= 60 ms) in each packet. This gives a packet rate of 16.67 packets per second and an IPv6/UDP/RTP overhead of 8 kbps. This leaves 16 kbps for redundancy transmission (max 200 % redundancy). However, the 'maxptime' parameter prevents using this amount of redundancy. With 'maxptime' set to 80 it is only possible to include 20 ms of redundant frames in each packet (2.67 kbps), which gives a redundancy level of 33.33 %. As such, the total bandwidth becomes 18.67 kbps.
- The lowest possible bandwidth is achieved when sending 8 non-redundant frames (80 ms) in the packets. This reduces the IPv6/UDP/RTP overhead to 6 kbps and the total bandwidth becomes 14 kbps.

  - This would leave 18 kbps that could be used for redundancy. However, the 'maxptime' parameter prevents adding redundant frames to the packets.

- When the UE is not adapting, then the bandwidth will be 32 kbps.
- When the UE is adapting, then the bandwidth can be anything between 14 and 32 kbps. The bandwidth is upwards limited by b=AS in the SDP answer and downwards limited by the 'maxptime' parameter.

For the B->A direction there are more possibilities:

- UE-B can send 1 frame in the packet which gives 100 packets per second. Such packetization would give an IPv6/UDP/RTP overhead of 48 kbps and the total bitrate would be 56 kbps, assuming that no redundancy is used.
- UE-B could use the remaining 32 kbps for redundancy and send 1 original frame and 4 redundant frames in the packets, while still sending 100 packets per second. This means a maximum redundancy of 400 %.
- If UE-B limits the packet rate to 50 packets per second then it could send up to 64 kbps media without exceeding the 88 kbps limit. However, since the 'maxptime' parameter is 80 ms then this allows sending only 8 frames in the packet out of which 2 frames are original frames. This leaves 6 redundant frames per packet which gives a redundancy level of 300 %. This gives a maximum media bandwidth of 32 kbps, an IPv6/UDP/RTP overhead of 24 kbps and a total maximum bandwidth of 56 kbps.
- With a larger 'maxptime' value the UE-B could use up to 700 % redundancy (56 kbps) without exceeding the 88 kbps bandwidth limit while still keeping a packet rate of 50 packets per second, but this would require a 'maxptime' value of 160 ms (2 non-redundant frames and 14 redundant frames, 10 ms each).
- With the 'maxptime' parameter set to 240 ms, as recommended in TS 26.114, then further variants are possible.
- The lowest bitrate that UE-B could achieve is when encapsulating 8 non-redundant frames in the packets. Such encapsulation would reduce the IPv6/UDP/RTP overhead to 6 kbps and the total bitrate would be 14 kbps.
- When the UE is not adapting, then the bandwidth will be 32 kbps.
- When the UE is adapting, then the bandwidth can be anything between 14 and 88 kbps.

6.3.1.3 Bearer allocation based on first SDP offer/answer

All these possibilities mean that the PCRF have many different options to consider when determining the session information parameters, but the two most probable options are likely:

- Option 1: Assume that the session will be asymmetric with max 32 kbps in the A->B direction and max 88 kbps in the B->A direction.
  - According to TS 29.213, clause 6.2, the P-CSCF will select Max-Requested-Bandwidth-UL according to b=AS in the SDP offer and Max-Requested-Bandwidth-DL according to b=AS in the SDP answer, but for other AFs this is only a recommendation.
  - The PCRF may still also override these values.
  - The use of b=AS in the resource reservation may lead to asymmetric bearers.
- Option 2: Assume that the session will be symmetric with max 32 kbps in both directions.
  - According to TS 29.213, clause 6.3, the PCRF prefers to select Max-Requested-Bandwidth-UL/DL based on the codec-specific algorithm for the codec for which the codec specific algorithm exist, regardless of how MBR UL/DL was set by the AF.
  - The use of codec specific algorithm in the resource reservation leads to symmetric bearers.

For the minimum bandwidth, the PCRFs could set this to anything from 14 kbps up to 32 kbps for A->B direction and anything from 14 kbps up to either 32 or 88 kbps for the B->A direction, depending on whether the PCRFs choose option 1 or option 2. The PCRF however selects one single value for the minimum bandwidth. This value needs to be derived from codec specific algorithms or from operator policies since there is no information in the SDP that tells what the UEs plan to use.

It should be noted that there is nothing in the SDPs that show how much redundancy the UEs are allowed to use or how much they plan to use. There is also no information about whether they plan to send 100 packets per second, 50 packets per second or something else. The 'ptime' parameter is only a recommendation and the 'maxprate' parameter defined in RFC 3890 [10] is not used in TS 26.114. However, TS 24.229 has defined the b=TIAS bandwidth modifier and the 'maxprate' parameter as optional SDP parameters. A client specification could introduce limitations on what the clients are allowed to do, for example how much redundancy that is allowed, but the AFs would then have to rely on other mechanisms like a feature tag, or similar, to determine which specification the UE follows (if any).

If both AFs choose option 1 then the PCRFs would set the Authorized IP QoS parameters to (the PCRF has to choose one single value but the tables indicate the range that could be considered):

Table 6.3.1.3-1: Bearer allocation when both networks do resource reservation according to option 1

Direction	QoS parameters A (asymmetric)	Rate	QoS parameters B (asymmetric)	Rate
A->B	MBR-UL_A	32 kbps	MBR-DL_B	32 kbps
	GBR-UL_A	14-32 kbps	GBR-DL_B	14-32 kbps
B->A	MBR-DL_A	88 kbps	MBR-UL_B	88 kbps
	GBR-DL_A	14-88 kbps	GBR-UL_B	14-88 kbps

If both AFs choose option 2 then the PCRFs would set the QoS parameters to:

Table 6.3.1.3-2: Bearer allocation when both networks do resource reservation according to option 2

Direction	QoS parameters A (symmetric)	Rate	QoS parameters B (symmetric)	Rate
A->B	MBR-UL_A	32 kbps	MBR-DL_B	32 kbps
	GBR-UL_A	14-32 kbps	GBR-DL_B	14-32 kbps
B->A	MBR-DL_A	32 kbps	MBR-UL_B	32 kbps
	GBR-DL_A	14-32 kbps	GBR-UL_B	14-32 kbps

Another possibility is that the AF in IMS-A chooses option 1 while the AF in IMS-B chooses option 2. The respective PCRFs would set the QoS parameters to:

Table 6.3.1.3-3: Bearer allocation when network A does resource reservation according to option 1 and network B does resource reservation according to option 2

Direction	QoS parameters A (asymmetric)	Rate	QoS parameters B (symmetric)	Rate
A->B	MBR-UL_A	32 kbps	MBR-DL_B	32 kbps
	GBR-UL_A	14-32 kbps	GBR-DL_B	14-32 kbps
B->A	MBR-DL_A	88 kbps	MBR-UL_B	32 kbps
	GBR-DL_A	14-88 kbps	GBR-UL_B	14-32 kbps

Correspondingly, if the AF in IMS-A chooses option 2 and the AF in IMS-B chooses option 1 then the respective PCRFs would set the QoS parameters to:

Table 6.3.1.3-4: Bearer allocation when network A does resource reservation according to option 2 and network B does resource reservation according to option 1

Direction	QoS parameters A (symmetric)	Rate	QoS parameters B (asymmetric)	Rate
A->B	MBR-UL_A	32 kbps	MBR-DL_B	32 kbps
	GBR-UL_A	14-32 kbps	GBR-DL_B	14-32 kbps
B->A	MBR-DL_A	32 kbps	MBR-UL_B	88 kbps
	GBR-DL_A	14-32 kbps	GBR-UL_B	14-88 kbps

Since local resource reservation is used in IMS this means that the RAN in network A allocates bearers according to the PCRF-A's request, and that RAN in network B allocates bearers according to PCRF-B's request.

When RAN allocates the bearers then it is expected to use:

- If an MBR=GBR bearer is allocated:
  - MBR-UL = MBR-UL for RAN A and B, respectively.
  - MBR-DL = MBR-DL for RAN A and B, respectively.
  - GBR-UL = MBR-UL for RAN A and B, respectively.
  - GBR-DL = MBR-DL for RAN A and B, respectively.
- If an MBR>GBR bearer is allocated:
  - MBR-UL = MBR-UL for RAN A and B, respectively.
  - MBR-DL = MBR-DL for RAN A and B, respectively.
  - GBR-UL = GBR-UL for RAN A and B, respectively.
  - GBR-DL = GBR-DL for RAN A and B, respectively.

6.3.1.4 Bearer allocation based on second SDP offer/answer

One of the identified potential solutions is that UE-A sends a second SDP offer/answer with only the selected codec. This allows UE-A to modify the bandwidth that she is willing to receive, see example below. This solution is already used in TS 26.114 clause 6.2.7 to align the media bandwidth for the receiving direction with the downlink QoS parameters. It is investigated here if the solution can also be used to align the media bandwidth for the sending direction with the uplink QoS parameters.

Table 6.3.1.4-1: Second SDP offer and SDP answer for use case B

Second SDP offer
m=audio 46000 RTP/AVP 18 b=AS:32 a=rtpmap:18 G729/8000/1 a=ptime:20 a=maxptime:80
Second SDP answer
m=audio 46002 RTP/AVP 18 b=AS:32 a=rtpmap:18 G729/8000/1 a=ptime:20 a=maxptime:80

This would mean that both Alice and Bob will be limited to sending max 32 kbps.

It is also expected that the AFs will set the Max-Requested-Bandwidth to 32 kbps, for both UL and DL. Correspondingly, both PCRFs will also set the MBR-UL and MBR-DL parameters to 32 kbps.

For the setting of GBR-UL and GBR-DL parameters the PCRF still need to rely on operator policies or codec-specific algorithms since the second SDP offer/answer still does not include any information about the minimum bitrate that the clients want when receiving or plan to use when sending.

If the MTSI client is QoS aware then it will know how GBR-UL/DL is set for the local access, but there is no mechanisms available in SDP that the MTSI client could use to inform the remote client about the local settings. This may have consequences for the adaptation. One example is when the local QoS parameters are sets to MBR=32 kbps and GBR=14 kbps while the remote RAN has allocated a bearer with MBR=GBR=32 kbps. If congestion occurs in the local downlink then the remote client has no way of knowing how it should adapt to reduce/remove the congestion.

6.3.2 Gap analysis after 1st SDP offer/answer

6.3.2.1 Common

The media rate (UE send rate) shown in the tables below is determined from the information in the SDPs. The case when the client uses also the QoS parameters for the UL bearer to limit the sending rate is commented below the tables, where needed.

The tables below indicate the bitrate range that can be considered for GBR. The PCRF chooses one value within this range.

The Gap analysis consists of separate analyses for Network A and Network B, Gap A and Gap B, respectively. It is judged whether the QoS parameter is optimal, over-allocated or under-allocated with respect to the local access.

An end-to-end Gap analysis is also made, Gap AB, with the following rules:

- If either Gap A or Gap B shows 'under-allocation', then Gap AB becomes 'under-allocation', regardless of the other Gap.
- If both Gaps show 'optimal', then Gap AB becomes 'optimal'.
- If one Gap shows 'over-allocation' and the other shows 'optimal' then Gap AB becomes 'over-allocation'.

6.3.2.2 IMS-A chooses Option 1; IMS-B chooses Option 1

Comparing the UEs send rates with the QoS parameters when both RANs set up bearers according to option 1 gives:

Table 6.3.2.2-1: Gap analysis when both networks do resource reservation according to option 1

Id	Direction	Media rate	QoS parameters A (asymmetric)	Gap A	QoS parameters B (asymmetric)	Gap B	Gap AB
A	A->B 50 packets/sec	14-32 kbps	MBR-UL_A=32 kbps GBR-UL_A=14-32 kbps	Optimal Optimal	MBR-DL_B=32 kbps GBR-DL_B=14-32 kbps	Optimal Optimal	Optimal Optimal
B1	B->A 50 packets/sec No redundancy	14-32 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Over-allocation Over-allocation	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Over-allocation Over-allocation	Over-allocation Over-allocation
B2	B->A 100 packets/sec No redundancy	14-56 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Over-allocation Over-allocation	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Over-allocation Over-allocation	Over-allocation Over-allocation
B3	B->A 50 packets/sec Up to 300 % redundancy	14-56 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Over-allocation Over-allocation	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Over-allocation Over-allocation	Over-allocation Over-allocation
B4	B->A 100 packets/sec Up to 400 % redundancy	14-88 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Optimal Optimal	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Optimal Optimal	Optimal Optimal

Knowledge about the QoS parameters for the bearers does not solve the over-allocation (case B1, B2, B3) because the bearers in the B->A direction are allocated based on what UE-A declares that it is capable of receiving when sending the SDP offer to UE-B, and there is no information about what bitrate UE-B wants to send in the SDP answer that UE-B sends to UE-A.

In general, over-allocation cannot be solved by the knowing the QoS parameters in the client. This can only be solved by providing more information from the clients to the network, which is not included in the 1st offer-answer.

6.3.2.3 IMS-A chooses Option 2; IMS-B chooses Option 2

Comparing the UEs send rates with the QoS parameters when both RANs set up bearers according to option 2 gives:

Table 6.3.2.3-1: Gap analysis when both networks do resource reservation according to option 2

Id	Direction	Media rate	QoS parameters A (symmetric)	Gap A	QoS parameters B (symmetric)	Gap B	Gap AB
A	A->B 50 packets/sec	14-32 kbps	MBR-UL_A=32 kbps GBR-UL_A=14-32 kbps	Optimal Optimal	MBR-DL_B=32 kbps GBR-DL_B=14-32 kbps	Optimal Optimal	Optimal Optimal
B1	B->A 50 packets/sec No redundancy	14-32 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Optimal Optimal	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Optimal Optimal	Optimal Optimal
B2	B->A 100 packets/sec No redundancy	14-56 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Under-allocation Under-allocation	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Under-allocation Under-allocation	Under-allocation Under-allocation
B3	B->A 50 packets/sec Up to 300 % redundancy	14-56 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Under-allocation Under-allocation	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Under-allocation Under-allocation	Under-allocation Under-allocation
B4	B->A 100 packets/sec Up to 400 % redundancy	14-88 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Under-allocation Under-allocation	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Under-allocation Under-allocation	Under-allocation Under-allocation

By knowing the QoS parameters for the bearers a client could handle the issues with under-allocation (case B2, B3, B4) in the local network (network A for UE-A and network B for UE-B, respectively). In this case, handling the under-allocation in one network also handles the under-allocation in the other network because the QoS parameters for RAN-A are the same as in RAN-B. However, in the general case, a client cannot handle under-allocation occurring in the remote network since it does not know how the bearers in the remote network are set up.

It should be noted that, according to TS 24.229 Annex L [12], a UE using LTE access is required to be QoS aware. In addition, according to TS 24.229 Annex B, a UE using UTRAN or GPRS access is also required to be QoS aware. An MTSI client in such a UE could then get the information about the local QoS parameters if cross-layer communication is used. However, there is no requirement in for example TS 26.114 that an API for cross-layer communication does exist.

6.3.2.4 IMS-A chooses Option 1; IMS-B chooses Option 2

Comparing the UEs send rates with the QoS parameters when RAN-A set up bearers according to option 1 and RAN-B set up bearers according to option 2 gives:

Table 6.3.2.4-1: Gap analysis when network A does resource reservation according to option 1 and network B does resource reservation according to option 2

Id	Direction	Media rate	QoS parameters A (asymmetric)	Gap A	QoS parameters B (symmetric)	Gap B	Gap AB
A	A->B 50 packets/sec	14-32 kbps	MBR-UL_A=32 kbps GBR-UL_A=14-32 kbps	Optimal Optimal	MBR-DL_B=32 kbps GBR-DL_B=14-32 kbps	Optimal Optimal	Optimal Optimal
B1	B->A 50 packets/sec No redundancy	14-32 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Over-allocation Over-allocation	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Optimal Optimal	Over-allocation Over-allocation
B2	B->A 100 packets/sec No redundancy	14-56 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Over-allocation Over-allocation	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Under-allocation Under-allocation	Under-allocation Under-allocation
B3	B->A 50 packets/sec Up to 300 % redundancy	14-56 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Over-allocation Over-allocation	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Under-allocation Under-allocation	Under-allocation Under-allocation
B4	B->A 100 packets/sec Up to 400 % redundancy	14-88 kbps	MBR-DL_A=88 kbps GBR-DL_A=14-88 kbps	Optimal Optimal	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Under-allocation Under-allocation	Under-allocation Under-allocation

If the VoIP client in UE-B knows the QoS parameters for the bearer then the client can handle the under-allocation in RAN-B (case B2, B3, B4).

However, UE-B cannot solve the over-allocation in RAN-A (case B2, B3, B4) since UE-B does not have any knowledge about the QoS parameters for UE-A. The only way to solve this is if UE-B provides information about how much it plans to send, so that IMS-A can make a more accurate resource reservation.

6.3.2.5 IMS-A chooses Option 2; IMS-B chooses Option 1

Comparing the UEs send rates with the QoS parameters when RAN-A set up bearers according to option 2 and RAN-B set up bearers according to option 1 gives:

Table 6.3.2.5-1: Gap analysis when network A does resource reservation according to option 2 and network B does resource reservation according to option 1

Id	Direction	Media rate	QoS parameters A (symmetric)	Gap A	QoS parameters B (asymmetric)	Gap B	Gap AB
A	A->B 50 packets/sec	14-32 kbps	MBR-UL_A=32 kbps GBR-UL_A=14-32 kbps	Optimal Optimal	MBR-DL_B=32 kbps GBR-DL_B=14-32 kbps	Optimal Optimal	Optimal Optimal
B1	B->A 50 packets/sec No redundancy	14-32 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Optimal Optimal	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Over-allocation Over-allocation	Over-allocation Over-allocation
B2	B->A 100 packets/sec No redundancy	14-56 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Under-allocation Under-allocation	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Over-allocation Over-allocation	Under-allocation Under-allocation
B3	B->A 50 packets/sec Up to 300 % redundancy	14-56 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Under-allocation Under-allocation	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Over-allocation Over-allocation	Under-allocation Under-allocation
B4	B->A 100 packets/sec Up to 400 % redundancy	14-88 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Under-allocation Under-allocation	MBR-UL_B=88 kbps GBR-UL_B=14-88 kbps	Optimal Optimal	Under-allocation Under-allocation

If the VoIP client in UE-B knows the QoS parameters for the bearer then this does not solve the problem with under-allocation in RAN-A (case B2, B3, B4) since UE-B has no knowledge about the QoS parameters for UE-A. The problem with over-allocation in RAN-B (case B1, B2, B3) can also not be solved by UE-B because the allocation is based on what UE-A has declared in the SDP offer.

To solve the problem with under-allocation in RAN-A a second SDP offer/answer, with only the selected codec, is required. This is further discussed in clause 8.

6.3.2.6 Root-cause analysis when b=AS is used for resource reservation

The use case description shows that when the SDP offer includes multiple codecs with different bitrates and when one of the lower-rate codecs is chosen for the session then this gives an ambiguity regarding how high bitrate UE-B is allowed to send. The root cause for this is the reason that the SDP offer only includes one single bandwidth value, which is set to support the offered codec which requires the highest bandwidth.

There is no information in the SDP offer that limits how much redundancy UE-B may use, except that according to RFC 4566 [8] and RFC 3264 [9] it is not allowed to exceed the b=AS bandwidth and the 'maxptime' value. In addition, there is no information in the SDP answer about whether UE-B plans to use the excessive bandwidth for redundancy, and how much.

The discussion in clauses 6.3.2.2 to 6.3.2.5 shows that both over-allocation and under-allocation may occur in the B->A direction.

Over-allocation may occur if the resource reservation is based on the bandwidth in the SDP offer when asymmetric session is assumed. Under-allocation may occur if the resource reservation is based on the bandwidth in the SDP answer when symmetric session is assumed. The reason for the misalignment is that the SDPs do not include sufficient information to draw the correct conclusions, especially about how high bandwidth the clients plan to send.

Neither over-allocation nor under-allocation can be solved with only one offer-answer but under-allocation occurring in the local network can be handled if the client is QoS aware. Detection of over-allocation and/or under-allocation in the local network can however be used to trigger a second offer-answer negotiation including only the codec selected in the first offer-answer. This is further discussed in clause 6.3.3.

A QoS-aware client can however not detect over-allocation or under-allocation in the remote network so this cannot be used to trigger a second offer-answer.

