

ETSI TR 136 933 V14.0.0 (2017-04)



LTE;
Study on Context Aware Service Delivery in RAN for LTE
(3GPP TR 36.933 version 14.0.0 Release 14)



Reference

RTR/TSGR-0336933ve00

Keywords

LTE

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 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:

<http://www.etsi.org/standards-search>

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status.

Information on the current status of this and other ETSI documents is available at

<https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx>

If you find errors in the present document, please send your comment to one of the following services:

<https://portal.etsi.org/People/CommiteeSupportStaff.aspx>

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.

The content of the PDF version shall not be modified without the written authorization of ETSI.

The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2017.

All rights reserved.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are Trade Marks of ETSI registered for the benefit of its Members.

3GPP™ and **LTE™** are Trade Marks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.

oneM2M logo is protected for the benefit of its Members

GSM® and the GSM logo are Trade Marks registered and owned by the GSM Association.

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is 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 IPR Policy, no investigation, 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.

Foreword

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, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under <http://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

Intellectual Property Rights	2
Foreword.....	2
Modal verbs terminology.....	2
Foreword.....	4
Introduction	4
1 Scope	5
2 References	5
3 Definitions, symbols and abbreviations	5
3.1 Definitions	5
3.2 Abbreviations	6
4 Definitions and problems description.....	6
4.1 Issue 1: Backhaul long latency	6
4.2 Issue 2: TCP E2E delay with throughput decreasing	6
4.3 Issue 3: Video transmission issue cases	6
4.4 Issue 4: UL Video transmission critical data discard	7
5 Solutions.....	7
5.1 General Principles	7
5.2 Solutions for issue 1	7
5.2.1 Solution 1: UE assisted local cache solution	7
5.2.2 Conclusion for Issue 1	9
5.3 Solutions for issue 2	9
5.3.1 TCP Performance Enhancing Proxy (TCP PEP)	9
5.3.2 Network based Radio-Aware TCP.....	10
5.3.3 UE Based Radio-Aware TCP Solution	10
5.3.4 eNB generates TCP ACK on behalf of the UE.....	10
5.3.5 Conclusion for Issue 2	11
5.4 Solutions for issue 3	11
5.4.1 Solutions for case 1.....	11
5.4.1.1 Solution 1: UE request a higher QoS profile.....	11
5.4.1.2 Solution 2: Video Context Aware Solution.....	11
5.4.2 Solutions for case 2.....	12
5.4.2.1 Solution 1: PCC-based Solution.....	12
5.4.2.2 Solution 2: Solution of DASH optimisation.....	13
5.4.2.3 Solution 3: Solution of DASH optimisation via local breakout	14
5.4.3 Solutions for case 3.....	15
5.4.3.1 Solution 1	15
5.4.4 Conclusion for Issue 3	16
5.5 Solutions for issue 4	16
5.5.1 Solution 1: Dynamic adaptation of LTE radio resource allocation during the call.....	16
5.5.2 Solution 2: Static acceptance or rejection of adaptation of LTE radio resource allocation for the call	16
Annex A (informative): Video playout buffer aware scheduling	17
Annex B (informative): Change history	18
History	19

Foreword

This Technical Specification 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

Voice was the major service and 3GPP defined voice specific cross-layer optimizations for 3G and 4G. Web based streaming video and Apps are expected to take majority of 4G/5G bandwidth, and are key source to further increase the ARPU for operators. MPEG and 3GPP SA4 have been considering cross-layer optimizations by RAN and App mutual aware, e.g. network assisted DASH, video-aware scheduling. It is worthwhile and right time to study whether deep cross-layer optimizations for Web based streaming video and Apps can deliver the desired benefits. It is also worthwhile to take into account the related work in other SDOs/Working groups, e.g. ETSI MEC work on mobile video delivery optimization and local content caching. It is also beneficial to study and analyse the potential impact to architecture, protocol, and signalling to support RAN based local cached delivery, local breakout; and support RAN optimizations based on context awareness.

1 Scope

The present document provides descriptions and possible solutions of use cases for the Context Aware Service Delivery in RAN for LTE, and also provides analysis of these solutions. Considerations with regards to requested functionality in scope of other 3GPP groups, if any, may be captured in this document as well.

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] IETF RFC3135, Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations
- [3] X. Xu, Y. Jiang, T. Flach, E. Katz-Bassett, D. Choffnes, and R. Govindan, "Investigating performance enhancing proxies in cellular networks," in AIMS: 6th Workshop on Active Internet Measurements, Mar. 2014.
- [4] M. Necker, M. Scharf, and A. Weber, "Performance of TCP and HTTP Proxies in UMTS Networks," in Wireless Conference 2005 - Next Generation Wireless and Mobile Communications and Services (European Wireless), 11th European, April 2005, pp. 1–7.
- [5] M. Meyer, J. Sachs, and M. Holzke, "Performance evaluation of a TCP proxy in WCDMA networks," Wireless Communications, IEEE, vol. 10, no. 5, pp. 70–79, Oct 2003.
- [6] F. Ren, X. Huang, F. Liu, and C. Lin, "Improving TCP Throughput over HSDPA Networks," Wireless Communications, IEEE Transactions on, vol. 7, no. 6, pp. 1993–1998, June 2008.
- [7] V. Farkas, B. Héder, and S. Nováczki, "A Split Connection TCP Proxy in LTE Networks," in Information and Communication Technologies. Springer Berlin Heidelberg, 2012, vol. 7479, pp. 263–274.
- [8] S4-170238, Revised Work Item on "Server and Network Assisted DASH (SAND) for 3GPP Multimedia Services"
- [9] 3GPP TR 26.957: "Study on Server And Network-assisted Dynamic Adaptive Streaming over HTTP (DASH) (SAND) for 3GPP multimedia services"

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in 3GPP 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 3GPP TR 21.905 [1].

