

Source: Ericsson, Huawei
Title: New SI proposal: Study on Latency reduction techniques for LTE
Document for: Approval
Agenda Item: 13.1.2

3GPP™ Work Item Description

For guidance, see [3GPP Working Procedures](#), article 39; and [3GPP TR 21.900](#).
Comprehensive instructions can be found at <http://www.3gpp.org/Work-Items>

Title: **SI: Study on Latency reduction techniques for LTE**

Acronym: FS_LTE_LATRED

Unique identifier:

NOTE: If this is a RAN WID including Core and Perf. part, then Title, Acronym and Unique identifier refer to the feature WI. Please tick (X) the applicable box(es) in the table below:

This WID includes a Core part	
This WID includes a Performance part	

1 3GPP Work Area

X	Radio Access
	Core Network
	Services

2 Classification of WI and linked work items

2.0 Primary classification

This work item is a ...

X	Study Item (go to 2.1)
	Feature (go to 2.2)
	Building Block (go to 2.3)
	Work Task (go to 2.4)

NOTE: Core, Performance and Testing parts of RAN WIs are usually Building Blocks.
If you are in doubt, please contact MCC.

2.1 Study Item

Related Work Item(s) (if any)		
Unique ID	Title	Nature of relationship

Go to §3.

2.2 Feature

Related Study Item or Feature (if any)		
Unique ID	Title	Nature of relationship

Go to §3.

2.3 Building Block

Parent Feature (or Study Item)		
Unique ID	Title	TS

This work item is ...

	Stage 1 (go to 2.3.1)
	Stage 2 (go to 2.3.2)
	Stage 3 (go to 2.3.3)
	Test spec (go to 2.3.4)
	Other (go to 2.3.5)

2.3.1 Stage 1

Source of external requirements (if any)		
Organization	Document	Remarks

Go to §3.

2.3.2 Stage 2

Corresponding stage 1 work item		
Unique ID	Title	TS

Other source of stage 1 information		
TS or CR(s)	Clause	Remarks

If no identified source of stage 1 information, justify:

Go to §3.

2.3.3 Stage 3

Corresponding stage 2 work item (if any)		
Unique ID	Title	TS

Else, corresponding stage 1 work item		
Unique ID	Title	TS

Other justification		
TS or CR(s) or external document	Clause	Remarks

If no identified source of stage 2 information, justify:

Go to §3.

2.3.4 Test spec

Related Work Item(s)		
Unique ID	Title	TS

Go to §3.

2.3.5 Other

Related Work Item(s)			
Unique ID	Title	Nature of relationship	TS / TR

Go to §3.

2.4 Work task

Parent Building Block		
Unique ID	Title	TS

3 Justification

Packet data latency is one of the performance metrics that vendors, operators and also end-users (via speed test applications) regularly measure. Latency measurements are done in all phases of a radio access network system lifetime, when verifying a new software release or system component, when deploying a system and when the system is in commercial operation.

Better latency than previous generations of 3GPP RATs was one performance metric that guided the design of LTE. LTE is also now recognized by the end-users to be a system that provides faster access to internet and lower data latencies than previous generations of mobile radio technologies.

In the 3GPP community, much effort has been put into increasing data rates from the first release of LTE (release 8) until the most recent one (release 12). Features like Carrier Aggregation (CA), 8x8 MIMO, 256 QAM have raised the technology potential of the L1 data rate from 300 Mbps to 4 Gbps. In Rel-13, 3GPP aims to introduce even higher bit rates by introducing up to 32 component carriers in CA.

However, w.r.t further improvements specifically targeting the delays in the system little has been done. Packet data latency is important not only for the perceived responsiveness of the system; it is also a parameter that indirectly influences the throughput. HTTP/TCP is the dominating application and transport layer protocol suite used on the internet today. According to HTTP Archive (<http://httparchive.org/trends.php>) the typical size of HTTP-based transactions over the internet are in the range from a few 10's of Kbytes up to 1 Mbyte. In this size range, the TCP slow start period is a significant part of the total transport period of the packet stream. During TCP slow start the performance is latency limited. Hence, improved latency can rather easily be shown to improve the average throughput, for this type of TCP-based data transactions. In addition, to achieve really high bit rates (in the range of Gbps with Rel-13 CA), UE L2 buffers need to be dimensioned correspondingly. The longer the RTT is, the bigger the buffers need to be. The only way to reduce buffering requirements in the UE and eNB side is to reduce latency.