6.3.2.7 Root-cause analysis when codec-specific information is used for resource reservation

If codec-specific information is used, e.g. based on TS 26.114 Annex E, then it is likely that the session setup will result in the same bearer allocation as if the resources were allocated based on the information in the SDP answer (option 2). If both networks allocate resources this way then this leads to symmetrically allocated bearers. However, since there are no mechanisms in SDP that the clients can use to express their desired or required sending rate then there is no way for the network to know if it was better to allocate resources based on codec-specific information in the SDP offer and the SDP answer than on b=AS in the SDP answer. In TS 29.213 it is defined that codec-specific information takes precedence over the bandwidth offered with b=AS in the SDP.

A QoS-aware client can align its transmission to the QoS parameters in the local network also in this case but it can still not know if this is optimal also for the remote network.

6.3.2.8 Handling of over-allocation and under-allocation

Over-allocation in the network may not be a significant issue for the UEs but it can lead to admitting fewer users, for example the number of simultaneous voice calls the RAN will accept.

Under-allocation in the network is likely a severe problem for the UEs that need to be handled, either by reducing the bitrate or by sending a new SIP message so that the network performs a new resource reservation with the new bitrate.

A UE that is QoS aware, as defined in TS 24.229 Annex B for UTRAN and GPRS or in Annex L for E-UTRAN, will likely detect if under-allocation or over-allocation has happened in the local network and is then expected to initiate a new SDP offer-answer negotiation, see also TS 26.114 clause 6.2.7.

6.3.3 Gap analysis after 2nd SDP offer/answer

When the maximum bandwidth is limited to 32 kbps in both directions then this gives the following Gap analysis.

Table 6.3.3-1: Gap analysis after second SDP offer-answer

Id	Direction	Media rate	QoS parameters A	Gap A	QoS parameters B	Gap B	Gap AB
A	A->B 50 packets/sec	14-32 kbps	MBR-UL_A=32 kbps GBR-UL_A=14-32 kbps	Optimal Optimal	MBR-DL_B=32 kbps GBR-DL_B=14-32 kbps	Optimal Optimal	Optimal Optimal
B	B->A 50 packets/sec	14-32 kbps	MBR-DL_A=32 kbps GBR-DL_A=14-32 kbps	Optimal Optimal	MBR-UL_B=32 kbps GBR-UL_B=14-32 kbps	Optimal Optimal	Optimal Optimal

It should be noted that the optimality indicated here refers only to the case when both networks assign GBR equally. This would, for example, be the case when both PCRFs use codec-specific algorithms that derive the GBR values from the same table. When this is not the case, for example if the PCRFs would use operator policies that are different, then there is no guarantee that Gap AB will be 'optimal' even if both Gap A and Gap B are 'optimal'. This is because the GBR value is selected from a range and different PCRFs may very well choose different values.

It is here also assumed that UE-A and UE-B wants to use the selected codec in the same way. If either UE would need to use for example redundancy then the client can only indicate this for the receiving direction by assigning a higher bandwidth for the b=AS value. This can lead to appropriate bearer allocation in the downlink. For the bearer allocation in the uplink the client will have to rely on the SDP that the remote client sends and that it asks for receiving a higher bandwidth. However, since there are no SDP mechanisms that a client can use to indicate to the remote client that it wants to or needs to send with a larger bandwidth then the likelihood that the remote client will ask for exactly the correct bandwidth that the local client wants to send is probably quite low.

6.4 Use case C: Single multi-rate speech codec (AMR), no extra bandwidth allocated for redundancy

6.4.1 General description

Alice and Bob are setting up a voice-only session. Both UEs support only the AMR codec (4.75-12.2 kbps, all codec modes) but follows TS 26.114 and therefore offers both bandwidth-efficient and octet-aligned. Both UEs propose to encapsulate 1 frame in each packet but allows for up to 12 frames per packet, out of which maximum 4 can be non-redundant frames. Alice sends the SDP offer and Bob sends the SDP answer as shown below. Bob accepts using the bandwidth-efficient payload format version.

Table 6.4.1-1: SDP offer-answer for use case C

SDP offer
m=audio 49152 RTP/AVP 97 98 b=AS:38 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:37 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 38 kbps.
- UE-A will send max 37 kbps.
- UE-B (Bob) wants to receive max 37 kbps.
- UE-B will send max 38 kbps.
- Adaptation is possible in three ways:
  - The AMR codec supports multiple codec modes between 4,75 kbps and 12,2 kbps which mean that the clients can do bitrate adaptation. If a mode-set is defined by the offer-answer then this may restrict the rate throughout to only some codec modes and/or a portion of this range.
  - Frame aggregation is also possible since both clients declare that they can receive up to 240 ms of media in each packet (a=maxptime:240).
  - Redundancy may also be used but both clients declare that they will not send redundant frames that are older than 220 ms (max-red=220).
- Since the UEs cannot inform the network or the other UE what minimum codec mode it wants to send and/or receive or what packetization they plan to use when sending or receiving, there is no guidance in the SDPs for how the IMS networks should select the minimum bitrate that is required for the session.

The table below gives a few examples for how the bitrate changes with the adaptation:

Table 6.4.1-2: Total bitrate as a function of codec mode and packetization

#	Codec mode [kbps]	Packetization [frames/ packet]	Redundancy level [%]	RTP payload size [bytes/ packet]	IPv6/UDP/RTP overhead [bytes/packet]	RTP packet size [bytes]	Packet rate [packets/ second]	Total bitrate [kbps]
1	12.2	1	0	32	60	92	50	36.8 (NOTE 1)
2	4.75	1	0	14	60	74	50	29.6
3	5.9	1	0	16	60	76	50	30.4
4	4.75	2	0	26	60	86	25	17.2
5	4.75	4	0	51	60	111	12.5	11.1
6	4.75	2	100	26	60	86	50	34.4
7	5.9	2	100	32	60	92	50	36.8
8	5.9	2	100 with 1 frame offset	33	60	93	50	37.2 (NOTE 2)
9	4.75	3	200	39	60	99	50	39.6 (NOTE 2)
10	5.9	3	200	47	60	107	50	42.8 (NOTE 2)
11	4.75	4	300	51	60	111	50	44.4 (NOTE 2)
12	5.9	4	300	63	60	123	50	49.2 (NOTE 2)

NOTE 1: This format is expected to be used during normal operating conditions, i.e. when no adaptation is needed to handle congestion, high packet loss rates, large jitter or other degraded operating conditions.
NOTE 2: TS 26.114 allows for using up to 300 % redundancy, see TS 26.114 clause 9.2.1, but it also recommends reducing the bitrate when adding redundancy. The clients will also ensure that the negotiated bandwidth is not exceeded when using redundancy. To use packetization formats 8-12 the clients would need to declare higher bandwidths than what is shown in Table 6.4.1-1.

From this table, it should be clear that using redundancy can lead to using both lower as well as higher bandwidth than the bandwidth used for the normal operation (AMR12.2, no redundancy). A client can only control the amount of bandwidth in the receiving direction since the bandwidth parameter applies only to the receiving direction. The amount of bandwidth that can be used in the sending direction depends on the bandwidth that the remote client has declared in the SDP that it is prepared to receive.

For a UE using LTE or HSPA access types, it is expected that the b=AS is set such that it allows for receiving a bandwidth corresponding to the normal operation (AMR12.2, 1 frame per packet, no redundancy). This bandwidth limitation would then apply independently of what bandwidth the other UE wants to send. This limits the amount of redundancy to 100 % and requires that the clients adapt down to AMR 5.9 kbps, or lower, when using redundancy.

For WiFi, and other access types where high packet loss rates can occur relatively frequently, it can be beneficial to allow for using redundancy, both in the sending and receiving direction, even if this results in using a higher bandwidth. The UE can then set the b=AS value to a larger value, but since the b=AS bandwidth applies only to the receiving direction then this would only enable using higher bandwidths in DL. For UL, there are no mechanisms available to indicate a higher bandwidth for the sending direction. The sending bandwidth is instead limited by the maximum bandwidth that the remote UE has declared that it wants to receive, which the remote UE decides without knowing the local UEs preferences. Hence, for sessions where one UE is using LTE or HSPA and the other is using WiFi then the resource reservation in the LTE/HSPA network may not be sufficient for the UE that uses WiFi.

For the minimum bitrate the situation is even worse since there are no mechanisms that either UE could use to indicate the minimum bitrate they want to receive or want to send.

The PCRFs will have to use codec information to try to guess what the UEs want to do. This can be done in several options, for example:

1) Option 1: A PCRF may assume that the minimum configuration is: AMR4.75, 1 frame per packet and no redundancy. This gives a minimum bitrate of 30 kbps (60 bytes IPv6/UDP/RTP header, 14 bytes for one AMR4.75 speech frame, 50 packets per second gives 29.6 kbps).

2) Option 2: A PCRF may assume that the minimum configuration is: AMR4.75, 4 frames per packet and no redundancy. This gives a minimum bitrate of 12 kbps (60 bytes IPv6/UDP/RTP header, 51 bytes for four AMR4.75 speech frames, 50/4 packets per second gives 11.1 kbps).

3) Option 3: There could also be an operator policy that decides to allocate resources based on some other configuration, for example: AMR5.9, two non-redundant speech frames in each packet and 200 % redundancy. This gives a minimum bitrate of 31 kbps (60 bytes IPv6/UDP/RTP header, 94 bytes for four AMR5.9 speech frames, 50/2 packets per second gives 30.8 kbps).

The lack of mechanisms to negotiate the minimum bitrate that the UEs want to use means that different networks may allocate resources differently. A few example combinations are shown in the tables below.

Table 6.4.1-3: Example bearer allocation, IMS-A allocates resources for GBR ($GBR\text{-}UL_A$, $GBR\text{-}DL_A$) according to option 1, IMS-B allocates resources for GBR ($GBR\text{-}UL_B$, $GBR\text{-}DL_B$) according to option 2

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	$MBR\text{-}UL_A$	37 kbps	$MBR\text{-}DL_B$	37 kbps
	$GBR\text{-}UL_A$	30 kbps	$GBR\text{-}DL_B$	12 kbps
B->A	$MBR\text{-}DL_A$	38 kbps	$MBR\text{-}UL_B$	38 kbps
	$GBR\text{-}DL_A$	30 kbps	$GBR\text{-}UL_B$	12 kbps

Assuming that the UEs are QoS aware this means that:

- UE-A may send with a bitrate between 30 and 37 kbps
- UE-B may send with a bitrate between 12 and 38 kbps.

Another possibility is:

Table 6.4.1-4: Example bearer allocation, IMS-A allocates resources for GBR ($GBR\text{-}UL_A$, $GBR\text{-}DL_A$) according to option 1, IMS-B allocates resources for GBR ($GBR\text{-}UL_B$, $GBR\text{-}DL_B$) according to option 3

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	$MBR\text{-}UL_A$	37 kbps	$MBR\text{-}DL_B$	37 kbps
	$GBR\text{-}UL_A$	30 kbps	$GBR\text{-}DL_B$	31 kbps
B->A	$MBR\text{-}DL_A$	38 kbps	$MBR\text{-}UL_B$	38 kbps
	$GBR\text{-}DL_A$	30 kbps	$GBR\text{-}UL_B$	31 kbps

Assuming that the UEs are QoS aware this means that:

- UE-A may send with a bitrate between 30 and 37 kbps.
- UE-B may send with a bitrate between 31 and 38 kbps.

Yet another possibility is:

Table 6.4.1-5: Example bearer allocation, IMS-A allocates resources for GBR ($GBR\text{-}UL_A$, $GBR\text{-}DL_A$) according to option 2, IMS-B allocates resources for GBR ($GBR\text{-}UL_B$, $GBR\text{-}DL_B$) according to option 3

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	$MBR\text{-}UL_A$	37 kbps	$MBR\text{-}DL_B$	37 kbps
	$GBR\text{-}UL_A$	12 kbps	$GBR\text{-}DL_B$	31 kbps
B->A	$MBR\text{-}DL_A$	38 kbps	$MBR\text{-}UL_B$	38 kbps
	$GBR\text{-}DL_A$	12 kbps	$GBR\text{-}UL_B$	31 kbps

Assuming that the UEs are QoS aware this means that:

- UE-A may send with a bitrate between 12 and 37 kbps.
- UE-B may send with a bitrate between 31 and 38 kbps.

6.4.2 Gap analysis

The difference in maximum rate (37 kbps vs. 38 kbps) gives the same issues as already described in clause 6.3 when multiple fixed-rate codecs with different bitrates are offered. The difference between 37 kbps and 38 kbps might seem insignificant. However, if the answerer would have limited the maximum codec mode to, for example, 5.9 kbps mode then the bandwidth indicated in the SDP answer would likely have been 31 kbps, which would give a much larger difference. This can also be handled with a second SDP offer-answer as discussed in clause 6.3.

The larger issue is instead what minimum bitrates that will be used, and if the UEs use this in the adaptation. Since each UE only knows the QoS parameter for the local access, and it does not know the QoS parameters for the remote access, then it cannot adjust the adaptation to the bitrates allowed in the remote network. For example, if the bearers are set up according to Table 6.4.1-4 then UE-A may choose to not reduce the bitrate below 30 kbps even though the GBR in network B is only 12 kbps. This can be expected to cause significant packet losses or packet delays in network B.

Additionally, since each UE have aligned the transmission to the local QoS parameters, then they have no incentive to send a new SDP offer to try to align the bitrates because the bitrates are already aligned. Even if UE-B would detect that UE-A is sending at a bitrate that is higher than GBR in network B then there are no mechanisms in the SDP to inform UE-A about this fact since the b=AS bandwidth is used to determine MBR and there are no other bandwidth modifiers related to GBR.

It is however expected that most clients will reduce their bitrate even below GBR, if possible, as long as poor operating conditions remain. Hence, poor quality as a consequence of poor operating conditions will likely be a temporary problem and the quality should recover after a while as clients adapt. It may however happen that clients use the GBR as a threshold in their adaptation, for example if the bitrate is above GBR then the client may adapt rapidly down to GBR, but further downwards adaptation may be slower. This is because the GBR rate is supposed to be "guaranteed" by the network. If the RAN would not be able to guarantee this bitrate then it would either reject the session setup or set GBR to a lower value, possibly to 0. In either case, if the session setup is granted then the clients have no real incentive to adapt to lower bitrates, except for exceptional cases.

6.5 Use case D: Single multi-rate speech codec (AMR) with extra bandwidth allocated for redundancy

6.5.1 General description

Alice and Bob are setting up a voice-only session similar to what is shown in use case C. The difference is that Bob is using an access which may have high packet loss rate. UE-B therefore wants to use up to 100 % redundancy with up to 4 frames offset in both uplink and downlink even when using AMR12.2. This means that each packet (bandwidth-efficient) will contain:

- 1 CMR at 4 bits
- 6 ToC entries at 6 bits each:
  - 1 ToC for the redundant frame
  - 4 ToCs for NO_DATA frames
  - 1 ToC for the non-redundant frame
- 2 AMR12.2 speech frames á 244 bits each
- In this case, no padding is needed to fill up to an integer number of octets

The total RTP payload size becomes 528 bits = 66 bytes which, combined with the IPv6/UDP/RTP overhead of 24 kbps, gives a bandwidth of 50.4 kbps. UE-B therefore sets the b=AS bandwidth to 51 kbps.

Table 6.5.1-1: SDP offer-answer for use case D

SDP offer
m=audio 49152 RTP/AVP 97 98 b=AS:38 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:51 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240

It should be noted here that the only difference between this SDP answer and the SDP answer shown in Table 6.4.1-1 is that the bandwidth value is different. The SDP answers are otherwise identical.

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 38 kbps (based on b=AS in SDP offer).
- UE-A will send max 51 kbps (based on b=AS in SDP answer).
- UE-B (Bob) wants to receive max 51 kbps (based on b=AS in SDP answer).
- UE-B will send max 38 kbps (based on b=AS in SDP offer).

The PCRFs can use the b=AS information from the SDP offer and the SDP answer to set up MBR but when assigning GBR they will have to use codec information to try to guess what the UEs want to do since there is no corresponding information in the SDPs. This means that the bearers may be set up in several ways. The three examples shown in clause 6.4.1 may be used also here, except that the maximum bandwidth in the A->B direction is 51 kbps.

This gives the following example bearer allocation, which corresponds to Table 6.4.1-3.

Table 6.5.1-2: Example bearer allocation after first offer/answer, IMS-A allocates resources for GBR (GBR-UL_A, GBR-DL_A) according to option 1, IMS-B allocates resources for GBR (GBR-UL_B, GBR-DL_B) according to option 2

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	MBR-UL_A	51 kbps	MBR-DL_B	51 kbps
	GBR-UL_A	30 kbps	GBR-DL_B	12 kbps
B->A	MBR-DL_A	38 kbps	MBR-UL_B	38 kbps
	GBR-DL_A	30 kbps	GBR-UL_B	12 kbps

This means that redundancy with AMR12.2 is possible in the A->B direction but not in the B->A direction.

Assuming that both UEs are QoS aware, the sending bitrates becomes:

- For UE-A:
  - Max bitrate = 51 kbps, allows for AMR12.2 with 100 % redundancy
  - Min bitrate = 30 kbps
- For UE-B:

  Max bitrate = 38 kbps, requires adapting the bitrate down to AMR5.9 to allow for redundancy

  Min bitrate = 12 kbps

If a second SDP offer/answer negotiation is performed then the SDP that UE-A sends may change the bandwidth in the B->A direction. However, UE-A has no knowledge about what UE-B wants to do, i.e. if UE-B wants to set up a symmetric session or an asymmetric session. Hence, UE-A assigns a bandwidth purely based on what itself want to do, i.e. AMR12.2 with bandwidth-efficient payload format and without redundancy, without taking into account what UE-B wants to do.

A second SDP offer/answer is therefore likely to give the following bearer allocation:

Table 6.5.1-3: Example bearer allocation after second offer/answer

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	MBR-ULA	51 kbps	MBR-DLB	51 kbps
	GBR-ULA	30 kbps	GBR-DLB	12 kbps
B->A	MBR-DLA	37 kbps	MBR-ULB	37 kbps
	GBR-DLA	30 kbps	GBR-ULB	12 kbps

If the PCRFs use codec information to assign bearers then this could result in assigning different values for MBR and GBR. However, since there is no information in the SDP that UE-B sends about its desired sending rate then the PCRFs have no more knowledge than what UE-A has.

6.5.2 Gap analysis

If one assume that both UEs are QoS aware then one get the following gap analysis after the second SDP offer-answer negotiation:

Table 6.5.2-1: Gap analysis after second SDP offer-answer

Id	Direction	Media rate	QoS parameters A	Gap A	QoS parameters B	Gap B	Gap AB
A	A->B	30-51 kbps	MBR-ULA=51 kbps GBR-ULA=30 kbps	Optimal Optimal	MBR-DLB=51 kbps GBR-DLB=12 kbps	Optimal Under-allocation	Optimal Under-allocation
B	B->A	12-37 kbps	MBR-DLA=37 kbps GBR-DLA=30 kbps	Optimal, but undesirable Over-allocation	MBR-ULB=37 kbps GBR-ULB=12 kbps	Optimal, but undesirable Over-allocation	Optimal, but undesirable Over-allocation

As shown in the gap analysis, the lack of information about what UE-B wants to send result in both under-allocation and over-allocation for the GBR values. The MBR allocation in the B->A direction is judged as optimal since UE-B can adjust the bitrate to the QoS parameters. However, the allocated MBR is not the desired maximum bitrate since it does not allow UE-B to use AMR12.2 with redundancy in the sending direction.

Since there are no mechanisms available in SDP to indicate the desired sending rate then additional SDP negotiations will not solve the problem. The same problem occurs if UE-B wants to use a lower encoding rate but more than 100 % redundancy.

6.6 Use case E: Several multi-rate speech codecs (AMR and AMR-WB)

6.6.1 General description

Alice and Bob are setting up a voice-only session. UE-A supports both the AMR codec (4.75-12.2 kbps, all codec modes) and the AMR-WB codec (6.60-23.85 kbps, all codec modes). UE-B supports only the AMR codec. Both UEs follow TS 26.114 and therefore offers both bandwidth-efficient and octet-aligned. Both UEs propose to encapsulate 1 frame in each packet but allows for up to 12 frames per packet, out of which maximum 4 frames can be non-redundant frames. Alice sends the SDP offer and Bob sends the SDP answer as shown below. Bob accepts using the bandwidth-efficient payload format version.