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP 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 3GPP TR 21.905 [1].

4 Definitions and problems description

4.1 Issue 1: Backhaul long latency

The issue may arise in cases where the distance between the RAN and the node hosting the application content is long or the number of routers on this route is high. In these cases long transportation latency may be experienced. Consequently certain kinds of service may be impacted significantly due to the long latency. For example, backhaul delay increases the TCP RTT, therefore if TCP is configured in a way that it cannot cope with such delays, TCP throughput can be affected.

4.2 Issue 2: TCP E2E delay with throughput decreasing

The behavior of TCP assumes that network congestion is the primary cause for packet loss and high delay. In cellular networks the bandwidth available for each UE can vary by an order of magnitude on a TTI basis due to changes in the underlying radio channel conditions. Such changes can be caused by the movement of devices or interference, as well as changes in system load due to bursty traffic sources or when other UEs enter and leave the network. TCP has difficulties adapting to these rapidly varying conditions.

If the E2E delay increases, the TCP RTT increases and the TCP throughput may decrease, which may impact the user experience.

4.3 Issue 3: Video transmission issue cases

The Operator video is a video service under the LTE operator's control. It is usually transmitted over a dedicated EPS bearer or using a dedicated QCI. However, there might be cases where an operator decides not to apply any dedicated QoS to a video service. These video services are named "Over-The-Top" (OTT) video and are video services that LTE operators have no control on. Such service traffic is usually treated in the same way as normal internet traffic, e.g. transmitted via default bearer, which may lead to poor QoE. Nevertheless, the QoS framework allows assignment of dedicated QCI for video.

Dedicated bearer and QCI is helpful in lessening the video issues. Below are some of the issues that may occur when operators decide to neither use dedicated bearers nor dedicated QCI for OTT video services.

Case 1: Empty buffer issue

The user is watching a streaming video. When the UE requests for some not yet buffered video segments e.g. by dragging a play scroll bar or when playout buffer is exhausted due to link throughput fluctuation and if the scheduling priority of the video content is not set accordingly, the video playing would probably stall depends on some condition, e.g. eNB's load and UE's QoS profile.

Case 2: Inaccurate throughput prediction for DASH issue

DASH client requests video quality based on downlink throughput prediction. Throughput prediction is based on implementation specific mechanisms. The accuracy of the prediction is dependent from the specific implementation. However, one factor that may affect the prediction is that the DASH client may base the prediction on previous downloads. The DASH client may not have an insight to whether the network conditions have changed, thus the current available throughput may be difficult to predict. Conservative requesting low data rate video segment leads to low video quality and aggressive requesting high data rate video segment leads to more video stalling.

Case 3: Long video delay issue

In HTTP based streaming, client first buffers some content, i.e. initial buffering, before playout in order to absorb the throughput and delay fluctuation. Assuming that scheduling priority is not appropriately set, a large buffer may cause long delay, thus lead to bad user experience.

4.4 Issue 4: UL Video transmission critical data discard

In conversational video (real-time streaming) the problem of PDCP discard of critical data in UL may occur. Critical data include I-frames of an H.264 video sequence and RTCP feedbacks for lost RTP packets. Both types of data are carried on the same bearer (dedicated, GBR, or non-GBR) and may be encapsulated in the same PDCP SDUs. Hence if the video bearer queue is highly loaded (e.g. in case of UL congestion), both types of data may be discarded due to expiry of the PDCP discard timer. Currently in AS, there are no means to prioritize I-frame data and RTCP feedback packets over P-frame data because they are carried on the same bearer. If these critical data are lost because of internal PDCP Discard on the sender device, the video stream may stop on the receiver side until these critical data are successfully retransmitted or until a new I-frame is transmitted to allow resynchronizing the video codecs and restore the video prediction chain. If forward error correction information is added to the H.264 payload the video may not be subject to interruptions and it may result in lower video bitrate.

NOTE1: This issue case does not require any specific video codec awareness impact in RAN3.

5 Solutions

5.1 General Principles

Impacts to the following LTE/EPC functions and principles of operation should preferably be avoided in order to ensure that the study can proceed within the RAN domain. Proposed solutions shall not be excluded solely on this basis, but sufficient benefits must be demonstrated in order to justify such impacts.

- Security: current LTE security mechanism shall not be jeopardized, e.g. User identity and device confidentiality, etc (TS 36.300, TS 33.401, TS 23.401).- Policy control: The Policy and Charging Enforcement Function shall remain in the PGW (TS 36.300, TS 23.203, TS 23.401).
- Charging: it shall remain unchanged that the SGW/PGW is responsible for interfacing with the charging system, and CDR generation (TS 36.300, TS 23.401).
- QoS: it shall follow current QoS control mechanism (TS 36.300, TS 23.203, TS 23.401).
- Mobility anchor: SGW is the local Mobility Anchor point for inter-eNodeB handover (TS 36.300, TS 23.401).
- UE IP address allocation: The UE's IP address is assigned by PGW (or L-GW for local breakout) during the default bearer activation, or after default bearer activation. The UE IP address allocation mechanism shall not be changed. (TS 36.300, TS 23.401)
- Lawful Interception: The lawful interception mechanism shall not be changed (TS 36.300, TS 23.401).

5.2 Solutions for issue 1

5.2.1 Solution 1: UE assisted local cache solution

Local cache is a solution to solve the long backhaul latency issue.

An example of how local caching could be achieved is described as follows.

Step 1: UE initiates a content request, including indicator, e.g., based on service information list

Step 2: Network needs to check whether the UE is legal for the selective acceleration (cache) request from core network.