Radio resource efficiency could also be positively impacted by latency reductions. Lower packet data latency could increase the number of transmission attempts possible within a certain delay bound; hence higher BLER targets could be used for the data transmissions, freeing up radio resources but still keeping the same level of robustness for users in poor radio conditions. The increased number of possible transmissions within a certain delay bound, could also translate into more robust transmissions of real-time data streams (e.g. VoLTE), if keeping the same BLER target. This would improve the VoLTE voice system capacity.

There are more over a number of existing applications that would be positively impacted by reduced latency in terms of increased perceived quality of experience: examples are gaming, real-time applications like VoLTE/OTT VoIP and video telephony/conferencing.

Going into the future, there will be a number of new applications that will be more and more delay critical. Examples include remote control/driving of vehicles, augmented reality applications in e.g. smart glasses, or specific machine communications requiring low latency as well as critical communications.

Various pre-scheduling strategies can be used to lower the latency to some extent, but similarly to shorter Scheduling Request (SR) interval introduced in Rel-9, they do not necessarily address all efficiency aspects.

It should also be noted that reduced latency of user plane data may also indirectly give shorter call set-up/bearer set-up times, due to faster transport of control signalling.

To ensure LTE evolution and competitiveness it appears therefore necessary to study and improve the packet data latencies.

4 Objective

4.1 Objective of SI or Core part WI or Testing part WI

The objective of this study item is:

- Study enhancements to the E-UTRAN radio system in order to:
 - Significantly reduce the packet data latency over the LTE Uu air interface for an active UE
 - Significantly reduce the packet data transport round trip latency for UEs that have been inactive for a longer period (in connected state).

The study area includes resource efficiency, including air interface capacity, battery lifetime, control channel resources, specification impact and technical feasibility. Both FDD and TDD duplex modes are considered.

As first aspect, potential gains like reduced response time and improved TCP throughput due to latency improvements on typical applications and use cases are identified and documented. In this evaluation RAN2 may assume latency reductions due to protocol enhancements as well as shortened TTIs. In conclusion, this aspect of the study is supposed to show what latency reductions would be desirable [RAN2].

As second aspect, the following areas should be studied and documented:

- Fast uplink access solutions [RAN2]:
 - for active UEs and UEs that have been inactive a longer time, but are kept in RRC Connected, focus should be on reducing user plane latency for the scheduled UL transmission and getting a more resource efficient solution with protocol and signaling enhancements, compared to the pre-scheduling solutions allowed by the standard today, both with and without preserving the current TTI length and processing times;
- From RAN1#83: TTI shortening and reduced processing times [RAN1]:
 - Assess specification impact and study feasibility and performance of TTI lengths between 0.5ms and one OFDM symbol, taking into account impact on reference signals and physical layer control signaling
 - backwards compatibility shall be preserved (thus allowing normal operation of pre-Rel 13 UEs on the same carrier);

4.2 Objective of Performance part WI

NOTE: Leave empty if the WI proposal does not contain a RAN performance part.

RAN time budget proposal

NOTE: For WIs/SIs under RAN WG5 leadership this section is not filled out. Otherwise:
For a not yet approved WI/SI the rapporteur has to fill out the last row of the table(s) below up to the target date of the WI/SI (if necessary add further tables): Indicate the number of time units (1 TU ~ 2h), i.e. one value for each session/field. If no time unit is needed, leave the field empty.
For WI/SI already approved in the past, the tables below will no longer be updated in the WI/SI description (i.e. the tables reflect the status of the initial approval). But changes can be proposed in the status report of the WI/SI.

RAN #67										Q2/2015										RAN #68			
R1L	R1U	R2L	R2U	R2J	R3	R4RF Core	R4RD Core	R4RF Perf	R4RD Perf	R1L	R1U	R2L	R2U	R2J	R3	R4RF Core	R4RD Core	R4RF Perf	R4RD Perf				
80bis	80bis	89bis	89bis	89bis	87bis	74bis	74bis	74bis	74bis	81	81	90	90	90	88	75	75	75	75				
												1											

RAN #68										Q3/2015										RAN #69			
R1L		R1U		R2L		R2U		R2J		R3		R4RF Core		R4RD Core		R4RF Perf		R4RD Perf					
82		82		91		91		91		89		76		76		76		76					
				1																			