Table 6.6.1-1: SDP offer-answer for use case E

SDP offer
m=audio 49152 RTP/AVP 99 100 97 98 b=AS:49 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:37 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 49 kbps.
- UE-A will send max 37 kbps.
- UE-B (Bob) wants to receive max 37 kbps.
- UE-B will send max 49 kbps.
- Adaptation is possible in same ways as described in clause 6.4.1 for use case C except that UE-B can use the higher bitrate to send more redundancy.
- Since the UEs cannot inform the network or the other UE what minimum codec mode it wants to send and/or receive or what packetization they plan to use when sending or receiving, there is no guidance for how the IMS networks should select the minimum bitrate that is required for the session.

The bandwidths for different configurations shown in Table 6.4.1-2 apply also here.

Similar to discussed in clause 6.5, the PCRFs may allocate MBR in several different ways. A difference here is that UE-B may choose to use all combinations with a bitrate up to 49 kbps (combinations 1-11) while UE-A can only use the combinations up to 37 kbps (combinations 1-7). This can be aligned with a second SDP offer-answer.

The discussion in clause 6.5 on how the PCRFs may allocate GBR holds also here. This cannot be solved with a second SDP offer-answer.

6.6.2 Gap analysis

The differences in MBR give the same gap as has been discussed above in clause 6.3. This can be solved with a second SDP offer-answer negotiation.

The problems caused by the GBRs have also been discussed in clause 6.4. As described in clause 6.4.2, this cannot be solved with a second SDP offer-answer negotiation since there are no SDP parameters available for negotiating this information.

6.7 Use case F: Single video codec, symmetric usage

6.7.1 General description

Alice and Bob are setting up a video telephony session including both speech and video. Both UEs support the minimum set of speech and video codecs defined in TS 26.114, i.e.:

- for speech: AMR (4.75-12.2 kbps); and:
- for video: H.264 Constrained Baseline Profile (CBP) level 1.2.

For speech, both UEs propose to encapsulate 1 frame in each packet but allows for up to 12 frames per packet, out of which maximum 4 can be non-redundant frames.

Alice sends the SDP offer and Bob sends the SDP answer as shown below. Bob accepts using the bandwidth-efficient payload format version. The SDPs do not include attributes for SDPCapNeg, AVPF feedback messages, image attribute and video orientation since these things make no difference for the current analysis.

Since this analysis is targeting issues found for video then the SDPs do not include audio. The SDP offer is aligned with the example SDP offer in TS 26.114, Table A.4.2a2.

Table 6.7.1-1: SDP offer-answer for use case F

SDP offer
m=video 49152 RTP/AVP 99 b=AS:315 b=RS:0 b=RR:2500 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e00c; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==
SDP answer
m=video 49152 RTP/AVP 99 b=AS:315 b=RS:0 b=RR:2500 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e00c; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 315 kbps.
- UE-B (Bob) wants to receive max 315 kbps.
- UE-A will send max 315 kbps.
- UE-B will send max 315 kbps.
- Rate adaptation is possible. However, there is nothing in the SDPs that the functions in the network can use to determine the bitrate range the clients are planning to use when adapting.

The Application Functions use the b=AS values from the SDP offer and the SDP answer sets the service information to:

- In IMS-A:
  - UE-A max send rate is 315 kbps.
  - UE-A min send rate is unknown and is therefore left undefined.
  - UE-A max receive rate is 315 kbps.
  - UE-A min receive rate is also unknown and is therefore left undefined.

- In IMS-B:
  - UE-B max send rate is 315 kbps.
  - UE-B min send rate is unknown and is therefore left undefined.
  - UE-B max receive rate is 315 kbps.
  - UE-B min receive rate is unknown and is therefore left undefined.

The AF sends these parameters together with the remaining media-related information to the PCRF.

The PCRFs then uses the session information, and possibly also the remaining media-related information, to determine the Authorized IP QoS parameters MBR-UL, MBR-DL, GBR-UL and GBR-DL. Since the AF does not provide any minimum bitrates that could be used to set the GBR-UL and GBR-DL parameters then the PCRF uses operator policies instead. If different operators have different policies then GBR will be set differently in the different networks. One example is given below:

Table 6.7.1-2: Example bearer allocation for video

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	$MBR\text{-}UL_A$	315 kbps	$MBR\text{-}DL_B$	315 kbps
	$GBR\text{-}UL_A$	100 kbps	$GBR\text{-}DL_B$	150 kbps
B->A	$MBR\text{-}DL_A$	315 kbps	$MBR\text{-}UL_B$	315 kbps
	$GBR\text{-}DL_A$	100 kbps	$GBR\text{-}UL_B$	150 kbps

The difference between speech and video is that:

- For speech, at least when using AMR or AMR-WB, the PCRF may use the codec mode information available in the SDP to set GBR-UL/DL to the bandwidth needed for the lowest codec mode.
- For video, there is no corresponding information in the SDPs that is related to the desired minimum bandwidth which the PCRF could use to set GBR. The PCRF therefore relies on operator policies or use a codec specific algorithm for this. The codec specific algorithm may use the QoS examples provided in TS 26.114 Annex E. If both PCRFs use the same method to determine GBR-UL/DL then it is likely that they choose the same values.

The lack of information in SDP can give a more or less unpredictable behaviour when the clients need to adapt the bitrate.

6.7.2 Gap analysis

A QoS aware client is expected to align the bitrate range (min-max bitrate) used for the adaptation to the allocated MBR and GBR values. However, a QoS aware client is only aware of the QoS parameter in the local access and has no knowledge about the QoS parameters defined for the bearers in the remote access. A UE in network A may therefore adapt down to 100 kbps, which should give no problems in network B. However, a UE in network B may adapt down to only 150 kbps, which can be expected to give problems in network A if the congestion occurred in that network. After adapting down to GBR further adaptation may or may not happen. The figure below illustrates how the downwards adaptation works if the client follows TS 26.114.

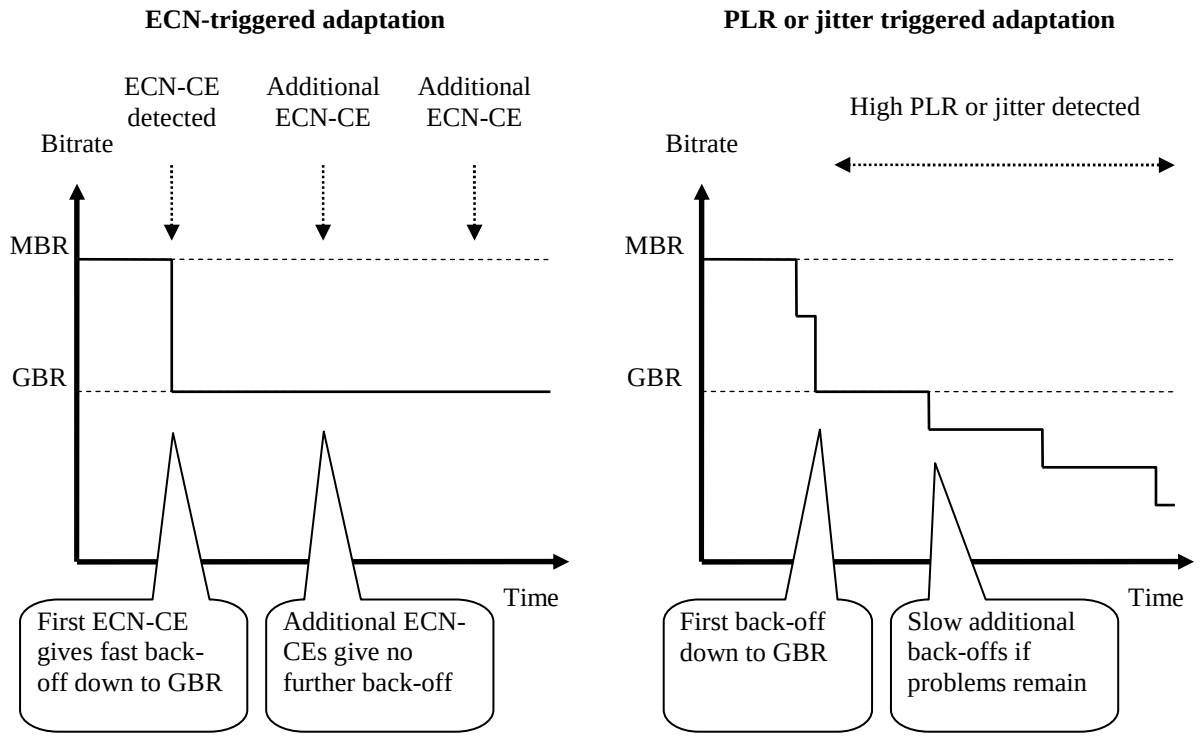

Figure 6.7.2-1: Examples of how MBR and GBR may be used by the MTSI client when adapting

When adaptation is triggered by ECN, then the sending client is expected to adapt down to ECN_min_rate, which is aligned with GBR (if known). This first downwards adaptation can either immediately switch down to ECN_min_rate or may switch down in steps, but in both cases this first downwards adaptation need to be fast. Further ECN-CE markings will not give any further downwards adaptation because ECN_min_rate is supposed to be "guaranteed", even in bad conditions. Further ECN-CE markings will however prevent the client from increasing the rate.

It is here assumed that ECN_min_rate is aligned with the GBR. If they are different then the adaptation is expected to adapt down to the lower of the two bitrates.

When adaptation is triggered by other means than ECN, e.g. packet losses or jitter, then a good reaction is to adapt fast down to GBR (if GBR<MBR). If the bad operating conditions remain then further downwards adaptation is needed but one can expect that this back-off will be slower than the back-off from MBR to GBR. This is again because GBR is supposed to be "guaranteed" and backing off slowly will (hopefully) force other sessions to back-off.

This is why GBR alignment between networks or accesses is important. The receiving client, who detects the bad performance and sends an adaptation request (e.g. TMMBR) back to the sending client, will likely request the sender to adapt down to GBR. However, this will be the GBR of the local access. If the congestion occurs on the sending side and if GBR is lower in the sender's access then the receiver will send a request which does not give enough back-off.

A QoS unaware client will not have any information about how GBR is set, neither for the local network, nor for the remote network. Hence, a QoS unaware client needs to be prepared to adapt down to virtually 0 kbps, which of course will give bad quality.

Hence, the lack of information in SDP can be expected to give a more or less unpredictable behaviour when the clients need to adapt the bitrate. Some UEs may adapt too little, which does not handle congestion properly. Other UEs may adapt too much, which gives under-utilization of the bearers.

In this case, sending a second SDP offer/answer will not help to align the GBRs because the b=AS bandwidths are used to set the MBRs and not the GBRs.

In the discussion above, adaptation triggered by ECN and adaptation triggered by other means is described as separate functions. This is only for the purpose of the discussion. In a real implementation, these functions can be merged into one adaptation function where, for example, the first back-off is triggered by ECN and subsequent back-offs is triggered by high packet loss rate.

6.7.3 Root-cause analysis

There are no bandwidth parameters similar to b=AS that the UE could set can use to indicate which minimum bitrate it wants to send and receive which the remote network (and intermediate networks) could use to align the GBR end-to-end.

6.8 Use case G: Single video codec, asymmetric usage, sending video with a bitrate matching the codec level

6.8.1 General description

Alice and Bob are setting up a video telephony session including both speech and video, as for Use case F above. The difference in this use case is that the clients support asymmetric video as follows:

- UE-A (Alice) supports: encoding with level 1.2 (max 384 kbps), decoding with level 3.1 (max 14 Mbps). UE-A is not capable of receiving video up to the maximum of the level and wants to reduce the bitrate in the receiving direction to 2 Mbps.

- UE-B (Bob) supports: encoding with level 1.3 (max 768 kbps), decoding with level 3.1 (max 14.0 Mbps). However, UE-B wants to limit video in the receiving direction to a lower rate than the maximum for the level, for example 3 Mbps.

UE-A sends the SDP offer to offer level 1.2 (~440 kbps including IPv6/UDP/RTP overhead) but also includes the 'max-recv-level' to indicate that it can receive up to level 3.1 (14 Mbps) but the bitrate in the receiving direction is limited to 2.0 Mbps with the b=AS parameter (~2.1 Mbps including overhead).

UE-B sends the SDP answer to offer level 1.3 (~810 kbps including IPv6/UDP/RTP overhead) but also includes the 'max-recv-level' to indicate that it can receive up to level 3.1. However, the b=AS value indicates that it wants to receive maximum video at 3.0 Mbps (~3.2 Mbps including overhead).

Since this analysis is targeting issues found for video then the SDPs do not include audio.

Table 6.8.1-1: SDP offer-answer for use case G

SDP offer
m=video 49152 RTP/AVP 99 b=AS:2100 b=RS:0 b=RR:5000 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e00c; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==; \ max-recv-level=e01f
SDP answer
m=video 49152 RTP/AVP 99 b=AS:3200 b=RS:0 b=RR:5000 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e00d; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==; \ max-recv-level=e01f

NOTE: The SDP offer is aligned with the SDP offer in TS 26.114 Table A.4.13. The SDP answer is however deliberately different to facilitate the discussion.

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 2100 kbps.

- UE-B (Bob) wants to receive max 3200 kbps.

- UE-A can send up to max ~440 kbps, if allowed by the UE-B.

- UE-B can send up to max ~810 kbps, if allowed by the UE-A.
- Rate adaptation is possible. However, similar to Use case F, there is nothing in the SDPs that the functions in the network can use to determine the bitrate range the clients are planning to use when adapting.

The Application Functions use the b=AS values from the SDP offer and the SDP answer sets the service information to:

- In IMS-A:
  - UE-A max send rate is 3200 kbps (based on b=AS in the SDP answer).
  - UE-A min send rate is unknown and is therefore left undefined.
  - UE-A max receive rate is 2100 kbps (based on b=AS in the SDP offer).
  - UE-A min receive rate is also unknown and is therefore left undefined.
- In IMS-B:
  - UE-B max send rate is 2100 kbps (based on b=AS in the SDP offer).
  - UE-B min send rate is unknown and is therefore left undefined.
  - UE-B max receive rate is 3200 kbps (based on b=AS in the SDP answer).
  - UE-B min receive rate is unknown and is therefore left undefined.

It should be noted that allocating bearers in this way would give quite large over-allocation since UE-A will send max ~440 kbps and that UE-B will send max ~810 kbps.

The PCRF would have to use the profile-level-id to set the Authorized IP QoS parameters to more reasonable values. If the PCRF does not check the codec-specific information then bearers will be seriously over-allocated.

As for Use case F, the PCRF would also have to set GBR-UL/DL based on operator policies or codec-specific algorithms since there is no information in the SDPs related to this.

6.8.2 Gap analysis

For optimal bearer allocation, the PCRF uses codec-specific information to analyse both the 'profile-level-id' parameter and the 'max-recv-level' parameter. If the PCRF does not do this then over-allocation is likely to happen. There are no generic bandwidth parameters that could be used for this.

The issue with different GBRs in different network, which causes problems for the adaptation, is the same as for Use case F.

Another issue is that UE-B knows what it wants to send but there are no mechanisms in SDP to communicate this to UE-B.

6.8.3 Root-cause analysis

There are no bandwidth parameters similar to b=AS that the UE could set can use to indicate the desired send rate.

This could possibly be solved with a second SDP offer/answer, but only if the clients analyse the received 'profile-level-id' parameter and assigns b=AS accordingly. However, there is nothing explicitly wrong with declaring that one can receive a higher bitrate that the other client is going to send, so there is no real motivation why a UE would send a second SDP offer/answer.

6.9 Use case H: Single video codec, asymmetric usage, sending video with a bitrate lower than the supported codec level

6.9.1 General description

This use case is very similar to Use case G, except that UE-B wants to send video using a bitrate that is lower than the supported H.264 profile and level and when there is no corresponding level defined for the bitrate that UE-B wants to use.

Alice and Bob are setting up a video telephony session including both speech and video, as for Use case F above. The difference in this use case is that the clients support asymmetric video as follows:

- UE-A (Alice) supports: encoding with level 1.2 (max 384 kbps), decoding with level 3.1 (14 Mbps) (max 14 Mbps). UE-B is not capable of receiving video up to the maximum of the level and wants to reduce the bitrate in the receiving direction to 2 Mbps.

- UE-B (Bob) supports: encoding with level 2.0 (max 2.0 Mbps), decoding with level 3.1 (max 14.0 Mbps). However, UE-B wants to limit video to 1 Mbps in the sending direction and to 5 Mbps in the receiving direction.

NOTE: There is no H.264 level that corresponds to 1 Mbps and the next higher level is 2.0 (2 Mbps). There is also no H.264 level (for Constrained Baseline Profile) that corresponds to 5 Mbps. To be able to receive at 5 Mbps the UE needs to support at least level 3.

UE-A sends the SDP offer to offer level 1.2 (~440 kbps including IPv6/UDP/RTP overhead) but also includes the 'max-recv-level' to indicate that it can receive up to level 3.1 (14 Mbps) but the bitrate in the receiving direction is limited to 2.0 Mbps with the b=AS parameter (~2.1 Mbps including overhead).

UE-B sends the SDP answer to offer level 2 (~2.1 Mbps including IPv6/UDP/RTP overhead) but also includes the 'max-recv-level' to indicate that it can receive up to level 3.1. However, the b=AS value indicates that it wants to limit the bandwidth in the receiving direction to max 5.2 Mbps (including overhead).

The main difference, compared to Use case G, is that UE-B wants to send video with a bitrate that is lower than the maximum for the level and that there is no level defined for the bitrate that UE-B wants to send.

Since this analysis is targeting issues found for video then the SDPs do not include audio.

Table 6.9.1-1: SDP offer-answer for use case H

SDP offer
m=video 49152 RTP/AVP 99 b=AS:2100 b=RS:0 b=RR:5000 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e00c; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==; \ max-recv-level=e01f
SDP answer
m=video 49152 RTP/AVP 99 b=AS:5200 b=RS:0 b=RR:5000 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e014; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==; \ max-recv-level=e01f

NOTE: The SDP offer is aligned with the SDP offer in TS 26.114 Table A.4.13. The SDP answer is however deliberately different to facilitate the discussion.

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 2.1 Mbps.

- UE-B (Bob) wants to receive max 5.2 Mbps.
- UE-A can send up to max ~440 kbps, if allowed by the UE-B.
- UE-B can send up to max ~2.1 Mbps. However, in this case, UE-B plans to send max ~1.1 Mbps.
- Rate adaptation is possible, as for Use case F.

The Application Functions use the b=AS values from the SDP offer and the SDP answer sets the service information to:

- In IMS-A:
  - UE-A max send rate is 5200 kbps (based on b=AS in the SDP answer).
  - UE-A min send rate is unknown and is therefore left undefined.
  - UE-A max receive rate is 2100 kbps (based on b=AS in the SDP offer).
  - UE-A min receive rate is also unknown and is therefore left undefined.
- In IMS-B:
  - UE-B max send rate is 2100 kbps (based on b=AS in the SDP offer).
  - UE-B min send rate is unknown and is therefore left undefined.
  - UE-B max receive rate is 5200 kbps (based on b=AS in the SDP answer).
  - UE-B min receive rate is unknown and is therefore left undefined.

The PCRFs then uses the session information, and possibly also the remaining media-related information, to determine the Authorized IP QoS parameters. Since the AF does not provide any minimum bitrates that could be used to set the GBR-UL and GBR-DL parameters then the PCRF uses operator policies or codec-specific algorithms instead. In this case, it is assumed that the operator policy defines that GBR-UL/DL is set to ~50 % of the MBR-UL/DL, respectively.

Table 6.9.1-2: Bearer allocation

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	$MBR\text{-}UL_A$	5200 kbps	$MBR\text{-}DL_B$	5200 kbps
	$GBR\text{-}UL_A$	2100 kbps	$GBR\text{-}DL_B$	2100 kbps
B->A	$MBR\text{-}DL_A$	2100 kbps	$MBR\text{-}UL_B$	2100 kbps
	$GBR\text{-}DL_A$	1100 kbps	$GBR\text{-}UL_B$	1100 kbps

Since UE-B is going to send max 1.1 Mbps and since $MBR\text{-}UL_B = MBR\text{-}DL_A$ = 2.1 Mbps this gives an over-allocation in the B->A direction in both networks.

Similarly, if a UE supports the same H.264 profile and level in both sending and receiving direction and if the UE will use a lower send rate than what is indicated with the level (and the b=AS value) and if the UE does not use the 'max-recv-level' parameter to indicate a higher level for the receiving direction, then the same over-allocation will also occur.

6.9.2 Gap analysis

The PCRFs has no way of knowing that a UE-B is planning to send video with a bitrate that is lower than the maximum bitrate for the supported H.264 profile and level.

UE-A also has no way of knowing what bandwidth UE-B will send, so initiating a new SDP offer-answer negotiation will not resolve the issue, because UE-A has no information that it could use to choose a more suitable b=AS value.

