Source: Huawei, HiSilicon

Document for: Information

Agenda Item: 13.1.2

Motivation of New SI Proposal Latency Reduction

www.huawei.com

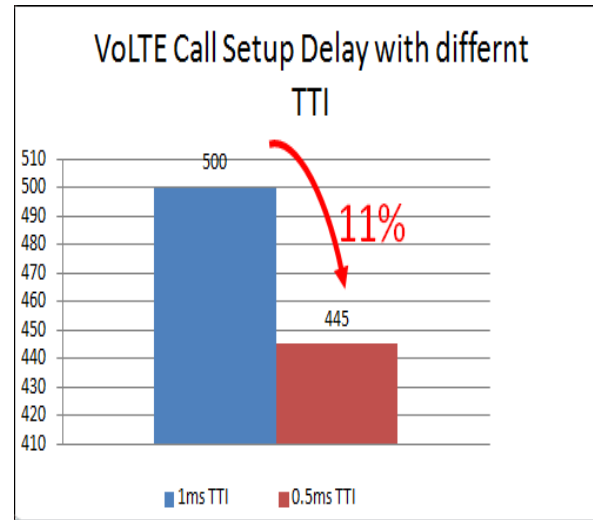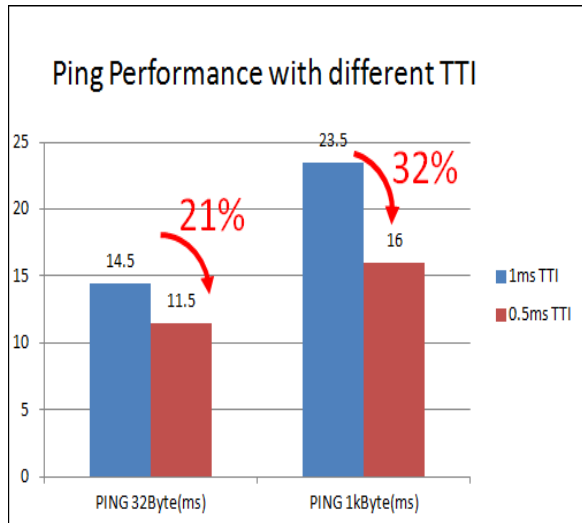HUAWEI TECHNOLOGIES CO., LTD.

Motivation for latency reduction

- Latency is crucial for mobile network applications and it is essential to reduce the time-to-wait for the user experience[8][9]

- Traditional services like VoIP and real time applications (e.g. interactive gaming) need low latency. Lower latency is also required by more and more emerging applications with small packet transmission

- The end to end latency requirements for the targeted applications:

  - Gaming on Sport Events (< 25 ms) [1]

  - Auto Pilot((Car-to-car Communications) (<30 ms) [1]

  - Remote control (RTT < 50 ms) [1]

  - Teleprotection in smart grid network (< 8 ms) [3]

  - Traffic efficiency and safety (<5 ms) [3]

  - Public safety (Call setup < 300 ms; E2E media transport <150 ms)
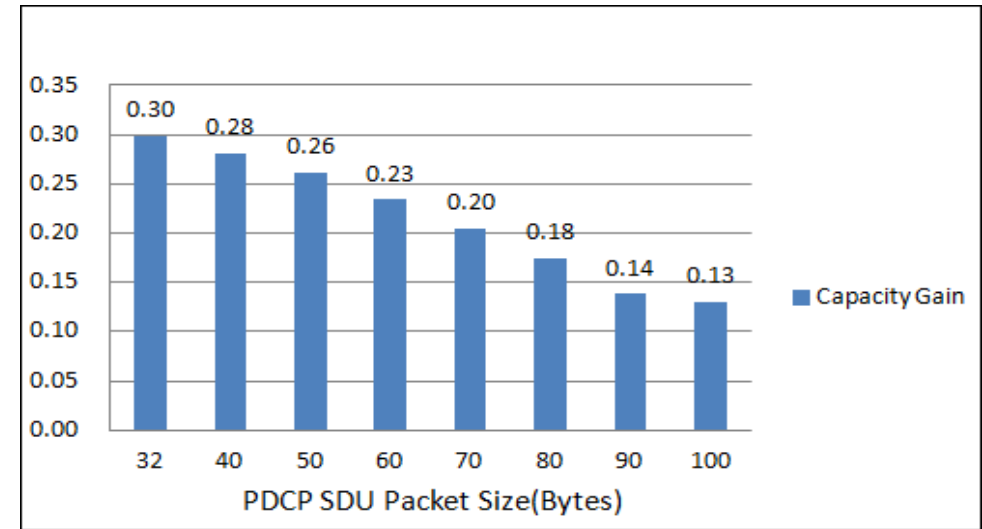
Latency reduction with shorter TTI