RAN #69										Q4/2015										RAN #70			
R1L	R1U	R2L	R2U	R2J	R3	R4RF Core	R4RD Core	R4RF Perf	R4RD Perf	R1L	R1U	R2L	R2U	R2J	R3	R4RF Core	R4RD Core	R4RF Perf	R4RD Perf				
82bis	82bis	91bis	91bis	91bis	89bis	76bis	76bis	76bis	76bis	83	83	92	92	92	90	77	77	77	77				
		1								[1]		1											

RAN #70										Q1/2016										RAN #71			
R1L		R1U		R2L		R2U		R2J		R3		R4RF Core		R4RD Core		R4RF Perf		R4RD Perf					
84		84		93		93		93		91		78		78		78		78					
2																							

RAN #71										Q2/2016										RAN #72			
R1L	R1U	R2L	R2U	R2J	R3	R4RF Core	R4RD Core	R4RF Perf	R4RD Perf	R1L	R1U	R2L	R2U	R2J	R3	R4RF Core	R4RD Core	R4RF Perf	R4RD Perf				
84bis	84bis	93bis	93bis	93bis	91bis	78bis	78bis	78bis	78bis	85	85	94	94	94	92	79	79	79	79				
2										2													

L: LTE, U: UMTS, J: Joint, RD: RRM/demodulation

NOTE: In case further explanation of the time budget proposal is needed, then please explain this below.

additional comments to the time budget proposal:

4.3 RAN time budget proposal

NOTE: For WIs/SIs under RAN WG5 leadership this section is not filled out. Otherwise:
For a not yet approved WI/SI the rapporteur has to fill out the last row of the table(s) below up to the target date of the WI/SI (if necessary add further tables): Indicate the number of time units (1 TU ~ 2h), i.e. one value for each session/field. If no time unit is needed, leave the field empty.
Once the WI/SI is approved, the tables below will no longer be updated in the WI/SI description (i.e. the tables reflect the status of the initial approval). But changes can be proposed in the status report of the WI/SI.

L: LTE, U: UMTS, J: Joint, RD: RRM/demodulation

NOTE: In case further explanation of the time budget proposal is needed, then please explain this below.

additional comments to the time budget proposal:

- 5 Service Aspects
- 6 MMI-Aspects
- 7 Charging Aspects
- 8 Security Aspects
- 9 Impacts

Affects:	UICC apps	ME	AN	CN	Others
Yes		X	X		
No	X				
Don't know				X	

10 Expected Output and Time scale

New specifications [If Study Item, one TR is anticipated]						
Spec No.	Title	1st rsp. WG	2nd rsp. WG(s)	Presented for information at plenary#	Approved at plenary #	Comments
TR 36.8yz	Study on Latency Improvements in LTE	RAN2	RAN1, RAN3		RAN#72, June 2016	

NOTE: If this is a RAN WID including Core and Perf. part, then all new Core part specs have to be listed first and then all new Perf. part specs. Indicate "Core part" or "Perf. part" under Comments for each spec. By default a new specs can only be new for one of both parts.

Affected existing specifications [None in the case of Study Items]				
Spec No.	CR	Subject of the CR	Approved at plenary#	Comments

NOTE: If this is a RAN WID including Core and Perf. part, then all new Core part specs have to be listed first and then all new Perf. part specs. Indicate "Core part" or "Perf. part" under Comments for each spec. If an existing spec is affected by both (Core part and Perf. part), then it has to be listed twice with appropriate approval dates.

11 Work item rapporteur(s)

Enbuske, Henrik
Company: Ericsson
Email: Henrik.enbuske@ericsson.com

12 Work item leadership

RAN WG2

NOTE: If this is a RAN WID including Core and Perf. part, then this WG specifies the WG leading the Core part. RAN WG4 is by default leading the Perf. part.

Supporting Individual Members

Supporting IM name
Ericsson
AT&T
Panasonic
SouthernLINC
Telefonica
TeliaSonera
Deutsche Telekom
Sprint
Telenor
Apple
InterDigital
NVIDIA
T-Mobile USA
AsusTek
Huawei
HiSilicon
HTC
III
Coolpad
Fujitsu
ITRI
Potevio
China Telecom
China Unicom
Telecom Italia
DISH network
Microsoft
Elektrobit Wireless Comm.
U.S. Doc
UK Home Office
LightSquare
Bell Mobility
Nokia Corporation
Nokia Networks
Qualcomm Incorporated
Samsung
CATT
US Cellular
KT
Intel Corporation
Telus