UE-B could also initiate a second SDP offer-answer, but since there are no SDP parameters to indicate the desired sending rate then this will not help.

6.9.3 Root-cause analysis

There is no bandwidth parameter similar to b=AS that the UE can use to indicate the desired sending rate.

Operator policies or codec-specific algorithms can be used but will only work if they are aligned with what the UE wants to do.

6.10 Use case I: Multiple video codecs

6.10.1 General description

Alice and Bob are setting up a video telephony session including both speech and video. Both UEs support AMR speech coding. For video, both UEs support video coding as follows:

- 5 inch display with 848x480 resolution.
- H.264 (AVC) Constrained Baseline Profile (CBP) level 3.1. For the given display size and resolution the H.264 (AVC) codec would need 690 kbps (including IPv6/UDP/RTP overhead).
- H.265 (HEVC) Main Profile, Main tier level 3.1. For the given display size and resolution the H.265 (HEVC) codec would need 540 kbps (including IPv6/UDP/RTP overhead).
- Both UEs want to use the more efficient H.265 video codec to reduce the bitrate to 540 kbps whenever H.265 can be used, i.e. when both UEs support the H.265 codec.

UE-A sends the SDP offer to offer with b=AS set to 690 kbps (including IPv6/UDP/RTP overhead) since the client needs to choose the higher of the bitrates needed for each respective codec. UE-B accepts H.265 and sets b=AS to 540 kbps. The SDP offer and the SDP answer are the same as in TS 26.114 Table A.4.16 and Table A.4.18, respectively.

Table 6.10.1-1: SDP offer-answer for use case I

SDP offer
m=video 49154 RTP/AVP 98 97 100 99 a=tcap:1 RTP/AVPF a=pcfg:1 t=1 b=AS:690 b=RS:0 b=RR:5000 a=rtpmap:100 H264/90000 a=fmtp:100 packetization-mode=0; profile-level-id=42e01f; \ sprop-parameter-sets=Z0KAHpWgNQ9oB/U=,aM46gA== a=imageattr:100 send [x=848,y=480] recv [x=848,y=480] a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e01f; \ sprop-parameter-sets=Z0KADZWgUH6Af1A=,aM46gA== a=imageattr:99 send [x=320,y=240] recv [x=320,y=240] a=rtpmap:98 H265/90000 a=fmtp:98 profile-id=1; level-id=5d; \ sprop-vps=QAEMAf//AWAAAAMAgAAAAwAAAwB4LAUg; \ sprop-sps=QgEBAWAAAAMAgAAAAwAAAwB4oAaiAeFlLktIvQB3CAQQ; \ sprop-pps=RAHAcYDZIA== a=imageattr:98 send [x=848,y=480] recv [x=848,y=480] a=rtpmap:97 H265/90000 a=fmtp:97 profile-id=1; level-id=5d; \ sprop-vps=QAEMAf//AWAAAAMAgAAAAwAAAwB4LAUg; \ sprop-sps=QgEBAWAAAAMAgAAAAwAAAwB4oAoIDxZS5LSL0AdwgEE=; \ sprop-pps=RAHAcYDZIA== a=imageattr:97 send [x=320,y=240] recv [x=320,y=240] a=rtcp-fb:* trr-int 5000 a=rtcp-fb:* nack a=rtcp-fb:* nack pli a=rtcp-fb:* ccm fir a=rtcp-fb:* ccm tmmbr a=extmap:4 urn:3gpp:video-orientation

SDP answer
m=video 49156 RTP/AVPF 98 a=acfg:1 t=1 b=AS:540 b=RS:0 b=RR:5000 a=rtpmap:98 H265/90000 a=fmtp:98 profile-id=1; level-id=5d; \ sprop-vps=QAEMAf//AWAAAAMAgAAAAwAAAwB4LAUg; \ sprop-sps=QgEBAWAAAAMAgAAAAwAAAwB4oAaiAeFlLktIvQB3CAQQ; \ sprop-pps=RAHAcYDZIA== a=imageattr:98 send [x=848,y=480] recv [x=848,y=480] a=rtcp-fb:* trr-int 5000 a=rtcp-fb:* nack a=rtcp-fb:* nack pli a=rtcp-fb:* ccm fir a=rtcp-fb:* ccm tmmbr a=extmap:4 urn:3gpp:video-orientation

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 690 kbps.
- UE-B (Bob) wants to receive max 540 kbps.
- UE-A can send up to max 540 kbps, since UE-B has limited the bitrate in its receiving direction to this.
- UE-B can send up to max 690 kbps, since UE-A has not introduced any further limitations.
- Rate adaptation is possible, as for Use case F.

The Application Functions use the b=AS values from the SDP offer and the SDP answer sets the service information to:

- In IMS-A:
  - UE-A max send rate is 540 kbps (based on b=AS in the SDP answer).
  - UE-A min send rate is unknown and is therefore left undefined.
  - UE-A max receive rate is 690 kbps (based on b=AS in the SDP offer).
  - UE-A min receive rate is also unknown and is therefore left undefined.
- In IMS-B:
  - UE-B max send rate is 690 kbps (based on b=AS in the SDP offer).
  - UE-B min send rate is unknown and is therefore left undefined.
  - UE-B max receive rate is 540 kbps (based on b=AS in the SDP answer).
  - UE-B min receive rate is unknown and is therefore left undefined.

The PCRFs then uses the session information, and possibly also the remaining media-related information, to determine the Authorized IP QoS parameters. Since the AF does not provide any minimum bitrates that could be used to set the GBR-UL and GBR-DL parameters then the PCRF uses operator policies instead. The MBRs and the GBRs for the A->B direction and the GBRs for the B->A direction are aligned with the QoS example in TS 26.114 clause E.24. The MBRs for the B->A direction are aligned with the b=AS in the SDP offer.

Table 6.10.1-2: Bearer allocation

Direction	QoS parameters A	Rate	QoS parameters B	Rate
A->B	$MBR\text{-}UL_A$	540 kbps	$MBR\text{-}DL_B$	540 kbps
	$GBR\text{-}UL_A$	64 kbps	$GBR\text{-}DL_B$	64 kbps
B->A	$MBR\text{-}DL_A$	690 kbps	$MBR\text{-}UL_B$	690 kbps
	$GBR\text{-}DL_A$	64 kbps	$GBR\text{-}UL_B$	64 kbps

This means that the improved coding gain is used for:

- lower bitrate in the A->B direction; and for:
- better quality in the B->A direction.

6.10.2 Gap analysis

The PCRFs has no way of knowing that a UE-A is planning to send video with a bitrate that is lower for H.265 than for H.264.

6.10.3 Root-cause analysis

There is no bandwidth parameter similar to b=AS that the UE can use to indicate different bandwidths for different codecs. This could possibly be solved with SDP Capability Negotiation RFC 5939 [14] and SDP Miscellaneous Capability Negotiation RFC 7006 [15].

A second SDP offer/answer could also be used where only the H.265 codec is included. UE-A could use this to indicate the appropriate maximum receiving rate for the H.265 codec.

6.11 Use case J: Single video codec, symmetric usage, bitrate variations

6.11.1 General description

This use case is identical to Use case F but the discussion here focuses on bitrate variations that may occur in the session, even under normal operating conditions with good channel condition and low network load so that there is no need for end-to-end bitrate adaptation. This use case therefore illustrates the bitrate variations that may be generated by a video codec if no restrictions are applied to the encoding process.

Alice and Bob are setting up a video telephony session including both speech and video. Both UEs support the minimum set of speech and video codecs defined in TS 26.114, i.e.:

- for speech: AMR (4.75-12.2 kbps); and:
- for video: H.264 Constrained Baseline Profile (CBP) level 1.2.

Alice sends the SDP offer and Bob sends the SDP answer as shown below. Since this analysis is targeting issues for bitrate variations for video then the SDPs do not include audio. The SDP examples shown below are also simplified versions without SDPCapNeg, AVPF feedback messages, image attribute and video orientation since these things make no difference for the current analysis.

Table 6.11.1-1: SDP offer-answer for use case J

SDP offer
m=video 49152 RTP/AVP 99 b=AS:315 b=RS:0 b=RR:2500 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e00c; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==
SDP answer
m=video 49152 RTP/AVP 99 b=AS:315 b=RS:0 b=RR:2500 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0; profile-level-id=42e00c; \ sprop-parameter-sets=J0LgDJWgUH6Af1A=,KM46gA==

This discussion focuses on video and speech is not considered any more in this use case.

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A (Alice) wants to receive max 315 kbps (on average).
- UE-B (Bob) wants to receive max 315 kbps (on average).
- UE-A will send max 315 kbps (on average).
- UE-B will send max 315 kbps (on average).

The Application Functions use the b=AS values from the SDP offer and the SDP answer sets the service information to:

- In IMS-A:
  - UE-A max send rate is 315 kbps (on average).
  - UE-A min send rate is unknown and is therefore left undefined.
  - UE-A max receive rate is 315 kbps (on average).
  - UE-A min receive rate is also unknown and is therefore left undefined.
- In IMS-B:
  - UE-B max send rate is 315 kbps (on average).
  - UE-B min send rate is unknown and is therefore left undefined.
  - UE-B max receive rate is 315 kbps (on average).
  - UE-B min receive rate is unknown and is therefore left undefined.

The AF sends these parameters together with the remaining media-related information to the PCRF.

In this case, it is assumed that the PCRFs suggest to set up a bearer with MBR=GBR. This gives the following bearer allocation.

Table 6.11.1-2: Bearer allocation for video

Direction	Parameter	Rate	Parameter	Rate
A->B	$Max\_DR\_UL_A$	315 kbps	$Max\_DR\_DL_B$	315 kbps
	$Gua\_DR\_UL_A$	315 kbps	$Gua\_DR\_DL_B$	315 kbps
B->A	$Max\_DR\_DL_A$	315 kbps	$Max\_DR\_UL_B$	315 kbps
	$Gua\_DR\_DL_A$	315 kbps	$Gua\_DR\_UL_B$	315 kbps

In this discussion it is assumed that these parameters are also used in the PGW to monitor that the UEs do not exceed the negotiated bandwidths.

Two trace files of video frame sizes are used to facilitate the discussion on bitrate variations. The files are described in the Table 6.11.1-3 below and are shown in Figures 6.11.1-1 and 6.11.1-2.

Table 6.11.1-3: File information

	File 1	File 2
Codec	H.264	H.264
Resolution	QCIF	QCIF
Frame rate	30 fps	30 fps
Bitrate (incl. IP, UDP and RTP overhead)	315 kbps	315 kbps
Number of frames	2692	2204
Average frame size	1312.5 bytes	1312.5 bytes
Min frame size	595 bytes	1006 bytes
Max frame size	2147 bytes	1771 bytes

These files do not include any I frames.

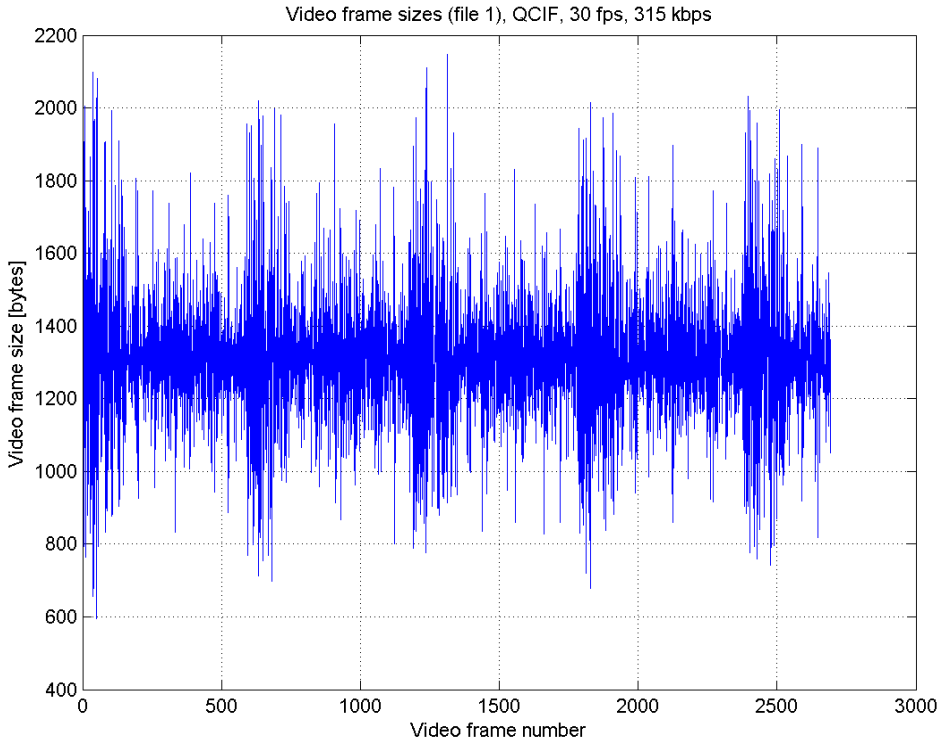

Figure 6.11.1-1: Video frame sizes for file 1

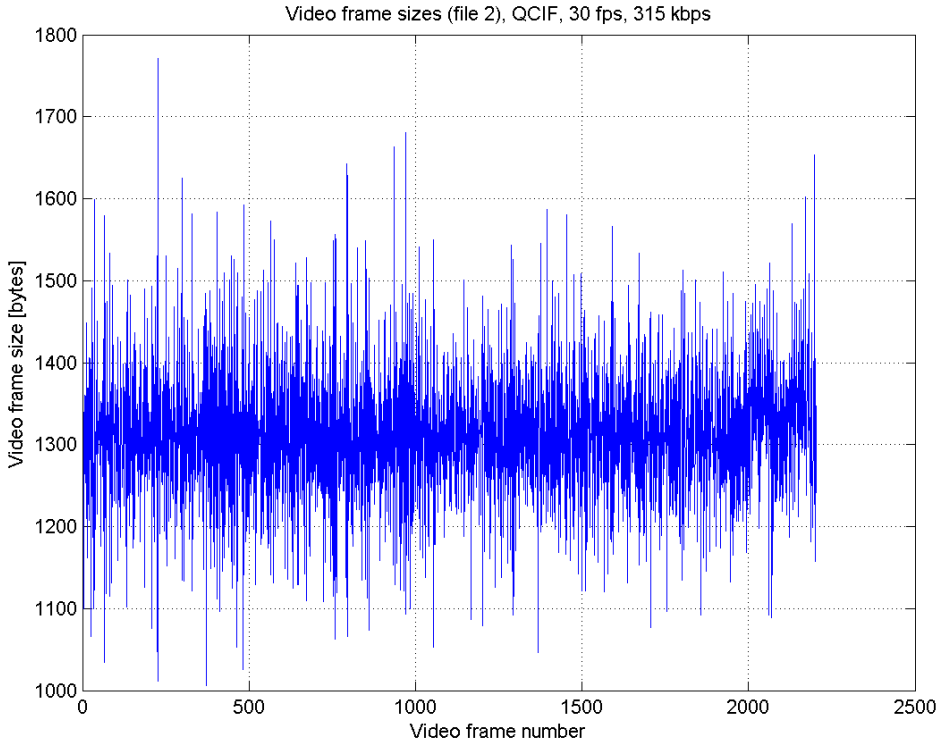

Figure 6.11.1-2: Video frame sizes for file 2

When measuring the used bitrate, e.g. in a policing function, then one need to average the instantaneous bitrates over some time to create a short-term average. In this analysis, an averaging window has been used and different lengths of the averaging window have been tested. The figure below shows a few examples of how the variations in the short-term bitrate average are reduced as the length of the averaging window increases.

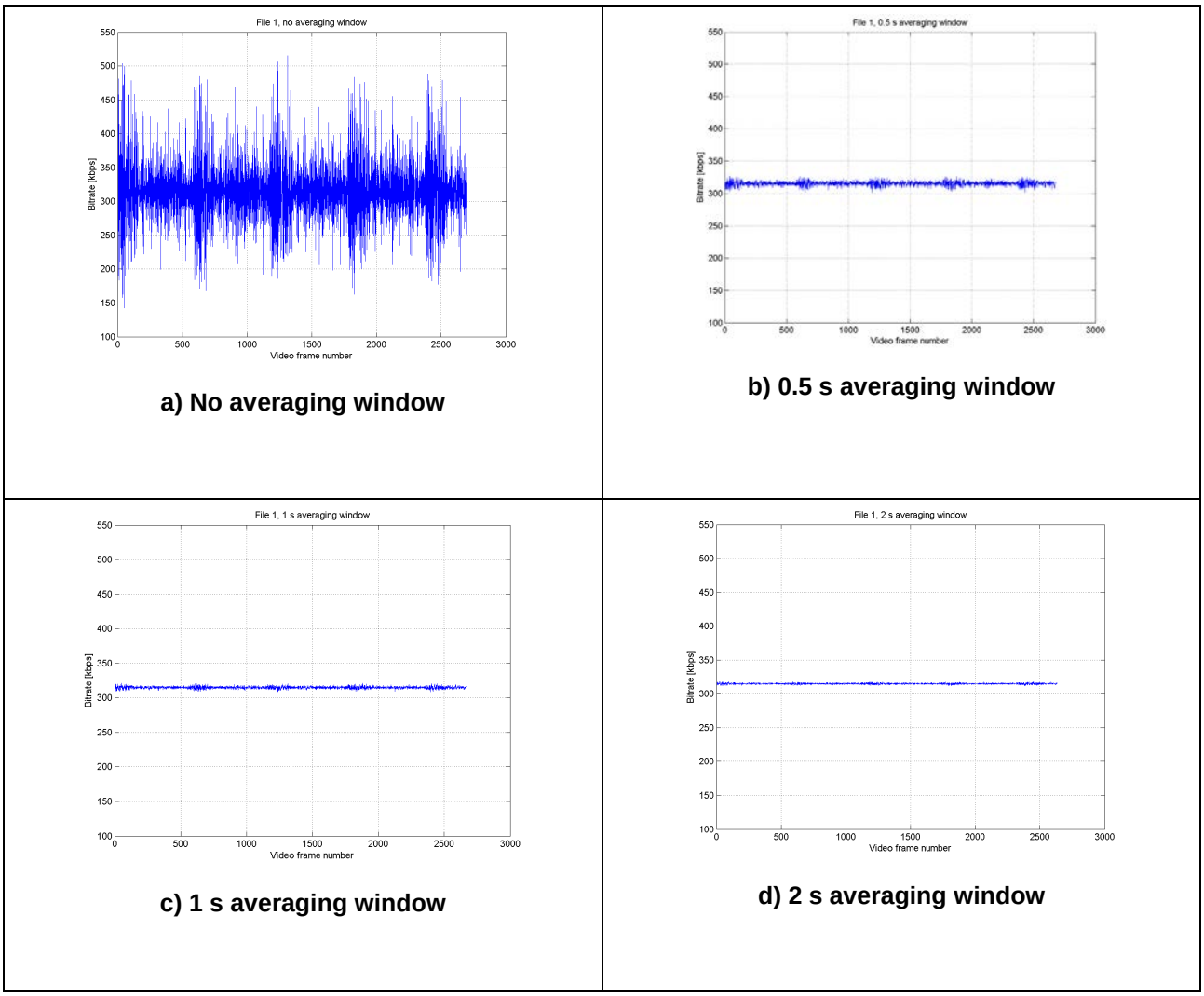

Figure 6.11.1-3: Bitrates for file 1 with different averaging windows

As can be seen in the figure above the variations in the short-term average are significantly reduced when applying an averaging window, even if the window is as short as 0.5 seconds. This is however excluding I frames. Statistics when applying different averaging window lengths are shown in Table 6.11.1-4.

I frames may be generated by the encoder for several reasons, for example generated at a regular interval to stop error propagation or generated when the receiver requests a Full Intra Refresh. I frames are usually much larger than the average frame size, often as large as 5 to 10 times larger. The effects of I frames on the short-term bitrate average have been analysed by manually inserting I frames every 15 second in the video trace files. Both I frames of 5 times and 10 times the average video frame size have been used. The video trace files were then re-scaled to maintain the 315 kbps average bitrate (measured over the whole trace file). The averaging windows were then applied in the same was as above. Table 6.11.1-4 shows the statistics for the two files when I frames have been added.