During handover procedure, the eligibility check result may be transferred to the target eNB to avoid second eligibility check. Meanwhile, the indicator in step 1 can also be transferred to the target eNB to assist the selective acceleration.

Step 3: If the UE is legal for the selective acceleration (cache) request, Network uses local path otherwise uses traditional path

It needs to be noted that the exchange of information to RAN regarding the eligibility of UEs for local cache services may be achieved in different alternative ways, e.g. at UE context setup. This solution can be applied to all the potential deployment options regardless of where the cache is deployed. The possible local caching solutions are listed as below:

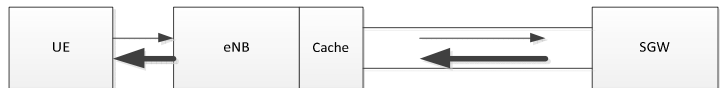
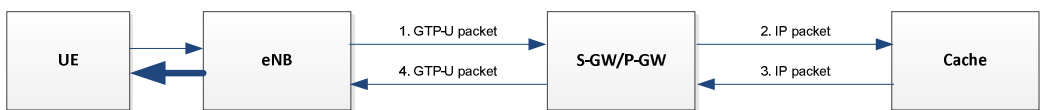
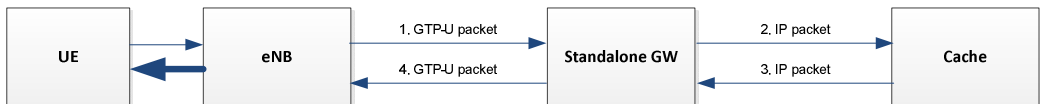


Figure 5.2.1-1 cache server collocated in the eNB

Note: This solution may not require any changes to the standard, and can be considered as the eNB’s implementation issue.



Standalone Cache Server and SIPTO above RAN



Standalone Cache Server and SIPTO@LN with standalone GW



Standalone Cache Server and SIPTO@LN with collocated L-GW

Figure 5.2.1-2 cache server after P-GW

Note: This solution may not require any changes to the standard, and can be considered as operator’s deployment option.

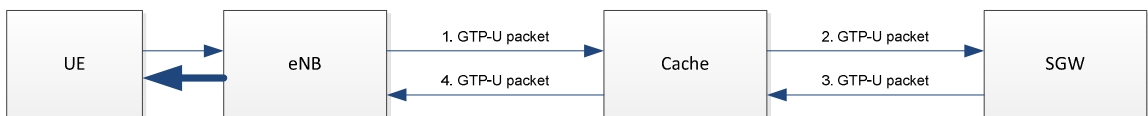


Figure 5.2.1-3 cache server in the middle of S1-U

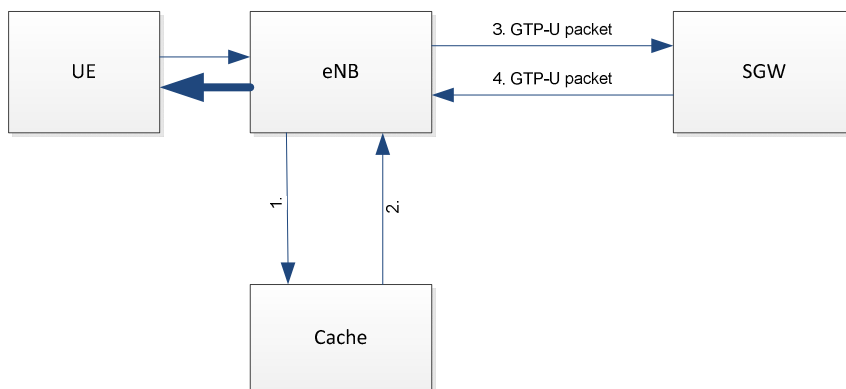


Figure 5.2.1-4 cache server off the S1-U path

5.2.2 Conclusion for Issue 1

The local caching can be used to address the issue on backhaul long latency. The local caching is feasible, if CN functionalities and principles when applied to local caching as defined in section 5.1 are respected. How to make it should be considered in the normative work with respect existing architecture.

5.3 Solutions for issue 2

5.3.1 TCP Performance Enhancing Proxy (TCP PEP)

TCP PEP [2] has been widely discussed in IETF Performance Implications of Link Characteristics WG (PILC). A PEP is used to improve the performance of the Internet protocols on network paths where native performance suffers due to characteristics of a link or subnetwork on the path. Distinct standards track recommendations for the performance mitigation of TCP over links with high error rates, links with low bandwidth, and so on, have been developed or are in development by the Performance Implications of Link Characteristics WG (PILC). The TCP PEP is a transport layer PEP implementations. The most common PEP variant is the split connection TCP Proxy, which intercepts the TCP handshake (connection establishment) between the User Equipment (UE) and the server, becomes the TCP endpoint towards the UE and establishes a TCP connection with the server on behalf of the UE. After the connection setup, the TCP Proxy operates the separate server side (upstream) and UE side (downstream) connections asynchronously. The TCP Proxy has been investigated in Wideband Code Division Multiple Access systems as well as in LTE [3][4][5][6][7].

There are two further options depend on where the TCP PEP is located.

- Option 1a: TCP PEP is deployed after P-GW/L-GW

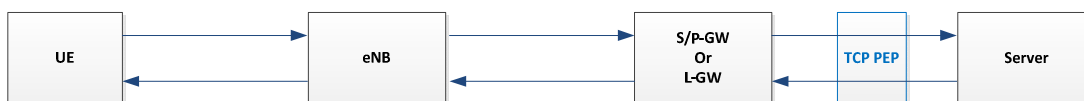


Figure 5.3.1-1 example for TCP PEP after P-GW

This option is transparent to the UE. It does not require changes to standard. It can also use the Commercial-Off-The-Shelf TCP PEP product.