Table 6.11.1-4: Bitrate variations after applying averaging window

File	Averaging window length [s]	Average bitrate [bps]	Maximum bitrate [bps]	Minimum bitrate [bps]	Max/Avg bitrate	Number of frames	Number of times short-term average is >25 % over long-term average	Number of times short-term average is >10 % over long-term average
File 1	-	314999.6	515280	142800	1.64	2692	160	584
File 1	0.17	315003.3	352272	280944	1.12	2687	0	7
File 1	0.33	314987.7	333648	294816	1.06	2682	0	0
File 1	0.50	314980.8	327392	302080	1.04	2677	0	0
File 1	0.67	314980.9	323400	306300	1.03	2672	0	0
File 1	1.00	314986.6	321208	308784	1.02	2662	0	0
File 1	2.00	314990.1	317848	311728	1.01	2632	0	0
File 1, 5x I	-	314998.5	1560960	141360	4.96	2692	152	546
File 1, 5x I	0.17	314630.5	597120	278400	1.90	2687	26	27
File 1, 5x I	0.33	314572.4	454608	292248	1.45	2682	51	51
File 1, 5x I	0.50	314554.2	408672	299424	1.30	2677	52	76
File 1, 5x I	0.67	314550.9	382284	303612	1.22	2672	0	101
File 1, 5x I	1.00	314557.6	359416	306016	1.14	2662	0	151
File 1, 5x I	2.00	314580.0	335960	308956	1.07	2632	0	0
File 1, 10x I	-	314998.1	3087840	139920	9.80	2692	134	490
File 1, 10x I	0.17	314180.8	899376	275328	2.86	2687	26	26
File 1, 10x I	0.33	314071.4	604056	289032	1.92	2682	51	51
File 1, 10x I	0.50	314039.7	507152	296144	1.61	2677	76	76
File 1, 10x I	0.67	314032.2	455340	300288	1.45	2672	101	101
File 1, 10x I	1.00	314040.1	406968	302696	1.30	2662	151	151
File 1, 10x I	2.00	314085.6	358032	305584	1.14	2632	0	301
File 2	-	315001.4	425040	241440	1.35	2204	5	167
File 2	0.17	315007.9	336240	297888	1.07	2199	0	0
File 2	0.33	315001.6	332328	306528	1.06	2194	0	0
File 2	0.50	315003.6	329104	308784	1.04	2189	0	0
File 2	0.67	315005.7	328092	310044	1.04	2184	0	0
File 2	1.00	315002.1	326824	311672	1.04	2174	0	0
File 2	2.00	314952.8	325616	313084	1.03	2144	0	0
File 2, 5x I	-	315001.5	1561200	239040	4.96	2204	8	137
File 2, 5x I	0.17	314554.7	576768	295104	1.83	2199	21	21
File 2, 5x I	0.33	314496.3	450408	303696	1.43	2194	41	41
File 2, 5x I	0.50	314484.4	402384	305984	1.28	2189	57	61
File 2, 5x I	0.67	314482.3	378408	307536	1.20	2184	0	81
File 2, 5x I	1.00	314479.8	356728	308840	1.13	2174	0	121
File 2, 5x I	2.00	314454.1	334512	310232	1.06	2144	0	0
File 2, 10x I	-	315001.3	3087600	236400	9.80	2204	7	105
File 2, 10x I	0.17	314005.8	879024	291792	2.80	2199	21	21
File 2, 10x I	0.33	313884.3	599688	300360	1.91	2194	41	41
File 2, 10x I	0.50	313855.6	500752	302576	1.60	2189	61	61
File 2, 10x I	0.67	313848.3	451356	304092	1.44	2184	81	81
File 2, 10x I	1.00	313847.2	404216	305384	1.29	2174	121	121
File 2, 10x I	2.00	313849.9	356512	306780	1.14	2144	0	241

As can be seen in the table, adding I frames has a large impact on how the averaged bitrate varies. To get a maximum short-term average (measured over the averaging window) that is reasonably close to the long-term average (measured over the whole file) then one need to have a long averaging window, at least a few seconds long.

Looking at the columns showing how many times the average exceeds 25 % and 10 % over the average one can see that this occurs quite frequently when 5x and 10x I frames are added, even for long averaging windows. If the network would drop a packet every time this happens then this would increase the packet loss rate with a few percent, in the worst cases with as much as 10 %.

Another observation is that the frequency of large short-term average actually seems to increase with increasing window length when large I frames are added.

6.11.2 Gap analysis

There is no information in the SDPs (offer or answer) and in the functions and protocols used in PCC and RAN for resource reservation (GDP, Rx, scheduler, etc.) about how large bitrate variations the clients want to use. There is also no information in the SDPs or in the QoS parameters for the bearers that informs the clients about how large bitrate variations the networks will allow.

In addition, for the MBR and GBR bitrates in EPC, [11], there is no definition for how these (average) bitrates should be calculated.

This means that the policing functions in the networks will need to be configured without knowing what will work for the clients. The clients should also implement some form of rate smoothing but a problem here is that client developers do not know how much smoothing that is required by the networks. In addition, the policing functions in different networks could very well be configured differently so a rate smoothing that works in one network is not guaranteed to work in other networks.

6.12 Use case K: Several multi-rate speech codecs (AMR, AMR-WB and EVS), small difference in EVS usage between operators

6.12.1 General description

Alice and Bob are setting up a voice-only session. Both UE-A and UE-B support: the AMR codec (4.75-12.2 kbps, all codec modes); the AMR-WB codec (6.60-23.85 kbps, all codec modes); and the EVS (5.9-128 kbps, all codec modes). Both operators want to use EVS for super-wideband (SWB), but in slightly different ways.

Operator A (originating side) wants to ensure SWB quality in the 9.6-24.4 kbps bitrate range for most operating conditions, while adaptation down to WB or even NB speech at lower bitrates is allowed but should happen quite rarely.

Operator B (terminating side) wants to use the EVS codec only for SWB and only with the 13.2 kbps bitrate.

Table 6.12.1-1: SDP offer-answer for use case K

SDP offer
m=audio 49152 RTP/AVP 101 99 100 97 98 b=AS:50 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:101 EVS/16000/1 a=fmtp:101 br=5.9-24.4; bw=nb-swb; mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 101 b=AS:38 a=rtpmap:101 EVS/16000/1 a=fmtp:101 br=13.2; bw=swb; mode-set=0,1,2; max-red=220 a=ptime:20 a=maxptime:240

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A wants to receive max 50 kbps, based on the b=AS in the SDP offer.
- UE-A will send max 38 kbps, based on the b=AS in the SDP answer.
- UE-B wants to receive max 38 kbps, based on the b=AS in the SDP answer.

- UE-B will send max 50 kbps, based on the b=AS in the SDP offer.
- There is however no information in the SDPs that can be used to set GBR. The br parameter for EVS primary mode in the SDP answer indicates that GBR should be set based on the 13.2 kbps bitrate. However, the mode-set parameter for EVS AMR-WB IO mode shows that all bitrates down to 6.6 kbps are supported.

When offering EVS, the RTP payload type is used for both EVS primary mode and EVS AMR-WB IO mode. Since the offer-answer rules are the same for EVS AMR-WB IO mode as for AMR-WB, an Open Offer is used also for EVS AMR-WB IO mode in the same way as used for AMR-WB. This means that the bandwidth in the SDP offer will set such that it allows for AMR-WB 23.85. However, the bandwidth may also be set to an even higher value if EVS primary modes higher than 24.4 are offered.

6.12.2 Gap analysis

As discussed for use case D, see clause 6.5, the PCRFs may allocate MBR and GBR in several different ways. The downlink MBR in the local access and the uplink MBR in the remote access can be aligned with a second SDP offer-answer. However, there are no mechanisms in SDP to align GBR.

In addition, it is not possible to indicate in SDP a desired bitrate range, if this is different from the allowed bitrate range. So UE-A or Operator A cannot declare that it primarily wants to use the EVS primary mode with at least 9.6 kbps.

6.13 Use case L: Several multi-rate speech codecs (AMR, AMR-WB and EVS), large difference in EVS usage between operators

6.13.1 General description

Alice and Bob are setting up a voice-only session. Both UE-A and UE-B support: the AMR codec (4.75-12.2 kbps, all codec modes); the AMR-WB codec (6.60-23.85 kbps, all codec modes); and the EVS (5.9-128 kbps, all codec modes). The operators want to use the EVS codec in different ways.

Operator A (originating side) wants to ensure SWB-FB quality in the 32-64 kbps bitrate range for most operating conditions, while adaptation down to WB or even NB speech at lower bitrates is allowed but should happen quite rarely. The higher bitrate range could, for example, be desirable in conference calls.

Operator B (terminating side) wants to use the EVS codec primarily between WB and SWB and in the 5.9-13.2 kbps bitrate range.

Table 6.13.1-1: SDP offer-answer for use case L

SDP offer
m=audio 49152 RTP/AVP 101 99 100 97 98 b=AS:89 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:101 EVS/16000/1 a=fmtp:101 br=5.9-64; bw=nb-fb; mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 101 b=AS:38 a=rtpmap:101 EVS/16000/1 a=fmtp:101 br=5.9-13.2; bw=nb-swb; mode-set=0,1,2; max-red=220 a=ptime:20 a=maxptime:240

For the media handling in the UEs, the SDP offer/answer negotiation means:

- UE-A wants to receive max 89 kbps, based on the b=AS in the SDP offer.
- UE-A will send max 38 kbps, based on the b=AS in the SDP answer.
- UE-B wants to receive max 38 kbps, based on the b=AS in the SDP answer.
- UE-B will send max 89 kbps, based on the b=AS in the SDP offer.

After a second offer-answer negotiation (if performed), it is expected that the maximum bandwidths become 38 kbps in both directions.

There are no mechanisms in SDP that originating client or Operator A can use to show to the other network and the terminating client that it really wants to use a much higher bitrate range (32-64 kbps instead of <=13.2 kbps).

6.13.2 Gap analysis

As discussed for use case K, see clause 6.12, there are no mechanisms in SDP to align GBR. In use case K, the differences between MBR/GBR and the desired operating bitrate ranges are relatively small. However, in this case the selected bitrate range is much lower than what Operator A would prefer to use.

7 Recommended requirements

7.1 Discussion on individual recommended requirements

7.1.1 General

In this section, atomic requirements for each individual use case are derived from the respective gaps. The atomic requirements are then merged into complete requirements where possible, which are summarized in clause 7.4.

7.1.2 Use case A: Single fixed-rate speech codec

Gap(s):

No gap identified.

Proposed requirement(s):

None.

7.1.3 Use case B: Several fixed-rate speech codecs

Gap(s) after first SDP offer/answer:

Both over-allocation and under-allocation can happen, since b=AS only indicates one single value and therefore has to be set to the maximum bandwidth needed for the codec that uses the highest bitrate.

Over-allocation typically happens when the resources are set up for a high-bitrate codec but then a lower bitrate codec is negotiated. Under-allocation typically happens if a network assigns too high MBR and GBR values and if the other networks assign MBR and GBR with lower values. This can lead to misalignment of both MBR and GBR between the different networks.

Gap(s) after second SDP offer/answer:

If both UEs use the selected codec in the same way, then it should be possible to avoid the gap.

However, if a UE wants to use the selected codec in a different way, for example with redundancy, then it is only possible to indicate this for the receiving direction. There is no possibility to indicate the maximum or desired bitrate for the sending direction.

Proposed requirement(s):

The following requirement is proposed: "*It should be possible to make the network aware of the minimum and maximum supported bandwidth requirements negotiated between the UEs for each media direction.*"

Comments:

The proposed requirement is an aggregation of several "atomic" requirements. The analysis of this use case supports the following atomic requirements:

- It should be possible to make the network aware of the minimum supported bandwidth requirements negotiated between the UEs for the receiving direction.
- It should be possible to make the network aware of the maximum supported bandwidth requirements negotiated between the UEs for the receiving direction.

The corresponding atomic requirements for the sending direction are discussed for Use case D, see clause 7.1.5.

7.1.4 Use case C: Single multi-rate speech codec (AMR), no extra bandwidth allocated for redundancy

Gap(s):

For AMR, the UE signals which minimum and maximum bitrates that are supported. This is either done by defining a mode-set, which then explicitly shows what lowest and highest codec modes that are allowed, or by not defining any mode-set, which then implicitly shows that the lowest and highest codec modes defined for the codec are allowed for the session.

It is here assumed that the maximum desired bitrate is the same as the maximum allowed bitrate and can thus be derived from the allowed mode-set. However, it is not always possible to derive the minimum desired bitrate since different operators in the path may want to ensure different quality levels even if they allow adapting to even lower bitrates.

There is no signalling in SDP of the minimum desired bitrate so the UEs will not know the desired minimum bitrate for the other UE, if it is different from the minimum supported bitrate. Correspondingly, the networks will also not know the minimum desired bitrates for the UEs and the other networks, unless it is the same as the minimum supported bitrate.

Comments:

For example, a session may be set up to allow for using the AMR 12.2, 7.4, 5.9 and 4.75 kbps codec modes, but UE-A or Operator A may want to use at the minimum the 5.9 kbps codec mode for most sessions and may consider using the 4.75 kbps codec mode only in the worst case. Correspondingly, UE-B or Operator B may want to use at the minimum the 7.4 kbps codec mode for most sessions but may consider using the 4.75 codec mode in the worst case. Operator A and Operator B will then set up the bearers to UE-A and UE-B differently. They should use the same value for MBR, but it can be expected that the GBR values are different and defined according to each operator's desired minimum bitrate.

Proposed requirement(s):

The following requirement is proposed: "*It should be possible to make the network aware of the minimum and maximum desired bandwidth requirements negotiated between the UEs for each media direction, if this is different from the supported bandwidths.*"

Comments:

This proposed requirement is also an aggregation of several "atomic" requirements. This use case supports the following atomic requirement:

- It should be possible to make the network aware of the minimum desired bandwidth requirements negotiated between the UEs for the receiving direction.

The corresponding atomic requirement for the minimum desired bandwidth for the sending direction is discussed for Use case D in clause 7.1.5.

The corresponding atomic requirements for the maximum desired bandwidths are discussed for Use case K, see clause 7.1.12.

7.1.5 Use case D: Single multi-rate speech codec (AMR) with extra bandwidth allocated for redundancy

Gap(s):

See gaps for Use case C. This use case identifies that there are no mechanisms in SDP to indicate the minimum desired bitrate for the sending direction.

Proposed requirement(s):

The proposed requirement is the same as for Use case C, see clause 7.1.4. No new requirements are needed for this use case.

Comments:

The atomic requirement for the minimum desired bandwidths for the receiving direction is supported by Use case C. This use case supports the atomic requirement for the minimum desired bandwidths for the sending direction.

7.1.6 Use case E: Several multi-rate speech codecs (AMR and AMR-WB)

Gap(s):

Same as for use cases B and C.

Proposed requirement(s):

The proposed requirement is the same as for Use cases B and C, see clauses 7.1.3 and 7.1.4, respectively. No new requirements are needed for this use case.

7.1.7 Use case F: Single video codec, symmetric usage

Gap(s):

Same as for use case C. However, for video codecs there is typically no signalling of the minimum supported bitrate. This means that a UE may adapt quite frequently down to quality levels (bitrate, frame rate) that are significantly lower than the operator preferences, especially if the remote network assigns a lower GBR value than what used in the local network.

Proposed requirement(s):

The proposed requirement is the same as for Use case C, see clause 7.1.4. No new requirements are needed for this use case.

7.1.8 Use case G: Single video codec, asymmetric usage, sending video with a bitrate matching the codec level

Gap(s):

Same as for use cases D. There are codec parameters to indicate a higher (but not lower) codec level for the receiving direction than for the sending direction. The b=AS bandwidth indicates only the bitrate in the receiving direction and there is no corresponding parameter for the sending direction. The maximum bitrate in the sending direction can however be derived from the codec level applicable to the sending direction.

Proposed requirement(s):

The proposed requirement is the same as for Use case D, see clause 7.1.5. No new requirements are needed for this use case.

7.1.9 Use case H: Single video codec, asymmetric usage, sending video with a bitrate lower than the supported codec level

Gap(s):

Similar to use cases D and G but the maximum bitrate for the sending direction cannot be derived from the codec level that is applicable to the sending direction.

Proposed requirement(s):

No new requirements are needed for this use case.

7.1.10 Use case I: Multiple video codecs

Gap(s):

Same as for use cases C and E. The difference from use case C is that there is no information about the minimum supported bitrates in the SDPs, so the network has less information that it can use when assigning resources.

Proposed requirement(s):

The proposed requirement is the same as for Use cases C and E, see clauses 7.1.4 and 7.1.6, respectively. No new requirements are needed for this use case.

7.1.11 Use case J: Single video codec, symmetric usage, bitrate variations

Gap(s):

There is no information in the SDPs that informs the networks about the bitrate variations that the UEs would like to utilize.

There is also no information in the SDPs or in the QoS parameters where the network can indicate how large bitrate variations that are allowed.

There is also no definition in the EPC specifications of how the (average) bitrate is calculated.

Proposed requirement(s):

The following requirement is proposed: "*It should be possible to make the clients aware of what bitrate variations are allowed or how the bitrate average is calculated, e.g. in policing functions.*"

7.1.12 Use case K: Several multi-rate speech codecs (AMR, AMR-WB and EVS), small difference in EVS usage between operators

Gap(s):

Same as for Use cases B, C, D, E and I with the addition that there is no information in SDP about the maximum desired bandwidths for sending and receiving directions.

Proposed requirement(s):

The proposed requirements are the same as described for Use cases B and C, see clauses 7.1.3 and 7.1.5, respectively. No new requirements are needed for this use case.

Comments:

The atomic requirements for the minimum desired bandwidths for receiving and sending directions are supported by Use cases C and D, respectively. This use case supports the atomic requirements for the maximum desired bandwidths for sending and receiving directions.

7.1.13 Use case L: Several multi-rate speech codecs (AMR, AMR-WB and EVS), large difference in EVS usage between operators

Gap(s):

Same type of gaps as identified for Use cases K although the gaps are larger for Use case L.

Proposed requirement(s):

The proposed requirements are the same as described for Use cases K.

Comments:

In Use case K, the differences between wanted bandwidth parameters and allocated bandwidth parameters should in most cases result in a quality that is close to the user expectations. In this case, the differences between bandwidth parameters and allocated bandwidth parameters are larger and this gives an increasing risk that the resulting quality level is not sufficient to meet the user expectations, which gives an increasing risk that users become unsatisfied. It is thus more important to solve the issues and to handle the bandwidth negotiation in a better way than how it is done in current systems.

7.2 Discussion on proposed requirements for new SDP attributes

The natural way of conveying information related to session and/or media bandwidths between the clients and the network nodes is to use SDP, especially for the end-to-end case. It is therefore foreseen that new SDP attributes may need to be designed. Such new SDP attributes should preferably be future proof and so that they can be extended to also fulfil the needs of services and media that are not considered at this point in time. This gives the following proposed requirement: "*New SDP attribute(s) should allow for future extensions.*"

Any new SDP attributes also need to be backwards compatible with other already existing SDP attributes. The introduction of new SDP attributes should not prevent the usage of other SDP attributes. This gives the following proposed requirement. "*New SDP attribute(s) should be backwards compatible with existing attributes and offer/answer negotiation process.*"

When introducing new SDP attributes, it is likely that this is done in a gradual, incremental fashion. Some networks may add support for the new SDP attributes early while it can be expected that other networks continue to use only the legacy SDP attributes, possibly for a long time. Existing resource allocation functions need to work as good or as bad as they work today also in the future when new SDP attributes are used. This gives the following proposed requirement: "*The existing functionality for resource allocation in legacy networks should not be affected by the introduction of new SDP attributes.*"

7.3 Discussion on proposed general requirements for the solution

The use cases in the present document have discussed speech and video media and a limited number of codecs. For any solution to become widely accepted and used, the solution needs to be designed in such a way that it is generic and can be reused for any service, any media, any codec and any codec configuration. This gives the following proposed requirement: "*The solution should be generic and reusable for all services, media types and codecs.*"

With a generic solution follows also a need to make it specific, so that it can be optimized to work well for any service, any media, any codec and any codec configuration. The solution therefore needs to be configurable. This gives the following proposed requirement: "*The solution should allow configuration in clients and/or network nodes to be able to adapt its usage to different services, media types and codecs.*"

A difference from the existing QoS mechanisms used for the local resource reservation is that the intention in this work is to find a solution that can be used by all networks in the path as well as the end-points. This gives the following proposed requirement: "*The solution should be usable by all networks and end-points.*"

7.4 Summary of proposed requirements

Proposed requirements:

A It should be possible for the SDP offerer to indicate the minimum and maximum supported bandwidth for each offered codec of a media component and for each media direction, and for the network to reduce the maximum supported bandwidth according to its policies.

B It should be possible for the SDP offerer to indicate the minimum and maximum desired bandwidth for each offered codec of a media component and for each media direction, and for the network to reduce the minimum and maximum desired bandwidth according to its policies.

NOTE 1: The minimum supported bandwidth provides a lower boundary for the network how far it can reduce the minimum desired bandwidth.

C It should be possible to make the network aware of the maximum supported bandwidth requirements for the negotiated codec of a media component and for each media direction, as negotiated between the UEs.

NOTE 2: The maximum supported bandwidth may be used to derive the MBR and/or for policy enforcement. For MBR=GBR bearers, the maximum supported bandwidth may also be used to derive the GBR. The b=AS SDP bandwidth modifier in the SDP offer applies to the set of codecs in the SDP offer rather than the negotiated codec.

D It should be possible for the SDP answerer to make the offerer aware of all bandwidth properties (minimum and maximum, supported and desired) for the negotiated codec of a media component and for each media direction.

E It should be possible to make the network aware of the minimum and maximum desired bandwidth requirements for the negotiated codec of a media component and for each media direction, as negotiated between the UEs.

NOTE 3: For MBR>GBR bearers, the minimum desired bandwidth may be used to derive the GBR. The minimum and/or maximum desired bandwidths may be used for the admission control, e.g. to determine how many users can be admitted in order to ensure a certain desired quality level.

F It should be possible for the clients to know what bitrate variations are allowed or how the bitrate average is calculated, e.g. in the policing functions.

Proposed requirements for the design of new SDP attributes:

G New SDP attribute(s) should allow for future extensions.

H New SDP attribute(s) should be backwards compatible with existing attributes and offer/answer negotiation process.

I The existing functionality for resource allocation in legacy networks should not be affected by the introduction of new SDP attributes.

NOTE 4: Since legacy networks are expected to ignore any new SDP attributes, the UEs cannot assume that all networks in the path use the information included in the new SDP attributes.

Proposed requirements for the solution design:

J The solution should be generic and reusable for all services, media types and codecs.

K The solution should allow configuration in clients and/or network nodes to be able to adapt its usage to different services, media types and codecs.

L The solution should be usable by all networks and end-points.

M The solution should give clear definitions of the new bandwidth information and procedures on how the information should be used by networks and clients.

7.5 Definition of new bandwidth information parameters

7.5.1 Overview

This section provides more detailed definitions for the new bandwidth parameters. It also describes how the parameters should be used at session setup and during the session. The impacts on media quality are also discussed.

The most commonly used bandwidths are in the range from the Minimum Desired Bandwidth to the Maximum Desired Bandwidth. The higher end of this range should preferably be used for most sessions.

Bandwidths below the Minimum Desired Bandwidth, down to the Minimum Supported Bandwidth, may be used during poor operating conditions, although should happen rarely. If the Minimum Supported Bandwidth cannot be maintained then the media may be dropped or the session may be closed.

Bandwidths above the Maximum Desired Bandwidth, up to the Maximum Supported Bandwidth, can be used to provide room for redundancy so that the media may survive during very bad operating conditions and when reducing the bandwidth is unable to provide sufficient quality. This range should be used rarely.

This means that the following relationships apply:

- Minimum Supported Bandwidth <= Minimum Desired Bandwidth
- Minimum Desired Bandwidth <= Maximum Desired Bandwidth
- Maximum Desired Bandwidth <= Maximum Supported Bandwidth

All bandwidth parameters describe here include the IP, UDP and RTP overhead. Corresponding bandwidth parameters excluding the IP, UDP and RTP overhead may also be defined but are not included here.

7.5.2 Maximum Supported Bandwidth

Definition:

Identifies the highest bandwidth that can be used in the session during non-normal operating conditions.

Should be used to set MBR.

Should also be used to set GBR for MBR=GBR bearers.

Usage during the session:

The additional bandwidth for redundancy should only be used if adapting the bitrate to lower values, down to the Minimum Supported Bandwidth, fails to provide sufficient quality.

Quality aspects:

When additional bandwidth is allocated for redundancy, the resilience against losses should be improved. It should however be expected that the end-to-end delay will be longer than for the normal operating range.

7.5.3 Maximum Desired Bandwidth

Definition:

Identifies the highest bandwidth that should be used in most cases during normal operating conditions.

Usage during the session:

The adaptation should ensure that bandwidths up to the Maximum Desired Bandwidth are used whenever possible.

Quality aspects:

Using bandwidths in the higher end of the Minimum Desired Bandwidth ~ Maximum Desired Bandwidth range should give the intended encoding quality and end-to-end delay.

7.5.4 Minimum Desired Bandwidth

Definition:

Identifies the lowest bandwidth that should be used in the session during relatively normal or slightly degraded operating conditions.

Used for setting GBR for MBR>GBR bearers.

Usage during the session:

Bandwidths in the lower end of the Minimum Desired Bandwidth ~ Maximum Desired Bandwidth should be used less frequently, mainly during periods with high load and/or degraded operating conditions.

The used bandwidth can be lower than the Minimum Desired Bandwidth, for example, during DTX periods for speech or when DTMF is being transmitted instead of speech.

Quality aspects:

Using bandwidths in the lower end of this range can give slightly reduced encoding quality but should not give increased end-to-end delay.

For video, this bandwidth should be selected such that the video is still relatively smooth.

7.5.5 Minimum Supported Bandwidth

Definition:

Identifies the lowest bandwidth that may be used in the session during non-normal operating conditions, primarily when limited by the bitrate.

Usage during the session:

Bandwidths below the Minimum Desired Bandwidth, down to the Minimum Supported Bandwidth, should be used quite rarely, mainly for severely degraded operating conditions.

If the Minimum Supported Bandwidth cannot be maintained then the session can be terminated.

The used bandwidth can be lower than the Minimum Supported Bandwidth, for example, during DTX periods for speech or when DTMF is being transmitted instead of speech.

Quality aspects:

It can be expected that the encoding quality is reduced for these bandwidths and also that the end-to-end delay is increased.

8 Potential solution(s)

8.1 Potential solution A: Session re-negotiation

8.1.1 Introduction

This solution describes potential new requirements and/or recommendations to TS 26.114 to clarify when a session re-negotiations should be performed.

In TS 26.114 clause 6.2, many requirements and recommendations are defined for the original session setup, i.e. for the first SDP offer-answer negotiation. There are also some requirements and recommendations for session re-negotiations when adding or removing media during a session. However, there are quite few requirements or recommendations for session re-negotiation for aligning the bandwidths to the selected codec. The same applies to the SDP examples in TS 26.114 Annex A.

8.1.2 Description of the solution

The proposed solution is to add requirements and/or recommendations in TS 26.114 clause 6.2 to clarify the conditions when an MTSI client should perform a session re-negotiation. For example, a re-negotiation may be useful for the following cases:

- If the SDP offer included multiple codecs and/or codec configurations requiring different bandwidths, and if the SDP answer includes a codec configuration that only needs a lower bandwidth.
- If the bandwidth in the SDP offer included room for redundancy but the SDP answer prevented using redundancy in the session.
- If the bandwidth in the SDP offer was selected to send speech with one frame per packet but the SDP answer indicated that frame aggregation is needed.
- If the original intention was to send media without redundancy ('max-red=0') but the remote end-point wants to receive redundancy and indicates this by sending an SDP answer with a bandwidth that is higher than what is needed for the codec.

Whether a re-negotiation is performed or not may also depend on whether the intention is to set up a symmetric or asymmetric session.

A few SDP examples may also be added in TS 26.114 Annex A to describe both the original negotiation and the re-negotiation.

The list above indicates only a few examples. More examples can be added later during the normative work.

In TS 26.114, clause 6.2.7, the following is defined:

> "If the MTSI client in terminal determines that the b=AS bandwidth(s) are not aligned with the MBR and the receiving capabilities of the MTSI client, then it should align the media-level b=AS bandwidth(s) to the MBR and its receiving capabilities by sending to the other party an SDP offer with the new b=AS bandwidth value(s). In the process of this alignment it is also likely that the session-level b=AS bandwidth needs to be updated. In addition, the MTSI client in terminal may modify other parts of the SDP, e.g. to replace the codecs or adjust codec parameters (such as the AMR or AMR-WB mode-set)."

This describes when a session re-negotiation should be performed to align the b=AS bandwidth with the downlink MBR. If new bandwidth modifiers or attributes are defined then it is likely that the above text may need to be updated accordingly. The modifications that are needed however depend on the chosen solution(s). The detailed modifications can therefore not be described at this point in time.

8.1.3 Compliance with proposed requirements

The above solution fulfils the requirement C on making the networks aware of the maximum supported bandwidth but only for the receiving direction of each end-point. This solution cannot itself fulfil the corresponding requirements for the sending direction because there are no bandwidth modifiers or attributes defined in SDP for this purpose. For the same reason, this solution can also not fulfil the requirement E on making the networks aware of the minimum supported or desired bandwidths, neither for the sending nor for the receiving direction. However, if combined with any of the proposed solution B to G, those requirements can also be addressed.

8.1.4 Impact on networks and terminals

The mechanisms needed to perform a session re-negotiation are supported already today. It is therefore expected that this solution have no impact on the architecture and the interfaces in the networks, for example for the PCC.

It is however expected that UEs and MGWs (and possibly some other nodes) use some form of decision logic to determine when a session re-negotiation should be performed. This decision logic may need to be updated.

TS 26.114 clause 6.2 targets terminals and MGWs, which means that updates to this clause would impact these nodes. Updates to TS 26.114 clause 6.2.7 should however not impact MGWs since they are not expected to have any knowledge about the QoS parameters for the bearers and would therefore not try to align any SDP information and any QoS parameters.

8.2 Potential solution B: New bandwidth modifiers in SDP offer and answer without SDP MiscCapNeg

8.2.1 Introduction

This solution describes how the clients can make the networks aware and each other of the negotiated maximum supported bandwidth, the minimum supported bandwidth, the maximum desired bandwidth and the minimum desired bandwidth for each direction by defining new bandwidth modifiers to carry the new bandwidth information.

8.2.2 Description of the solution

8.2.2.1 General solution

The general solution is to add information in the SDPs about the maximum supported bandwidth, minimum supported bandwidth, maximum desired bandwidth and the minimum desired bandwidth, for sending and receiving directions, respectively. The clients would negotiate these bandwidths in the same way as they negotiate other configuration parameters. The networks would use these bandwidths in the session setup and session re-negotiations, both for the admission control and for resource reservation.

8.2.2.2 New bandwidth modifiers

The following new bandwidth modifiers are needed:

- `b=AS_max_des_recv:<value>` - maximum desired bandwidth in receiving direction
- `b=AS_max_des_send:<value>` - maximum desired bandwidth in sending direction
- `b=AS_min_des_recv:<value>` - minimum desired bandwidth in receiving direction
- `b=AS_min_des_send:<value>` - minimum desired bandwidth in sending direction
- `b=AS_max_sup_recv:<value>` - maximum supported bandwidth in receiving direction (same as b=AS)
- `b=AS_max_sup_send:<value>` - maximum supported bandwidth in sending direction
- `b=AS_min_sup_recv:<value>` - minimum supported bandwidth in receiving direction
- `b=AS_min_sup_send:<value>` - minimum supported bandwidth in sending direction

The names of the new bandwidth modifiers can of course be changed.

One limitation with defining new bandwidth modifiers is the syntax for bandwidth modifiers defined in SDP [8]:

```
b=<bwtype>:<bandwidth>
```

This syntax prevents defining different bandwidths for different RTP payload types, which could be solved by using SDP miscellaneous capability negotiation (SDPMiscCapNeg) [14]. Such a solution is described in clause 8.3.

The bandwidth value is expressed in kbps since this is the default unit for bandwidth modifiers, which is also used for the b=AS value.

8.2.2.3 Procedures

The SDP offer can contain several RTP payload types corresponding to different codecs and configurations, where the different configurations have different bandwidth needs. The originating client includes the maximum values required for any of the offered codec and configurations as bandwidth information.

NOTE: According to the general semantics of bandwidth modifiers in IETF RFC 4566, they relate to the entire m-line. With only one set of new bandwidth modifiers it is not possible to identify the bandwidth needs for each offered codec and configuration.

The answerer selects a codec and configuration which complies with the received bandwidth information within the SDP offer (i.e. the required bandwidth is equal or below the received information for each type of bandwidth information). In the SDP answer, the answerer provides bandwidth information for both the send and receive direction for the media line that matches the needs of the codec and configuration it selects. The answerer provides bandwidth information which is equal or below the corresponding received bandwidth information within the SDP offer for the maximum supported, maximum desired and minimum desired bandwidths.

8.2.2.4 Session negotiation example

An example of how the new bandwidth modifiers can be used in the session negotiation is shown below. This example is based on Use case E where both AMR-WB and AMR are offered but where AMR is negotiated, see clause 6.6 and Table 6.6.1-1. A difference from Use case E is that the offer allows for using 100% redundancy even when the highest codec mode is used.

The new bandwidth modifiers are highlighted with bold font.

NOTE: The numerical values in those examples have not been carefully checked and can contain errors.

Table 8.2.2.4-1: SDP offer-answer for proposed solution B with new bandwidth modifiers

SDP offer
m=audio 49152 RTP/AVP 100 99 98 97 b=AS:73 **b=AS_max_des_recv:49** **b=AS_max_des_send:49** **b=AS_min_des_recv:34** **b=AS_min_des_send:34** **b=AS_max_sup_recv:73** **b=AS_max_sup_send:73** **b=AS_min_sup_recv:13** **b=AS_min_sup_send:13** a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:50 **b=AS_max_des_recv:37** **b=AS_max_des_send:37** **b=AS_min_des_recv:31** **b=AS_min_des_send:31** **b=AS_max_sup_recv:50** **b=AS_max_sup_send:50** **b=AS_min_sup_recv:12** **b=AS_min_sup_send:12** a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240

The bandwidth value for the b=AS parameter in the SDP offer is derived using existing rules, which in this case means using RTP payload type 99, i.e. AMR-WB, max 23.85 kbps and octet-aligned payload format. No extra bandwidth is allocated for redundancy.

In this case, a symmetric session is assumed. The new bandwidth values are therefore the same for the sending and receiving directions.

The values for new bandwidth modifiers shown in the SDP offer are derived for the most preferred configuration (100), i.e. AMR-WB, 23.85 kbps and bandwidth-efficient payload format:

- `b=AS_max_des_recv:49` - AMR-WB 23.85 kbps with no redundancy
- `b=AS_max_des_send:49` - AMR-WB 23.85 kbps with no redundancy
- `b=AS_min_des_recv:34` - AMR-WB 8.85 kbps with no redundancy
- `b=AS_min_des_send:34` - AMR-WB 8.85 kbps with no redundancy
- `b=AS_max_sup_recv:73` - AMR-WB 23.85 kbps with 100% redundancy, note that this is different from the b=AS value
- `b=AS_max_sup_send:73` - AMR-WB 23.85 kbps with 100% redundancy
- `b=AS_min_sup_recv:13` - AMR-WB 6.60 kbps with no redundancy, 4 frames per packet
- `b=AS_min_sup_send:13` - AMR-WB 6.60 kbps with no redundancy, 4 frames per packet

The bandwidth value for the b=AS parameter in the SDP answer is also derived using existing rules, which in this case means using RTP payload type 97, i.e. AMR, max 123.2 kbps and bandwidth-efficient payload format. No extra bandwidth is allocated for redundancy.

The values for new bandwidth modifiers shown in the SDP answer are derived from the selected configuration, i.e. AMR, max 12.2 kbps, bandwidth-efficient payload format:

- `b=AS_max_des_recv:37` - AMR 12.2 kbps with no redundancy
- `b=AS_max_des_send:37` - AMR 12.2 kbps with no redundancy
- `b=AS_min_des_recv:31` - AMR 5.9 kbps with no redundancy
- `b=AS_min_des_send:31` - AMR 5.9 kbps with no redundancy
- `b=AS_max_sup_recv:50` - AMR 12.2 kbps with 100% redundancy, note that this is different from the b=AS value
- `b=AS_max_sup_send:50` - AMR 12.2 kbps with 100% redundancy
- `b=AS_min_sup_recv:12` - AMR 4.75 kbps with no redundancy, 4 frames per packet
- `b=AS_min_sup_send:12` - AMR 4.75 kbps with no redundancy, 4 frames per packet

If the originating client accepts the bandwidths proposed by the terminating client then no further SDP offer-answer negotiations are needed, at least not for the reason of negotiating the bandwidths. However, since the terminating client selected a configuration that was not the most preferred by the originating client, it can happen that the originating client is not fully satisfied with the proposed bandwidths shown in the SDP answer. In this case, the originating client would need to send a SIP update to initiate a new SDP offer-answer negotiation.

In the worst case, one may even need several additional offer-answer negotiations to conclude on the configuration to use for the session. This would however increase the session setup time, add load on the SIP bearer and also add load to the SIP servers, which is undesirable.

8.2.2.5 Modifying the bandwidth information in the SDP offer

In the SDP offer, the first node can decrease the maximum supported, maximum desired and minimum desired bandwidth according to network policies. However, the first node does normally not increase the maximum supported and maximum desired and minimum desired bandwidth except when required to correct undesired or erroneous UE behavior, when adding codecs accessible via transcoding, or for IPv4/IPv6 transport interworking. The first node can increase the minimum supported bandwidth according to network policies. However, the first node does normally not decrease the minimum supported bandwidth except when required to correct UE misbehavior, when removing offered codecs, or for IPv4/IPv6 transport interworking.

In the SDP offer, subsequent nodes can decrease the maximum supported, maximum desired and minimum desired bandwidth according to network policies. However, they are supposed to not increase the maximum supported and maximum desired and minimum desired bandwidth except when adding codecs accessible via transcoding, or for IPv4/IPv6 transport interworking. They can increase the minimum supported bandwidth according to network policies. However, they are supposed to not decrease the minimum supported bandwidth except when removing offered codecs, or for IPv4/IPv6 transport interworking.

The following relationships are maintained by any network node when modifying bandwidths:

Minimum Supported Bandwidth <= Minimum Desired Bandwidth

Minimum Desired Bandwidth <= Maximum Desired Bandwidth

Maximum Desired Bandwidth <= Maximum Supported Bandwidth

If a network node desires to use a MBR=GBR bearer, it preferably decreases maximum supported bandwidth down to the maximum desired bandwidth in the SDP offer.

NOTE 1: These rules allow all both the originating and terminating operators in the call setup direction to implement certain policies, but avoid that subsequent operators in the call setup chain counteract the policies of the first operator, and guarantee that the used bandwidths remain in the supported range of the originating UE.
For instance, the following policies are supported:
An operator desiring to set a lower limit to the acceptable QoS can increase the Minimum Supported Bandwidth.
An operator desiring to limit the GBR for MBR>GBR bearers can decrease the Minimum Desired Bandwidth.
An operator desiring to limit the GBR for MBR=GBR bearers can decrease the Maximum Supported Bandwidth.

In the SDP answer, the first node does normally not modify the bandwidth values except when required to correct UE misbehavior, when replacing the selected codec and configuration in the SDP answer due to transcoding, or due to IPv4/IPv6 transport interworking. Subsequent node also are supposed to not modify the bandwidth values except when replacing the selected codec and configuration in the SDP answer due to transcoding, or due to IPv4/IPv6 transport interworking.

NOTE 2: To update the terminating client, the originating client would need to send a SIP UPDATE including the new bandwidth information. However, the originating client does not know if the bandwidth information in the SDP answer came from the terminating client or if the network changed this information, so it does not know that a SIP UPDATE would be needed.

Networks also have the possibility to reject the SDPs if the indicated bandwidths are unreasonable, as can be done already today.

8.2.2.6 Resource reservation in different networks

Since the bandwidth information in both directions is included in the SDP answer, all networks in the path have the same information and can use this instead of proprietary codec-specific algorithms both for the admission control and for resource reservation. Thereby, it is possible to align QoS end-to-end.

It should be noted that this does not prevent using operator policies, even if the operator policies would use different bandwidths than indicated in the SDP offer. However, in this case, it would be beneficial to modify the SDP offer before forwarding it to the next network so that the bandwidth information in the SDP offer is aligned with the selected QoS parameters.

8.2.3 Compliance with proposed requirements

The described proposed solution addresses the proposed requirements A to E. However, those requirements are only partially met, as the bandwidth information is provided per media component, rather than per codec and configuration. The proposed solution is compliant with the solution design requirements J to M. SDP attribute related requirements G to I are not applicable to this solution.

8.2.4 Impact on networks and terminals

Adding new information in SDP means that terminals and networks would need to support the new SDP parameters in order to make the solution useful.