- Option 1b: TCP PEP is collocated with the eNB

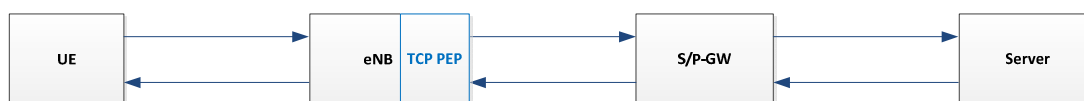


Figure 5.3.1-2 – example for TCP PEP collocated with the eNB

This option is transparent to the UE. This option can be considered as eNB's implementation, and not require changes to the standard.

5.3.2 Network based Radio-Aware TCP

Achievable data rate (ADR) could be used to improve the accuracy of TCP congestion control. In this solution, eNB provides the ADR to server via CN.

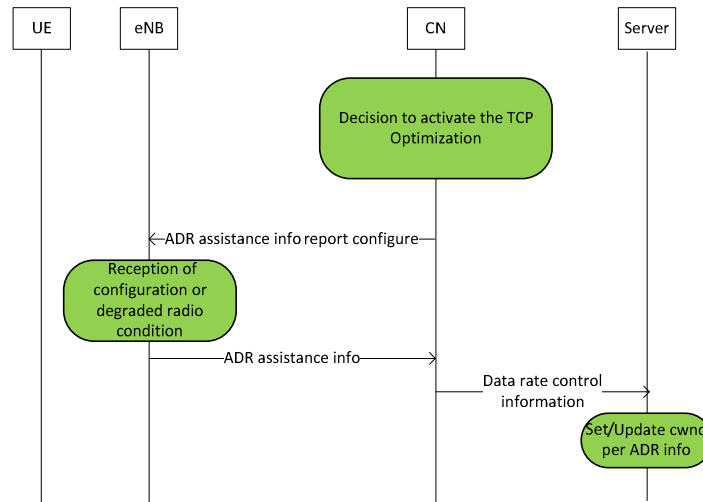


Figure 5.3.2-1: Network assisted radio-aware TCP solution

1. Upon decision to perform TCP optimisation, the CN will request eNB to report ADR, for this UE. In this step, the CN may also configure the reporting criteria and other related information. During handover procedure, CN may configure the target eNB to report optimization parameters by CN request or source eNB directly transfers related ADR report configuration to target eNB.
2. According to the ADR assistance info report configuration, eNB reports the required ADR assistance info to CN. This can be reported periodically or event triggered by CN configuration (e.g. ADR drops below a certain threshold).
3. CN sends ADR assistance info to TCP server; then TCP server set/update its congestion window accordingly.

Eligibility check similar to that provided for solution 1 of issue 1 may be applied, in case TCP optimization is only applied to some kind of user subscription, e.g. VIP users.

Potential impact:

1. CN sends ADR assistance information report configuration to eNB;
2. ADR report configuration transferred to the target eNB;
3. eNB sends required ADR assistance information to CN;
4. Eligibility check related impacts as already presented in solution 1 for issue 1.

5.3.3 UE Based Radio-Aware TCP Solution

In this solution, UE provides assistance information e.g. ADR to the server. The assistance information is derived by UE with assistance from eNB.

5.3.4 eNB generates TCP ACK on behalf of the UE

In this option, the eNB monitors the DL TCP packet. After the eNB receives the RLC ACK from the UE, the eNB generates the TCP ACK. To avoid the UE and eNB generates TCP ACK for a same TCP packet, the UE can request the eNB to activate/deactivate this function. This option mainly reduces the period for the UE to send the TCP ACK.

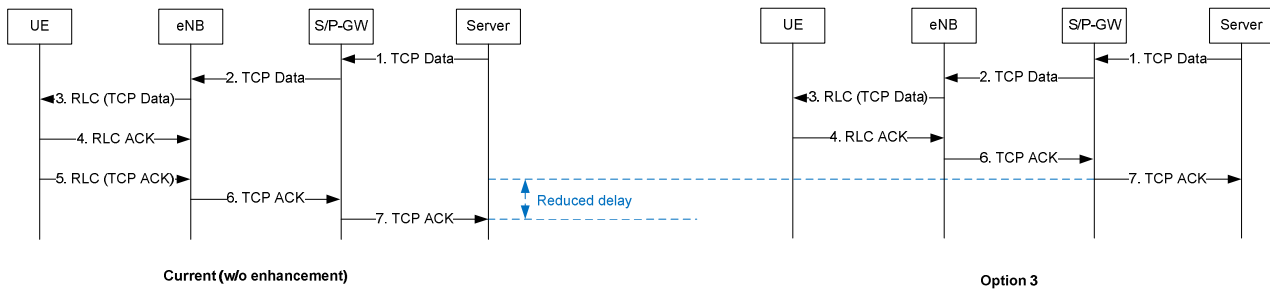


Figure 5.3.4-1: example for eNB generate TCP ACK on behalf of the UE

5.3.5 Conclusion for Issue 2

There is no consensus on whether TCP PEP is enough for LTE, or possibility if the TCP behavior needs to be changed.

5.4 Solutions for issue 3

RAN needs to check whether the UE is legal for the DASH optimisation from core network.

5.4.1 Solutions for case 1

5.4.1.1 Solution 1: UE request a higher QoS profile

Video transmission issues resulting from empty buffer can be mitigated by supplying UE’s higher QoS profile. If the operator decides to send video content over a non-GBR bearer initially, it may be useful to initiate bearer modification procedure, by supplying GBR bearer or higher priority non-GBR bearer, to serve the UE with higher QoS profile for OTT service. Even for non-GBR bearer, current TS23.203 defines various QCIs, e.g. QCI 8 for “premium bearer” and QCI 9 for “default bearer”.