For the networks, the AF (P-CSCF) would need to extract the new information from the SDP answerer and provide corresponding session information to the PCRF. Compared to the existing procedures, the AF or PCRF would no longer have to use information in the SDP offer to derive bandwidth for resource reservation purposes, but would have to extract more bandwidth information from the SDP answer. The PCRF would no longer need to use a codec-specific algorithm. The related detailed impacts on the Rx interface and AF and PCRF procedures will be determined by CT3 during their normative work. The PCRF would then use the session information to set the QoS parameters on the Gx interface. The QoS parameters that are used on the Gx interface are the same as in the existing specifications. This means that there is no need to change the PCEF, the RAN or the interfaces to these nodes.

Nodes in the network applying existing policies to modify SDP offer for transcoding purposes would need to be updated to adjust the new bandwidth information accordingly.

Transcoding nodes within the network would need to be updated to take into account the new bandwidth information when encoding media. The new bandwidth information would need to be provided to the IMS-AGW, TrGW and MRFP via the Iq interface, Ix interface and Mp interfaces, respectively.

Adding new SDP parameters also gives automatic fallback to the legacy solution whenever the new SDP parameters are not supported. This ensures backwards compatibility as long as the SDP still contains the old information, i.e. the b=AS bandwidth modifier.

8.3 Potential solution C: New bandwidth modifiers and SDPMiscCapNeg in SDP offer and answer

8.3.1 Introduction

In this solution, the new bandwidth modifiers from solution B are used together with SDP Miscellaneous Capability Negotiation (SDPMiscCapNeg) [14] to be able to identify different bandwidth for different RTP payload types in the SDP offer.

8.3.2 Description of the solution

8.3.2.1 General solution

The general solution is to add information in the SDPs about the maximum supported bandwidth, minimum supported bandwidth, maximum desired bandwidth and the minimum desired bandwidth, for sending and receiving directions, respectively, by defining a SDP bandwidth modifier to carry the new bandwidth information. SDP Miscellaneous Capability Negotiation (SDPMiscCapNeg) [14] is used to identify different bandwidth modifiers for different RTP payload types. The clients would negotiate these bandwidths in the same way as they negotiate other configuration parameters. The networks would use these bandwidths in the session setup and session re-negotiations, both for the admission control and for resource reservation.

8.3.2.2 Procedures

The SDP offer can contain several RTP payload types corresponding to different codecs and configurations, where the different configurations have different bandwidth needs. The originating client includes bandwidth values required for each of the offered codec and configurations as bandwidth information, using separate SDP capneg configuration with the new SDP bandwidth modifiers encapsulated in SDP "a=bcap" attributes.

The answerer selects a codec and configuration. In the SDP answer, the answerer provides bandwidth information for both the send and receive direction for the codec and configuration it selects. The answerer provides bandwidth information which is equal or below the corresponding received bandwidth information within the SDP offer for maximum values and for the minimum desired value, and which is equal or higher for the minimum supported value.

8.3.2.3 Session negotiation example

The new bandwidth modifiers and the new attributes for SDPMiscCapNeg are highlighted with bold font. The new bandwidth modifiers are included here in the same way as shown in solution B. This is to ensure compatibility with clients that do not support SPDMiscCapNeg, but it may not always be possible to do this.

NOTE: The numerical values in those examples have not been carefully checked and can contain errors.

Table 8.3.2.3-1: SDP offer-answer for proposed solution C with new bandwidth modifiers

SDP offer
a=csup:cap-v0,med-v0,bcap-v0
m=audio 49152 RTP/AVP 100 99 98 97
b=AS:73
b=AS_max_des_recv:49
b=AS_max_des_send:49
b=AS_min_des_recv:34
b=AS_min_des_send:34
b=AS_max_sup_recv:73
b=AS_max_sup_send:73
b=AS_min_sup_recv:13
b=AS_min_sup_send:13
// AMR, bandwidth-efficient
a=bcap:3 AS:50
a=bcap:31 AS_max_des_recv:37
a=bcap:32 AS_max_des_send:37
a=bcap:33 AS_min_des_recv:31
a=bcap:34 AS_min_des_send:31
a=bcap:35 AS_max_sup_recv:50
a=bcap:36 AS_max_sup_send:50
a=bcap:37 AS_min_sup_recv:12
a=bcap:38 AS_min_sup_send:12
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rmcap:3 AMR/8000/1
a=mfcap:3 mode-change-capability=2; max-red=220
// AMR, octet-aligned
a=bcap:4 AS:50
a=bcap:41 AS_max_des_recv:38
a=bcap:42 AS_max_des_send:38
a=bcap:43 AS_min_des_recv:31
a=bcap:44 AS_min_des_send:31
a=bcap:45 AS_max_sup_recv:50
a=bcap:46 AS_max_sup_send:50
a=bcap:47 AS_min_sup_recv:12
a=bcap:48 AS_min_sup_send:12
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=rmcap:4 AMR/8000/1
a=mfcap:4 mode-change-capability=2; max-red=220; octet-align=1
// AMR-WB, bandwidth-efficient
a=bcap:1 AS:73
a=bcap:11 AS_max_des_recv:49
a=bcap:12 AS_max_des_send:49
a=bcap:13 AS_min_des_recv:34
a=bcap:14 AS_min_des_send:34
a=bcap:15 AS_max_sup_recv:73
a=bcap:16 AS_max_sup_send:73

```
a=bcap:17 AS_min_sup_recv:13
a=bcap:18 AS_min_sup_send:13
a=rtpmap:99 AMR-WB/16000/1
a=fmtp:99 mode-change-capability=2; max-red=220
a=rmcap:1 AMR-WB/16000/1
a=mfcap:1 mode-change-capability=2; max-red=220

// AMR-WB, octet-aligned
a=bcap:2 AS:74
a=bcap:21 AS_max_des_recv:49
a=bcap:22 AS_max_des_send:49
a=bcap:23 AS_min_des_recv:34
a=bcap:24 AS_min_des_send:34
a=bcap:25 AS_max_sup_recv:74
a=bcap:26 AS_max_sup_send:74
a=bcap:27 AS_min_sup_recv:14
a=bcap:28 AS_min_sup_send:14
a=rtpmap:100 AMR-WB/16000/1
a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1
a=rmcap:2 AMR-WB/16000/1
a=mfcap:2 mode-change-capability=2; max-red=220; octet-align=1

a=ptime:20
a=maxptime:240

a=pcfg:1 m=1 a=-m b=1,11,12,13,14,15,16,17,18 pt=1:99
a=pcfg:2 m=2 a=-m b=2,21,22,23,24,25,26,27,28 pt=2:100
a=pcfg:3 m=3 a=-m b=3,31,32,33,34,35,36,37,38 pt=3:97
a=pcfg:4 m=4 a=-m b=4,41,42,43,44,45,46,47,48 pt=4:98
```

SDP answer

```
m=audio 49152 RTP/AVP 97
b=AS:50
b=AS_max_des_recv:37
b=AS_max_des_send:37
b=AS_min_des_recv:31
b=AS_min_des_send:31
b=AS_max_sup_recv:50
b=AS_max_sup_send:50
b=AS_min_sup_recv:12
b=AS_min_sup_send:12
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=ptime:20
a=maxptime:240

a=acfg:1 m=1 b=3,31,32,33,34,35,36,37,38 pt=1:97
```

The new bandwidth information is derived in the same way as done for solution B.

8.3.2.4 Modifying the bandwidth information in the SDP offer

Networks nodes in the path have the possibilities to modify the new bandwidth information in the SDP offer for each codec and configuration in the following manner:

- The first node can decrease the maximum supported, maximum desired and minimum desired bandwidth according to network policies. However, the first node does normally not increase the maximum supported, maximum desired and minimum desired bandwidth except when required to correct undesired or erroneous UE behavior, or for IPv4/IPv6 transport interworking.

- The first node can increase the minimum supported bandwidth according to network policies. However, the first node does normally not decrease the minimum supported bandwidth except when required to correct UE misbehavior, or for IPv4/IPv6 transport interworking.

- Subsequent nodes can decrease the maximum supported, maximum desired and minimum desired bandwidths and increase the minimum supported bandwidth, but are supposed to not increase the maximum supported, maximum desired and minimum desired bandwidths, nor to decrease the minimum supported bandwidth except when required due to IPv4/IPv6 transport interworking.

- Networks nodes may remove an unwanted codec or configuration together with all related bandwidth information.

- Networks nodes may add codecs or configurations accessible via transcoding together with all related bandwidth information.

- If a network node desires to use a MBR=GBR bearer, it preferably decreases maximum supported bandwidth down to the maximum desired bandwidth in the SDP offer.

- The following relationships are maintained by any network node when modifying bandwidths:

  Minimum Supported Bandwidth <= Minimum Desired Bandwidth

  Minimum Desired Bandwidth <= Maximum Desired Bandwidth

  Maximum Desired Bandwidth <= Maximum Supported Bandwidth

NOTE: These rules allow all both the originating and terminating operators in the call setup direction to implement certain policies, but avoid that subsequent operators in the call setup chain counteract the policies of the first operator, and guarantee that the used bandwidths remain in the supported range of the originating UE.
For instance, the following policies are supported:
An operator desiring to set a lower limit to the acceptable QoS can increase the Minimum Supported Bandwidth.
An operator desiring to limit the GBR for MBR>GBR bearers can decrease the Minimum Desired Bandwidth.
An operator desiring to limit the GBR for MBR=GBR bearers can decrease the Maximum Supported Bandwidth.

In the SDP answer, the first node does normally not modify the bandwidth values except when required to correct UE misbehavior, when replacing the selected codec and configuration in the SDP answer due to transcoding, or due to IPv4/IPv6 transport interworking. Subsequent node also are supposed to not modify the bandwidth values except when replacing the selected codec and configuration in the SDP answer due to transcoding, or due to IPv4/IPv6 transport interworking.

8.3.2.5 Resource reservation in different networks

Since the bandwidth information for both directions is included in the SDP answer, all networks in the path have the same information and can use this instead of proprietary codec-specific algorithms both for the admission control and for resource reservation. Thereby, it is possible to align QoS end-to-end.

It should be noted that this does not prevent using operator policies, even if the operator policies would use different bandwidths than indicated in the SDP offer. However, in this case, it would be beneficial to modify the SDP offer before forwarding it to the next network so that the bandwidth information in the SDP offer is aligned with the selected QoS parameters.

8.3.3 Compliance with proposed requirements

The described proposed solution solves the proposed requirements A to E. The proposed solution is compliant with the solution design requirements J to M. SDP attribute related requirements G to I are not applicable to this solution.

8.3.4 Impact on networks and terminals

The impacts on networks and terminals are the same as for solution B. In additions, networks and terminals need to implement SDPMiscCapNeg, which also mean that they need to implement SDPCapNeg [13] and SDPMediaCapNeg [15].

8.4 Potential solution D: New attribute for bandwidth information in SDP offer and answer for each RTP payload type

8.4.1 Introduction

This solution describes how the clients can make the networks aware of the negotiated maximum supported bandwidth, the minimum supported bandwidth, the maximum desired bandwidth and the minimum desired bandwidth for each direction by defining a new attribute to carry the new bandwidth information.

8.4.2 Description of the solution

8.4.2.1 General solution

The general solution is to add information in the SDPs about the maximum supported bandwidth, minimum supported bandwidth, maximum desired bandwidth and the minimum desired bandwidth, for sending and receiving directions, respectively. The clients would negotiate these bandwidths in the same way as they negotiate other configuration parameters. The networks would use these bandwidths in the session setup and session re-negotiations, both for the admission control and for resource reservation.

8.4.2.2 New attribute

The syntax for the new SDP attribute can be defined in several ways. One example is shown below:

```
a=bw:<pt-list> send=<maxs>,<maxd>,<mind>,<mins>; recv=<maxs>,<maxd>,<mind>,<mins>
```

where:

The attribute can be used either on media level or on session level.

`<pt-list>` identifies the RTP payload type(s) for which the current bandwidth declaration applies,

- A wild card ('*') can be used to make the bandwidth definition apply to all RTP payload types for the given media scope or for the entire session
- `pt-list` can be a comma-separated list of RTP payload type numbers, i.e. `a=bw:96,97,105 ...`
- `pt-list` can also include be a range RTP payload type numbers, i.e. `a=bw:96-99 ...`
- `pt-list` can even include a combination of individual RTP payload type number(s) and range(s), i.e. `a=bw:96-99,105,107-110 ...`

`send` or `recv` defines the direction for which the bandwidth declaration applies

`<maxs>,<maxd>,<mind>,<mins>` is the bandwidth declaration for the given direction, containing the maximum supported bandwidth, maximum desired bandwidth, minimum desired bandwidth and the minimum supported bandwidth.

This is probably the simplest possible syntax to support signalling the bandwidth information identified in this study.

A benefit with defining a new SDP attribute is that the syntax can be defined in whatever way needed (for instance, compare with the syntax in proposed solution F that could also be used here if a payload type is added). The syntax can also be defined to allow for future extensions, even though this is not shown in the definition above.

NOTE 1: Using 'bw' for the new bandwidth information can be confusing since the EVS codec uses the same name for the codec parameter that identifies the audio bandwidth. If this solution is selected then a different name should be used.

NOTE 2: This clause gives only a brief description for the proposed new SDP attribute. If this solution is selected then further details on the encoding need to be defined.

NOTE 3: An alternative encoding of the new attribute is presented in clause 8.6.2.2.

8.4.2.3 Procedures

The SDP offer can contain several RTP payload types corresponding to different codecs and configurations, where the different configurations have different bandwidth needs. The originating client includes bandwidth values required for each of the offered codec and configurations as bandwidth information, using a separate instance of the a=bw attribute for each RTP payload type.

The answerer selects a codec and configuration. In the SDP answer, the answerer provides bandwidth information for both the send and receive direction for the codec and configuration it selects. The answerer provides bandwidth information which is equal or below the corresponding received bandwidth information within the SDP offer for maximum values and for the minimum desired value, and which is equal or higher for the minimum supported value.

8.4.2.4 Session negotiation example

An example of how the new attribute can be used in the session negotiation is shown below. This example is based on Use case E where both AMR-WB and AMR are offered but where AMR is negotiated, see clause 6.6 and Table 6.6.1-1. A difference from Use case E is that the offer allows for using 100% redundancy even when the highest codec mode is used.

The SDP offer contains several RTP payload types corresponding to different codecs and configurations, where the different configurations have different bandwidth needs. With a new attribute it is possible to identify the bandwidth needs for each configuration.

The new attribute lines are highlighted with bold font.

NOTE: The numerical values in those examples have not been carefully checked and can contain errors.

Table 8.4.2.4-1: SDP offer-answer for proposed solution D with new bandwidth modifiers

SDP offer
m=audio 49152 RTP/AVP 100 99 98 97 b=AS:49 **a=bw:97 send=50,37,31,12 recv=50,37,31,12** a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 **a=bw:98 send=50,38,31,12 recv=50,38,31,12** a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 **a=bw:99 send=73,49,34,13 recv=73,49,34,13** a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 **a=bw:100 send=74,49,34,14 recv=74,49,34,14** a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:37 **a=bw:* send=50,37,31,12 recv=50,37,31,12** a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240

The new bandwidth information is derived in the same way as done for solution C.

A wildcard (*) is used in the SDP answer to indicate that the bandwidth information applies to the entire media. In this case, there is no need to use SDPMiscCapNeg to indicate different bandwidths for different payload types.

8.4.2.5 Modifying the bandwidth information in the SDP offer

Networks in the path have the same possibilities to modify the new bandwidth information as possible with solution C.

8.4.2.6 Resource reservation in different networks

Same as for solution C.

8.4.3 Compliance with proposed requirements

The described proposed solution solves the proposed requirements A to E. The proposed solution is compliant with the solution design requirements J to M and the SDP attribute related requirements G to I.

8.4.4 Impact on networks and terminals

Same as for solution B.

8.5 Potential solution E: New bandwidth modifiers only in SDP answer

8.5.1 Introduction

A variant of proposed solution B is to only include the new bandwidth modifiers in the SDP answer, since this shows what codec and configuration that has been negotiated, but not to include anything new in the SDP offer.

8.5.2 Description of the solution

8.5.2.1 General solution

This solution describes how the answering client can make the networks aware of the maximum supported bandwidth, the minimum supported bandwidth, the maximum desired bandwidth and the minimum desired bandwidth for the negotiated codec and configuration for each direction by defining new bandwidth modifiers to carry the new bandwidth information. The new bandwidth modifiers are only included in the SDP answer.

The reason for including the new bandwidth information only in the SDP answer is that it shows which codec and configuration that has been negotiated.

8.5.2.2 New bandwidth modifiers

Same as for solution B, see clause 8.2.2.2, except that it is only allowed to use the new bandwidth modifiers in the SDP answer.

8.5.2.3 Session negotiation example

The conditions for this example are the same as used for solution B

NOTE: The numerical values in those examples have not been carefully checked and can contain errors.

Table 8.5.2.3-1: SDP offer-answer for proposed solution E with new bandwidth modifiers only in the SDP answer

SDP offer
m=audio 49152 RTP/AVP 100 99 98 97 b=AS:49 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240

SDP answer
m=audio 49152 RTP/AVP 97 b=AS:37 **b=AS_max_des_recv:37** **b=AS_max_des_send:37** **b=AS_min_des_recv:31** **b=AS_min_des_send:31** **b=AS_max_sup_recv:50** **b=AS_max_sup_send:50** **b=AS_min_sup_recv:12** **b=AS_min_sup_send:12** a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240

The new bandwidth information is derived in the same way as done for solution B but only for the codec and configuration included in the SDP answer.

8.5.2.4 Modifying the bandwidth information in the SDP offer

Since the SDP offer does not include the new bandwidth information it becomes impossible for the networks to modify this information.

8.5.2.5 Resource reservation in different networks

Since the bandwidth information for both directions is included in the SDP answer, all networks in the path have the same information and can use this instead of proprietary codec-specific algorithms both for the admission control and for resource reservation. Thereby, it is possible to align QoS end-to-end.

8.5.3 Compliance with proposed requirements

This solution fulfils the proposed requirements C to E. However, it does not fulfil the proposed requirements A and B. The proposed solution is compliant with the solution design requirements J to M. SDP attribute related requirements G to I are not applicable to this solution.

8.5.4 Impact on networks and terminals

The implementation impacts on networks and terminals are the virtually same as for solution B.

8.6 Potential solution F: New SDP attribute in SDP offer and answer for entire media line

8.6.1 Introduction

A variant of proposed solution D is to only include the new bandwidth modifiers SDP attribute not for each RTP payload type but the entire media line.

8.6.2 Description of the solution

8.6.2.1 General solution

This solution describes how the offering client proposes bandwidth requirements for the entire media line. The network can modify the bandwidth information in the SDP offer according to its policies for the media type. The answering client takes this information into account in the codec selection and makes the networks and offerer aware of the maximum supported bandwidth, the minimum supported bandwidth, the maximum desired bandwidth and the minimum supported bandwidth for the negotiated codec and configuration for each direction. A new SDP attribute is defined to carry the new bandwidth information.

8.6.2.2 New SDP attribute

```
a=bwinfo:<type> <value>*(,<type> <value>)

<type>=mss/mds/iss/ids/msr/mdr/isr/idr/<token>
```

where:

<type> can be one of mss/mds/iss/ids/msr/mdr/isr/idr and identifies the type of the bandwidth. The defined types are maximum supported bandwidth in send direction ("mss"), maximum desired bandwidth in send direction ("mds"), minimum supported bandwidth in send direction ("iss"), minimum desired bandwidth in send direction ("ids"), maximum supported bandwidth in receive direction ("msr"), maximum desired bandwidth in receive direction ("mdr"), minimum supported bandwidth in receive direction ("isr"), minimum desired bandwidth in receive direction ("idr"). More types may be defined in the future and unknown types are ignored.

<value> is an integer denoting the applicable bandwidth value for a bandwidth type in kilobytes/sec.

8.6.2.3 Procedures

The SDP offer can contain several RTP payload types corresponding to different codecs and configurations, where the different configurations have different bandwidth needs. The originating client includes the maximum values required for any of the offered codec and configurations as bandwidth information.

The answerer selects a codec and configuration which complies with the received bandwidth information within the SDP offer (i.e. the required bandwidth is equal or below the received information for each type of bandwidth information). In the SDP answer, the answerer provides bandwidth information for both the send and receive direction for the media line that matches the needs of the codec and configuration it selects. The answerer provides bandwidth information which is equal or below the corresponding received bandwidth information within the SDP offer for maximum values and for the minimum desired value, and which is equal or higher for the minimum supported value.

8.6.2.4 Session negotiation example

The conditions for this example are the same as used for solution B.

NOTE: The numerical values in those examples have not been carefully checked and can contain errors.

Table 8.6.2.4-1: SDP offer-answer for proposed solution F with new SDP attribute only in the SDP answer

SDP offer
c= m=audio 49152 RTP/AVP 100 99 98 97 b=AS:74 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240 **a=bwinfo:mss 73,mds 49,iss 13,ids 34,msr 73,mdr 49,isr 13,idr 34**
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:52 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240 **a=bwinfo:mss 52,mds 31,iss 12,ids 31,msr 52,mdr 31,isr 12,idr 31**

8.6.2.5 Modifying the bandwidth information in the SDP offer

Networks in the path have the same possibilities to modify the new bandwidth information as possible with solution B.

8.6.2.6 Resource reservation in different networks

Since the bandwidth information for both directions is included in the SDP answer, all networks in the path have the same information and can use this instead of proprietary codec-specific algorithms for resource reservation. Thereby, it is possible to align QoS end-to-end.

8.6.3 Compliance with proposed requirements

The described proposed solution addresses the proposed requirements A to E. However, those requirements are only partially met, as the bandwidth information is provided per media component, rather than per codec and configuration. The proposed solution is compliant with the solution design requirements J to M and the SDP attribute related requirements G to I.

8.6.4 Impact on networks and terminals

Adding new information in SDP means that terminals and networks would need to support the new SDP parameters in order to make the solution useful.

For the networks, the AF (P-CSCF) would need to extract the new information from the SDP answers and provide corresponding session information to the PCRF. Compared to the existing procedures, the AF or PCRF would no longer have to use information in the SDP offer to derive bandwidth for resource reservation purposes, but would have to extract more bandwidth information from the SDP answer. The PCRF would no longer need to use a codec-specific algorithm. The related detailed impacts on the Rx interface and AF and PCRF procedures will be determined by CT3 during their normative work. The PCRF would then use the session information to set the QoS parameters on the Gx interface. The QoS parameters that are used on the Gx interface are the same as in the existing specifications. The only difference is the values that the PCEF would use. This means that there is no need to change the PCEF, the RAN or the interfaces to these nodes.

Nodes in the network applying existing policies to modify SDP offer for transcoding purposes would need to be updated to adjust the new bandwidth information accordingly.

Transcoding nodes within the network would need to be updated to take into account the new bandwidth information when encoding media. The new bandwidth information would need to be provided to the IMS-AGW, TrGW and MRFP via the Iq interface, Ix interface and Mp interfaces, respectively.

Adding new SDP attributes also gives automatic fallback to the legacy solution whenever the new SDP attributes are not supported.

Compared to solution D, the SDP processing load is smaller, as the new attribute appears only one time for each media line (and not potentially multiple times in the SDP offer).

8.7 Potential solution G: New SDP attribute only in SDP answer

8.7.1 Introduction

A variant of proposed solution F is to only include the new bandwidth modifiers in the SDP answer, since this shows what codec and configuration that has been negotiated, but not to include anything new in the SDP offer.

8.7.2 Description of the solution

8.7.2.1 General solution

This solution describes how the answering client can make the networks and the offerer aware of the maximum supported bandwidth, the minimum supported bandwidth, the maximum desired bandwidth and the minimum supported bandwidth for the negotiated codec and configuration for each direction by defining a new SDP attribute to carry the new bandwidth information. The new SDP attribute is only included in the SDP answer.

The reason for including the new bandwidth information only in the SDP answer is that it shows which codec and configuration that has been negotiated.

8.7.2.2 New SDP attribute

See Subclause 8.6.7.2.2.

8.7.2.3 Procedures

The SDP offer can contain several RTP payload types corresponding to different codecs and configurations, where the different configurations have different bandwidth needs.

The answerer selects a codec and configuration. In the SDP answer, the answerer provides bandwidth information for both the send and receive direction for the media line. The answerer provides bandwidth information which matches the needs of the codec and configuration it selects.

8.7.2.4 Session negotiation example

The conditions for this example are the same as used for solution B.

NOTE: The numerical values in those examples have not been carefully checked and can contain errors.

Table 8.7.2.4-1: SDP offer-answer for proposed solution G with new SDP attribute only in the SDP answer

SDP offer
c= m=audio 49152 RTP/AVP 100 99 98 97 b=AS:74 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR-WB/16000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR-WB/16000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 b=AS:52 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=ptime:20 a=maxptime:240 **a=bwinfo:mss 52,mds 31,iss 12,ids 31,msr 52,mdr 31,isr 12,idr 31**

8.7.2.5 Modifying the bandwidth information in the SDP offer

Since the SDP offer does not include the new bandwidth information it becomes impossible for the networks to modify this. The network can influence the required bandwidth by modifying the offered payload types or by reducing the existing b=AS bandwidth modifier.

8.7.2.6 Resource reservation in different networks

Since the bandwidth information for both directions is included in the SDP answer, all networks in the path have the same information and can use this instead of proprietary codec-specific algorithms for the resource reservation. Thereby, it is possible to align QoS end-to-end.

8.7.3 Compliance with proposed requirements

This solution fulfils the proposed requirements C to E on making the networks aware of the minimum/maximum supported/desired bandwidths for the negotiated media and codecs. However, it does not fulfil the proposed requirements A and B on allowing the bandwidth properties to be a part of the negotiation process. The proposed solution is compliant with the solution design requirements J to M and the SDP attribute related requirements G to I.

8.7.4 Impact on networks and terminals

Adding new information in SDP means that terminals and networks would need to support the new SDP parameters in order to make the solution useful.

For the networks, the AF (P-CSCF) or PCRF would need to extract the new information from the SDP answers and provide corresponding session information to the PCRF. Compared to the existing procedures, the AF or PCRF would no longer have to use information in the SDP offer to derive bandwidth for resource reservation purposes, but would have to extract more bandwidth information from the SDP answer. The PCRF would no longer need to use a codec-specific algorithm. The related detailed impacts on the Rx interface and AF and PCRF procedures will be determined by CT3 during their normative work. The PCRF would then use the session information to set the QoS parameters on the Gx interface. The QoS parameters that are used on the Gx interface are the same as in the existing specifications. The only difference is the values that the PCEF would use. This means that there is no need to change the PCEF, the RAN or the interfaces to these nodes.

Nodes in the network applying existing policies to modify SDP offer for transcoding purposes would need to be updated to adjust the new bandwidth information accordingly.

Transcoding nodes within the network would need to be updated to take into account the new bandwidth information when encoding media. The new bandwidth information would need to be provided to the IMS-AGW, TrGW and MRFP via the Iq interface, Ix interface and Mp interfaces, respectively.

Adding new SDP attributes also gives automatic fallback to the legacy solution whenever the new SDP attributes are not supported.

Compared to solution D, the SDP processing load is smaller, as the new attribute appears only one time for each media line in the SDP answer (and not potentially multiple times in the SDP offer). Existing policies in the network to modify SDP offer continue to be supported.

8.8 Potential solution H: Bitrate variations

8.8.1 Introduction

This solution defines that an averaging window should be used when calculating the used bitrate. The length (in time) of the averaging window is then selected such that even large bitrate variations are smoothed sufficiently to avoid risking packet losses.

8.8.2 Description of the solution

The procedure for how the used bitrate should be calculated is defined in 3GPP specifications. The procedure uses an averaging window over a specified time period 'T' such that an average over the given time period is calculated. This gives a smoothing effect such that clients that need to send one or more large packets, for example for a large I frame, have time to compensate for this by sending smaller packets afterwards.

The solution defines the time period that is used in somewhat different ways depending on whether the entity is generating media or whether the entity is monitoring the media:

- Entities generating media, e.g. codecs, should generate packets such that the average bitrate measured over a time period 'Te' that is shorter than or equal to T.
- Entities monitoring the media, e.g. policing functions, should calculate the average bitrate over a time period that 'Tp' that is longer than or equal to T.

The length of the averaging window 'T' is proposed to be 2 seconds for MBR calculation (see Annex A).

NOTE: It will be determined during the normative phase of the work whether the recommended 2sec averaging window also applies to the GBR.

This solution does not use any signalling between clients and networks, or between different networks nodes. This means that the implementation is local in the respective node. This also means that the actual implementation could be different and does not use an averaging window as long as the performance is equivalent to what is defined above. For example, a client generating media could use a packet pacing function to avoid sending several large packets too closely to each other, which would create a high peak bitrate and would risk triggering the policing function.

The time period T can be made dependent on the QCI and thus media specific or service specific by defining it in media or service specifications, for example in TS 26.114. It is FFS whether a time period T dependant on the QCI offers significant advantages over a fixed time period T that is sufficiently long for all services. It is also FFS to determine whether a separate value of T should be defined for source-controlled Variable-Bit-Rate (VBR) operation vs. non-VBR codec modes, e.g. for EVS using VBR operation.

If a generic definition is desired, which is then used for all media and all services, then it may be better to define the time period in PCC or EPC specifications.

8.8.3 Compliance with proposed requirements

This solution fulfils the proposed requirement F on bitrate variations.

8.8.4 Impact on networks and terminals

It is well known that encoders, especially for video, generate media with large bitrate variations already today. This means that both networks and terminals should already use some form of averaging when calculating the bitrate. Policing functions need to do this to avoid dropping packets unnecessarily. Terminals also need this to reduce the variations to avoid triggering packet dropping in policing functions.

When introducing this solution it should therefore be relatively easy to implement it also in networks and terminals. This solution also means no changes to the architecture or the interfaces, which further simplifies the implementation.

9 Conclusion and recommendations

This study has identified a number of use cases where the media handling is currently suboptimal since the resources are independently allocated in different networks. This can lead to either over-allocation, under-allocation or even both. These use cases would benefit from end-to-end alignment of the QoS settings.

A gap analysis has been performed for the use cases to derive a set of proposed requirements. Several proposed solutions have been analysed with respect to their compliance with the requirements, and also with respect to impact on networks and terminals.

The following areas are found to have weaknesses:

- It is not described in TS 26.114 for which conditions session re-negotiation can be performed to align codec and codec configurations end-to-end.
- Mechanisms for aligning bandwidth properties end-to-end are undefined to a large degree.
- Definitions are lacking for bandwidth variations.

The recommendations of the present document are:

- To perform work addressing the above areas.
  - Solution A on session re-negotiation should be addressed in the specification phase, where it can be outlined in more detail.
  - Solution C and D are the only ones fulfilling all requirements A-E on new bandwidth information. Solution D has lower network impact and is suggested to be the basis for the future work.
  - Solution H on bitrate variations should also be pursued in the normative work.
- To coordinate this work with other 3GPP groups and IETF, in particular with SA2, RAN2, CT3 and CT4.

Annex A:
Determining suitable averaging window length

A.1 Introduction

This annex describes one method for how a suitable averaging window length can be determined for variable bitrate media. The method included in clause A.2 uses video encoding for real-time conversational video as an example but it is easy to extend this method to other media and other services. A discussion on how the method can be modified for other media is included in clause A.3.

A.2 Suitable averaging window for video

A.2.1 Reasonable video encoder configuration

The video codec used in this analysis is configured as described in the table below.

Table A.2.1-1: Parameters used in this analysis

Property	Value(s)	Comments
Video codec	H.264	The video codec that is use is actually unimportant for this analysis
Bit rate	384 kbps	The average bitrate that is actually unimportant for this analysis
Frame rate	10, 15, 20, 30 Hz	
Average frame data size	2400 bytes/frame	The average frame data size is actually unimportant for this analysis
Frame data sizes, I frames	5x, 10x average frame data size	
Packetization	2 RTP packets per video frame on average Up to 10-20 RTP packets for I frames	This analysis assumes that all RTP packets are transmitted at once, which makes the packetization unimportant for this analysis
IPv6/UDP/RTP overhead	40+8+12 = 60 bytes	The IP overhead is actually unimportant for this analysis

A.2.2 Normal transmission

The normal transmission behaviour means that there is no need to send any large I frames. The actual frame data size will vary but the expectation value of the frame data size is still constant and depends only on the bitrate (encoding bitrate plus IP/UDP/RTP overhead) and the frame rate:

$$E[\,frameDataSize] = \frac{(encodingBitrate + overhead)}{frameRate} = X \qquad \text{(A.2.2-1)}$$

This is illustrated in the table below.

Table A.2.2-1: Description of expected normal sending behaviour

Frame number	-m	…	-3	-2	-1	0	+1	+2	+3	…	+m
E[fds]	X	…	X	X	X	X	X	X	X	…	X
Average	X		X	X	X	X	X	X	X		X

This is independent of the window length and the position.

A.2.3 Sending a large I frame

If a large I frame (5x the average frame data size) is transmitted and if the surrounding frames are not modified then the average depends on the length of the window and also the position. In the table below it is assumed that the length of the window is 5 frames and the average frame data size at position 'p' calculated over the last 5 frames.

Table A.2.3-1: Average frame data size with different positioning

Frame number	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6
E[fds]	X	X	X	X	X	X	5X	X	X	X	X	X	X
Avg@-6	X												
Avg@-5		X											
Avg@-4			X										
Avg@-3				X									
Avg@-2					X								
Avg@-1						X							
Avg@0							1.8X						
Avg@+1								1.8X					
Avg@+2									1.8X				
Avg@+3										1.8X			
Avg@+4											1.8X		
Avg@+5												X	
Avg@+6													X

It is here assumed that the policing allows for X. This means that for the Nth frame and for the 4 subsequent frames the average will exceed the MBR and the policing may drop some packets. This is highlighted in the table above.

A.2.4 Compensating for a large I frame

To compensate for the large I frame, the frame before and after would need to be encoded at a lower bitrate. If the averaging window is 5 frames and the size of the I frame is 5X then 4 frames before and after would need to be encoded at 0 bits. This is shown in the table below.

Table A.2.4-1: Compensating for a large I frame

Frame number	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6
E[fds]	X	X	0	0	0	0	5X	0	0	0	0	X	X
Avg@-6	X												
Avg@-5		X											
Avg@-4			0.8X										
Avg@-3				0.6X									
Avg@-2					0.4X								
Avg@-1						0.2X							
Avg@0							X						
Avg@+1								X					
Avg@+2									X				
Avg@+3										X			
Avg@+4											X		
Avg@+5												0.2X	
Avg@+6													0.4X

This now fulfils the MBR requirement and the policing should not need to drop any packets.

An obvious drawback of this is that 8 out of 9 frames are not transmitted so the video is frozen for a quite many frames, both before and after the I frame. How long freezing time one will get before and after the I frame depends on the frame rate:

- 10 fps → 4*1/10 s = 0.4 s
- 15 fps → 4*1/15 s = 0.267 s
- 20 fps → 4*1/20 s = 0.2 s
- 30 fps → 4*1/30 s = 0.133 s

The total impacted period is 9 frames (freezing before the I frame + the I frame + freezing after the I frame):

- 10 fps → 9*1/10 s = 0.9 s
- 15 fps → 9*1/15 s = 0.6 s
- 20 fps → 9*1/20 s = 0.45 s
- 30 fps → 9*1/30 s = 0.3 s

A.2.5 Sending an even larger I frame

Sending a 10x I frame while keeping the window length at 5 frames means that it becomes impossible to send all the date at once. To avoid exceeding the MBR one have to:

- skip the transmission of the 4 frames before the I frame
- send half of the data for the I frame (similar to what is shown above)
- skip the transmission of the next 4 frames
- send the remaining half of the data for the I frame
- skip the transmission of the 4 frames subsequent

It should be noted that the I frame cannot be completely decoded until the second half of the data has arrived, so the freezing time before the presentation of the I frame will be 9 frame periods. This corresponds to 0.3-0.9 seconds, depending on the frame rate.

This extra delay to send the second half of the data for the I frame is in the present document called "packet pacing" but other terminologies can also be relevant. This packet pacing would come automatically if a token bucket were used for the transmission. The drawback with the packet pacing is that it increases the end-to-end delay. If one want to maintain a smooth and regular play-out rate then this extra delay would impact all frames.

A.2.6 Using a longer averaging window

Extending the averaging window to 10 frames when sending a 5x frames allows for sending the frames immediately before and after the I frame. However, those frames need to be encoded at a lower bitrate. One example is shown in the table below where the surrounding frames are encoded with 50% of the average bitrate.

Table A2.6-1: Compensating for a large I frame

Frame number	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	+1
E[fds]	1	1	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	5	0.5
Avg	1	1	1	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6	1	0.95
Frame number	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14
E[fds]	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	1	1	1	1	1
Avg	0.95	0.95	0.95	0.95	0.95	0.95	0.95	1	0.6	0.65	0.7	0.75	0.8
Frame number	+15	+16	+17	+18	+19	+20	+21						
E[fds]	1	1	1	1	1	1	1						
Avg	0.85	0.9	0.95	1	1	1	1						

The above description can be easily expanded to other variants.

A.2.7 Selection of averaging window length

From the discussion above it can be derived that the window length, expressed in number of frames, depends on the following parameters:

- the relationship between the average bitrate and the recovery bitrate used before and after the I frame
- the relationship between the maximum frame data size that one want to handle in relation to the average frame data size

This can be expressed with the following formula:

$$avgWinLength = \frac{avgBitrate}{avgBitrate - recoveryBitrate} * \frac{maxFrameDataSize}{avgFrameDataSize} \quad \text{(A.2.7-1)}$$

A few examples are shown in the table below:

Table A.2.7-1: Examples of averaging window lengths

avgBitrate	recoveryBitrate	maxFrameDataSize	avgFrameDataSize	avgWinLength
B	B/2	5X	X	10
B	B/3	5X	X	7.5
B	B/4	5X	X	3.75
B	0	5X	X	5
B	B/2	10X	X	20
B	B/3	10X	X	15
B	B/4	10X	X	7.5
B	0	10X	X	10

Assuming that MTSI should allow for I frames up to 10 times the size of the average frame and assuming that MTSI should allow for using 50% of the average bitrate during the recovery period then the length of the averaging window becomes 20 frames.

Assuming further that frame rate should normally be between 10 and 30, the worst case (10 fps) gives T = 2 seconds.

A.3 Other services

A.3.1 Conversational speech using EVS VBR

In most cases (fixed-rate codecs, AMR, AMR-WB, EVS with CBR), the bearer setup is set to the highest codec mode, or higher if application layer redundancy is needed. Using an averaging window of 2 seconds should work well.

EVS however also includes the 5.9 kbps VBR mode. The 5.9 kbps average rate is only applicable to speech. If the signal is non-speech it can happen that the higher bitrate modes (7.2 or 8.0 kbps) are used for longer periods of time.

It is not obvious how to handle this with an averaging window in the policing function because at session setup, when the codec mode is decided, one do not know what type of signal that will be injected into the encoder.

There is however a quite straight-forward solution, which is to set the MBR (or GBR for MBR>GBR bearers) to 8.0 kbps or higher whenever the 5.9VBR is negotiated for the session, even if 5.9VBR is the only allowed mode.

A.3.2 Streaming and other non-conversational services

Streaming services and non-conversational services allow for longer end-to-end delay than conversational services. This gives more room for using packet pacing in the client to give a smoother bitrate.

Other services may be more similar to file transfer and may show very large bitrate variations than what is discussed above for video. These may therefore need a longer averaging window.

Also, if a vendor wants to use the same averaging window for all services then he should select the maximum one. If MTSI clients assume that the averaging window is 2 seconds then this should still work well. This is why the following statements are important:

- Policing functions in the network should use an averaging window that is at least 2 seconds.
- UEs should assume that the averaging window in policing functions is no longer than 2 seconds.

With this, the different service specifications could define different window lengths. It can then be left for the implementers to decide if they want to use one single averaging window length or if they want to have different averaging window lengths for different QCIs.

Annex B:
Change history

Change history							
Date	**TSG SA#**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2015-03	67	SP-150096			Presented to TSG SA #67 for information		1.0.0
2015-09	69	SP-150450			Presented to TSG SA #69 for approval	1.0.0	2.0.0
2015-09	69				Version 13.0.0	2.0.0	13.0.0
2016-12					LTE logo updated	13.0.0	13.0.1

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2017-03	75					Version for Release 14	14.0.0
2018-06	80					Version for Release 15	15.0.0
2020-07	88-e					Version for Release 16	16.1.0
2022-04	-	-	-	-	-	Update to Rel-17 version (MCC)	**17.0.0**
2024-03	-	-	-	-	-	Update to Rel-18 version (MCC)	**18.0.0**

History

Document history		
V18.0.0	May 2024	Publication