Table 5.4.1.1-1: Standardized QCI characteristics

QCI	Resource Type	Priority Level	Packet Delay Budget (NOTE 13)	Packet Error Loss Rate (NOTE 2)	Example Services
8 (NOTE 5)		8	300 ms (NOTE 1)	10 ⁻⁶	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9 (NOTE 6)		9			
NOTE 5: This QCI could be used for a <u>dedicated "premium bearer"</u> (e.g. associated with premium content) for any subscriber / subscriber group. Also in this case, the SDF aggregate's uplink / downlink packet filters are known at the point in time when the SDF aggregate is authorized. Alternatively, this QCI could be used for the default bearer of a UE/PDN for "premium subscribers". NOTE 6: This QCI is typically used for <u>the default bearer of a UE/PDN for non privileged subscribers</u> . Note that AMBR can be used as a "tool" to provide subscriber differentiation between subscriber groups connected to the same PDN with the same QCI on the default bearer.					

5.4.1.2 Solution 2: Video Context Aware Solution

For video on demand, the application packets request will be triggered intermittently which depends on the buffer status for progressive download mode. Currently, the eNB does not know the UE’s playout status, e.g. buffer status. This may lead to unnecessary stalling. To solve the problem, the context aware service delivery solution is depicted in the figure 5.4.1.2-1 below:

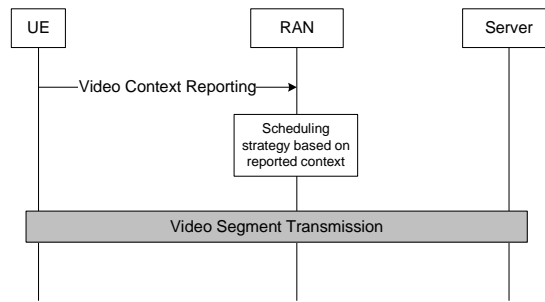


Figure 5.4.1.2-1 video context reporting procedure

Step1: The UE reports the video context information to the eNB.

This can be reported periodically or triggered by the configured event which depends on UE configuration. For example, the UE can report the context before or when requesting the next video segment. The video context information may include following parameters:

- **Playout status:** to indicate whether it is an initial request or a request during playout.
- **Playout buffer status:** refers to level of the buffer in milliseconds to evaluate the stalling probability of the playout.

Step2: eNB decides when and how to schedule the requested video segment when it comes.

It is up to the eNB's implementation on how to use these parameters to adjust the scheduling, and the duration to adjust its scheduling.

The eNB scheduler may also consider the QoS parameters, cell load, receiving quality of the UE into account to make the final scheduling decision.

5.4.2 Solutions for case 2

5.4.2.1 Solution 1: PCC-based Solution

This solution is based on the current specification of the 3GPP system. In current 3GPP system, PCC-level signalling can already accomplish the communication of network QoS information to the client device (user equipment or UE) and therefore the DASH client can locally (within the UE) obtain the QoS information via internal APIs. The DASH server act as Application, which have knowledge of the application Function, type and of the MPD.

Following figure depicts an example PCC architecture delivering end-to-end QoS support for DASH services with the capability to interpret the media presentation description (MPD) in order to gain information on the application-layer parameters for DASH content. In the current PCC architecture, the application function (AF) interacts with the applications requiring dynamic policy and charging control. Hence, in order to provide QoS for DASH services, the AF can extract DASH content information from the MPD, map it into the appropriate attribute-value pairs (AVPs), and provide the AVPs to the policy and charging rules function (PCRF) over the Rx reference point. TS 26.247 Annex I describes such mapping. The PCRF combines the DASH-related AVPs received over the Rx reference point and the input received from the Gx and Gxa/Gxc reference points with user-specific policies data from the subscriber profile repository (SPR) to form session-level policy decisions and provides those to the PCEF and BBERF. In other words, the PCRF takes the subscriber information into account when setting QoS. Access-specific QoS parameters are then communicated to the UE from PCEF/BBERF. In particular, TS23.401 describes how the UE acquires QoS information during dedicated bearer activation and bearer modification with bearer QoS update. It is also noted in TS23.401 that "The UE may provide EPS Bearer QoS parameters to the application handling the traffic flow. The application usage of the EPS Bearer QoS is implementation dependent."

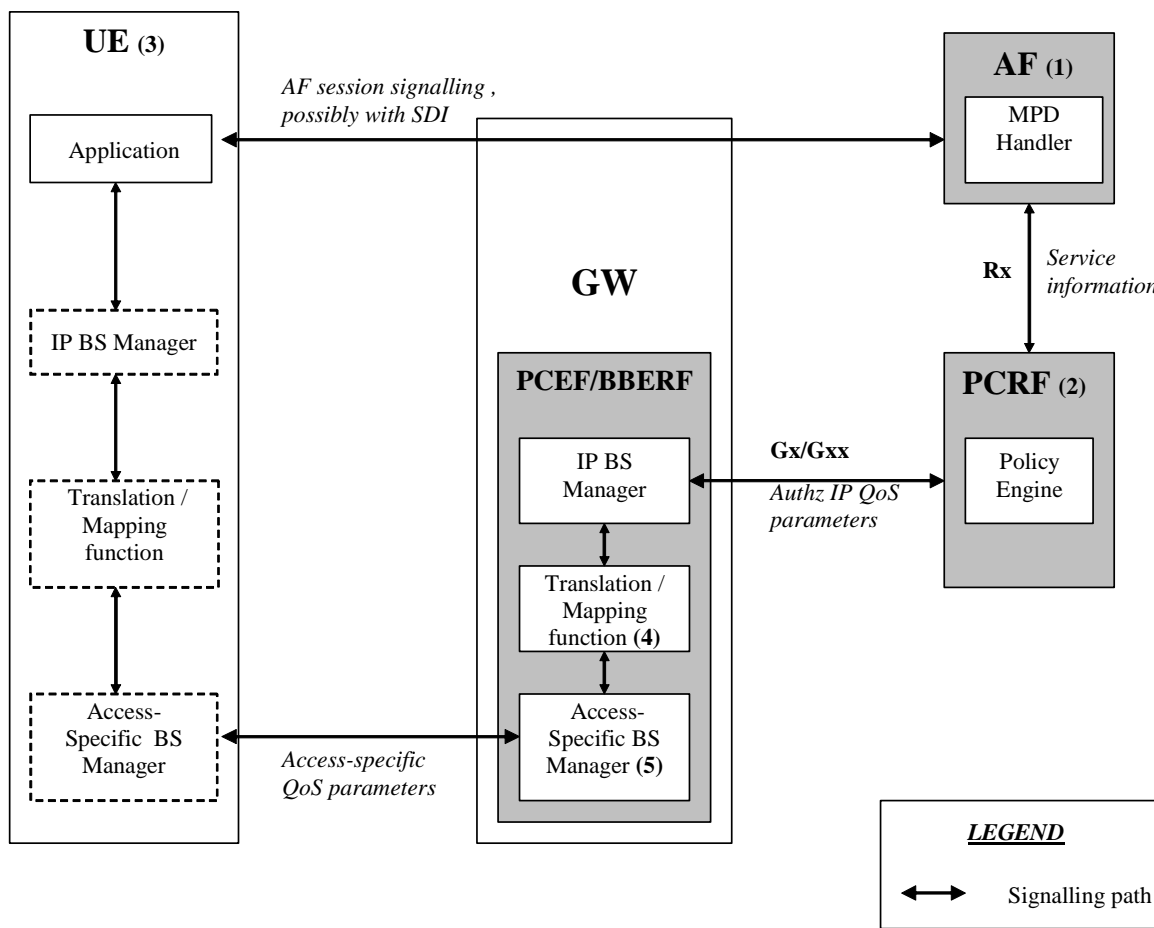


Figure 5.4.2.1-1: An example policy and charging control (PCC) architecture to deliver QoS for DASH services

5.4.2.2 Solution 2: Solution of DASH optimisation

In this solution, the RAN provides Predicted bandwidth to the client. The Predicted bandwidth for a specific UE is the throughput estimated to be available at the radio downlink interface in a period of time.

The whole procedure can be found in figure 5.4.2.2-1 below:

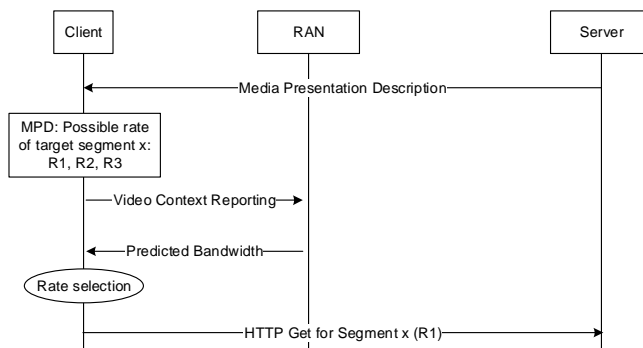


Figure 5.4.2.2-1: RAN Assistance Rate Selection

At first, the Client receives the MPD file of requested video from the Server side, which may include more than one rate for a segment.

And then, the Client may report the video context information to the eNB. The video context information includes the bandwidth range of the target segment and optionally with playout buffer status which will help RAN to give UE the predicted bandwidth more accurate. Details of these aspects were not addressed during the study

Finally, the RAN sends the predicted bandwidth to the Client. With the predicted bandwidth, the Client can select the rate of target segment suitably.

5.4.2.3 Solution 3: Solution of DASH optimisation via local breakout

In this solution network assistance information is provided in the form of a recommended media (audio/video) rate. This information is generated by the eNB, which is the node holding the information needed to recommend a media streaming bitrate.

In this solution the DANE is part of the eNB and therefore it has immediate access to the network assistance information derived from RAN resource monitoring.

The DANE is able to communicate the network assistance information via existing DASH signalling mechanisms, i.e. via HTTP signalling.

Impact due to initial setup of signalling connection between DANE and DASH Client were not addressed during the study.

The use of a local breakout bearer reduces over the air signalling between the RAN and the UE (hosting the DASH client) and it allows ready consumption of the network assistance information at the DASH client for an optimised QoE.

In order to receive the IP address of the DANE the UE could be configured with a URL pointing at the DANE. The UE could make a DNS lookup for the configured URL and receive an IP address that corresponds to the DANE.

The media content (i.e. video segments) is sent from the DASH server to the DASH Client upon opportune request of a particular bitrate from the DASH Client to the DASH Server. The request is based on the recommended media rate provided by the DANE to the DASH Client. Examples of the messages that could be exchanged between DANE and DASH Client for the purpose of network assistance information request and response can be found in section 6.4 of TR26.957.

Whether and how buffer level and video start information may be used by the DANE was not addressed during the study.

The solution is described in Figure 5.4.2.3-1:

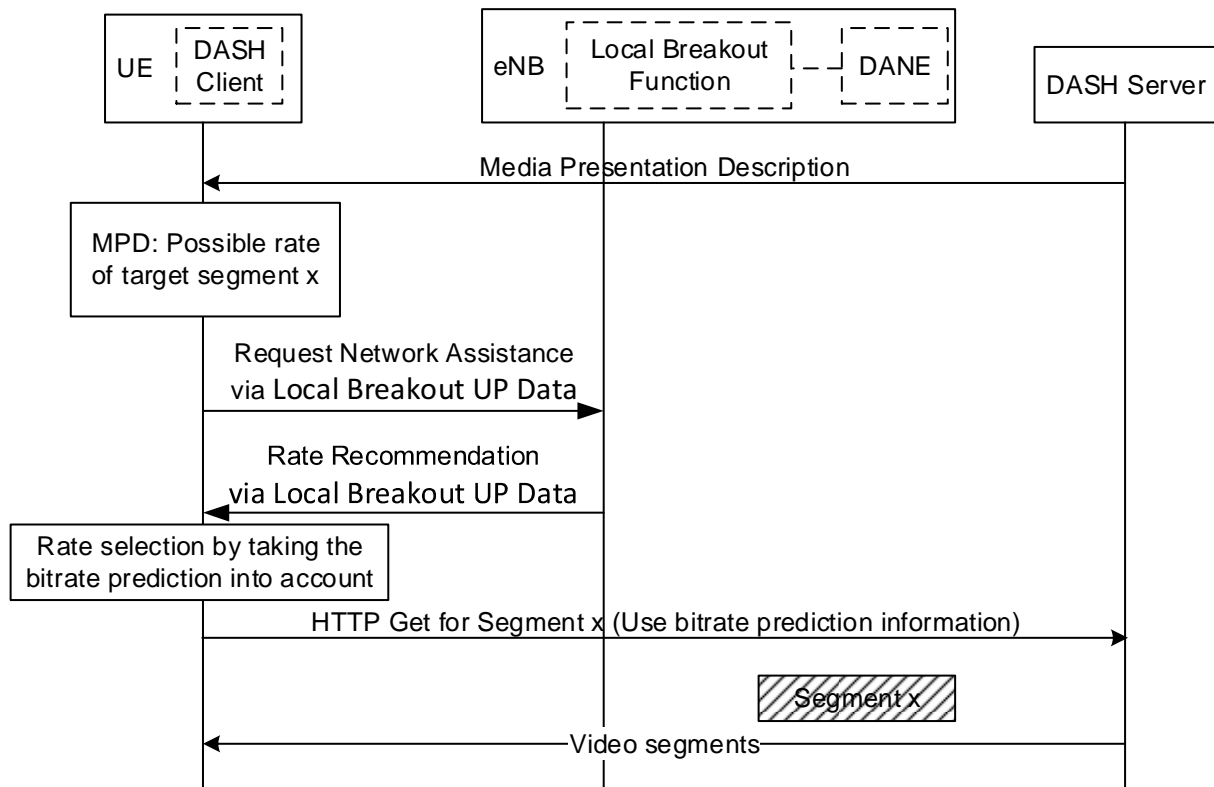


Figure 5.4.2.3-1: Solution for recommended rate information provisioning via DANE at RAN and Local Breakout UP bearers

5.4.3 Solutions for case 3

5.4.3.1 Solution 1

According to the issue description, the Empty buffer issue is caused by link throughput fluctuation. In case the OTT video uses GBR bearer, this may be very rare. It only happens when OTT video uses non-GBR bearer. SA4 already analysed this issue in TR23.976. The conclusion is

6.12.4 Conclusions and Gap Analysis

...

- such operation can be supported by the supplying a GBR bearer at a bitrate that enables real-time delivery of at least a minimum set (typically audio and video) of lower quality Representations.
- if the operator decides to send DASH content over a non-GBR bearer initially, it may be useful to initiate QoS update procedure with GBR QoS parameter once the network gets more loaded.

In addition, due to the nature of non-GBR bearer, using non-GBR bearer cannot guarantee the UE can play the video in a continuous way. Temporarily prioritize the video packets can only partially address the issue. The UE may frequently have empty buffer if the guaranteed bitrate is below the required bitrate for the lower quality representations.

Thus, to ensure a continuous play of video, GBR bearer should be used. The GBR bearer should have a bitrate larger than the minimum data rate required by the lower quality representations. If the operator decides to send video content over a non-GBR bearer initially, it may be useful to initiate bearer modification procedure, by supplying GBR bearer or higher priority non-GBR bearer, to serve the UE with higher QoS profile for OTT service.

5.4.4 Conclusion for Issue 3

For case 1, the solutions described in section 5.4.1.1 and 5.4.1.2 are feasible. For case 3, the solutions described in section 5.4.3.1 and 5.4.1.2 are feasible. For case 2, SA4 is defining solutions for the same video issue in work item “Server and Network Assisted DASH (SAND) for 3GPP Multimedia Services” [8] based on study in 3GPP TR 26.957 [9]. As this work develops further in SA4, there may be a need to address associated aspects in RAN3 scope if required (e.g. triggered by LS, new WI/building blocks, or otherwise).

It is also concluded that it may be beneficial for the UE to provide its buffer status to the RAN and it is up to RAN implementation on how to use such information for resource management optimisation. RAN2 should evaluate the benefits of this solution.

5.5 Solutions for issue 4

5.5.1 Solution 1: Dynamic adaptation of LTE radio resource allocation during the call

The UE sends a notification about the presence of a critical data on uplink bearer’s queue. The E-UTRAN accepts or declines to adapt its scheduling algorithm upon each notification based on its own mechanisms, and policy.

5.5.2 Solution 2: Static acceptance or rejection of adaptation of LTE radio resource allocation for the call

At multimedia call establishment, the E-UTRAN can inform IMS multimedia gateway or other entities in the Core network about its capability of receiving and handling accordingly, a notification from UE indicating critical multimedia data is pending to be sent on UE bearer’s queue. The IMS multimedia gateway or other entities in the Core network may then decide to send an indication to the E-UTRAN to accept or reject the notification from UE. This will be based on its own policy (e.g. depending on user subscription, network load, etc.) and applied to all notifications received from UE.

Note: The problem of how to validate the UE’s notification to protect the network against malicious UEs that request for prioritized resource allocation from the E-UTRAN was identified and not addressed during the study.

Annex A (informative): Video playout buffer aware scheduling

In current video transmission, video is either always not prioritized or always prioritized. Here is a simulation analysis for the performance of conditional prioritization for video i.e. buffer-aware scheduling:

- The scheduling for the UE is prioritized only when its video playout buffer level is low;
- When the UE's buffer level is high, the eNB may deprioritize the schedule to this UE.

Figure A-1 is the simulation result for the video capacity and stall probability under PF scheduling algorithm and playout buffer-aware scheduling.

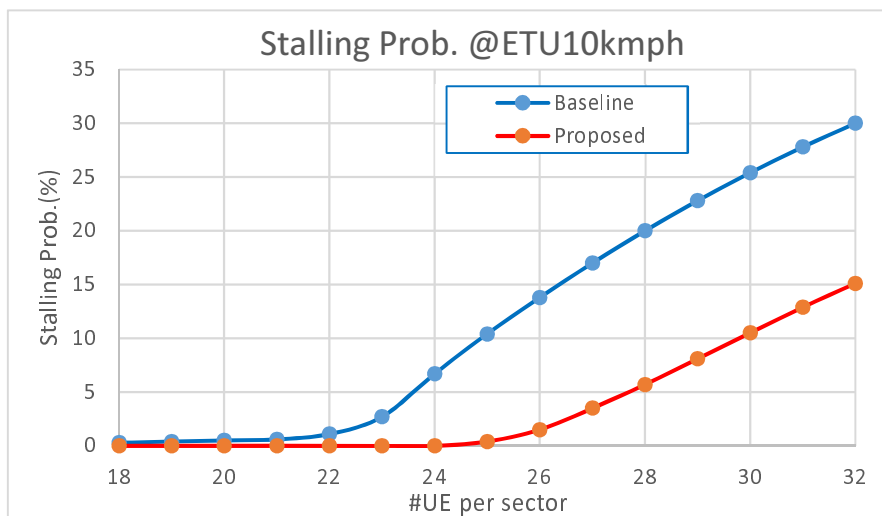


Figure A-1: Video Stall Probability Comparison

PF (Proportional Fair) is used as baseline. In PF, the scheduling priority is: $P=R/T$, where:

- R: the estimated data rate, if the UE is scheduled;
- T: the average data rate of this UE.

In playout aware scheduling, the scheduling priority is: $P=R/\max(B-I, 0.001)$, where:

- B: time to buffer empty at UE side, i.e. buffer level
- I: Initial playout buffer level, 1 sec is assumed in the simulation.

The simulation assumptions are:

- 10MHz system bandwidth
- All UEs have the same geometry 8.7dB,
- All the video segments have same size, and the data rate is 417kbps,
- eNB knows buffer level in real time.

The result shows that at 1% stalling probability level, PF can support 20 UEs, while playout buffer aware scheduling can support 25 UEs, i.e. 25% capacity gain, which may be the lower bound of the gain.

Annex B (informative): Change history

Change history							
Date	Meeting	TDoc.	CR	Rev	Cat	Subject/Comment	New
2016-04	RAN3#91 bis	R3-160812				TR Draft TR 36933 skeleton for Study on Context Aware Service Delivery in RAN for LTE	0.0.1
2016-04	RAN3#91 bis	R3-160969				TR Draft TR 36933 skeleton for Study on Context Aware Service Delivery in RAN for LTE	0.0.2
2016-04	RAN3#91 bis	R3-161027				Integration of agreed TP from R3-161026.	0.1.0
2016-05	RAN3#92	R3-161382				Presentation of the draft TR 36933_v0.1.1	0.1.1
2016-06	RAN3#92	R3-161471				Inclusion of agreed TP on general principles	0.2.0
2016-06	RAN3#92	R3-161477				Inclusion of agreed TP for TCP E2E delay	0.2.0
2016-06	RAN3#92	R3-161536				Inclusion of agreed TP on Video Transmission Issues	0.2.0
2016-09	RAN3#93	R3-162075				Integration of TP on UL Video transmission critical data discard issue from R3-162021	0.3.0
2016-09	RAN3#93	R3-162075				Integration of TP on DASH issue update R3-162022	0.3.0
2016-09	RAN3#93	R3-162075				Integration of TP on PCC based DASH solution from R3-162025	0.3.0
2016-09	RAN3#93	R3-162075				Integration of TP on Solution of DASH optimisation from R3-162026	0.3.0
2016-09	RAN3#93	R3-162075				Integration of TP on Video playout buffer aware scheduling from R3-162027	0.3.0
2016-09	RAN#73	RP-161768				Presentation of Report to TSG RAN73 of TR36933 Version 100	1.0.0
2016-10	RAN3#93 bis	R3-162603				Corrections to Solution 1 for DASH Optimisation	1.1.0
2016-10	RAN3#93 bis	R3-162607				Text Proposal for solutions to address issue 3 case 1 and case 3	1.1.0
2016-11	RAN3#94	R3-162885				Editorial update of TR 36933	1.1.1
2016-11	RAN3#94	R3-163210				Update of Draft TR 36933 (integration of agreed TPs)	1.2.0
2017-02	RAN3#95	R3-170727				Editorial update of TR 36933	1.2.1
2017-02	RAN3#95	R3-170759				Inclusion of TPs from R3-170760 , R3-170761 , R3-170889, R3-170772 . FFS clean up	1.3.0
2017-03	RAN#75	RP-170362				Presentation of Report to TSG RAN 73 of TR36933 Version 200 for Approval	2.0.0
2017-03	RAN#75					TR Approved by RAN plenary	14.0.0

History

Document history		
V14.0.0	April 2017	Publication