- Using shorter transmission time interval can reduce the RTT(i.e. Round Trip Time) over air interface and then shorten the latency
  - The current TTI design of LTE works well for the delay tolerant services.
  - Shorter TTI can enable more efficient support for the packets with stringent latency requirements, especially for the small packet transmissions [Reference 2][sec 7.1 of Reference 7]
  - In order to further reduce the latency, shorter TTI needs to be supported in both downlink and uplink
- The advantages of shorter TTI based transmission
  - Reduce call setup delay and user plane latency
  - More efficient resource utilization
  - Improve TCP throughput
  - Increase transmission reliability

KPI and resource efficiency improvement via shorter TTI

Ping Performance with different TTI

VoLTE Call Setup Delay with differnt TTI
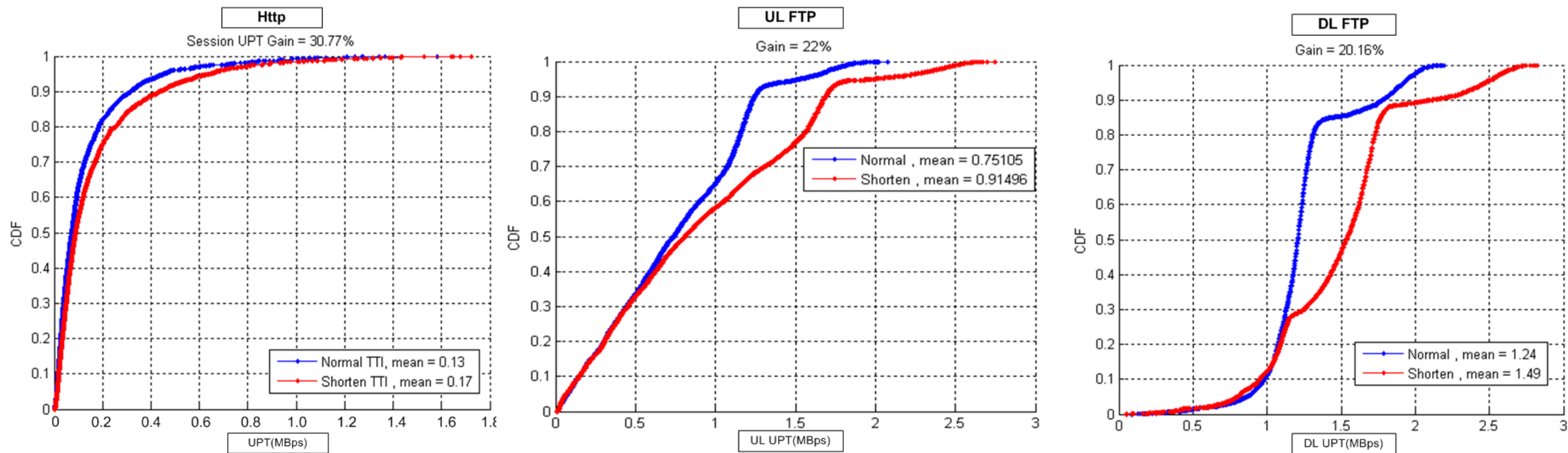

PDCP SDU Packet Size(Bytes)

Observation

- The user plane latency can be reduced with the utilization of shorter TTI
- The performance of Ping-Test can be speed up due to low air interface RTT
- The end to end VoLTE call set-up can also benefit from shorter TTI (no core network optimization is assumed in the calculation)

Observation

- Short TTI can reduce the padding proportion and increase the resource efficiency, i.e. system capacity.
- The average capacity gain is up to 30%
- If ROHC is used, the expected average gain can be even larger
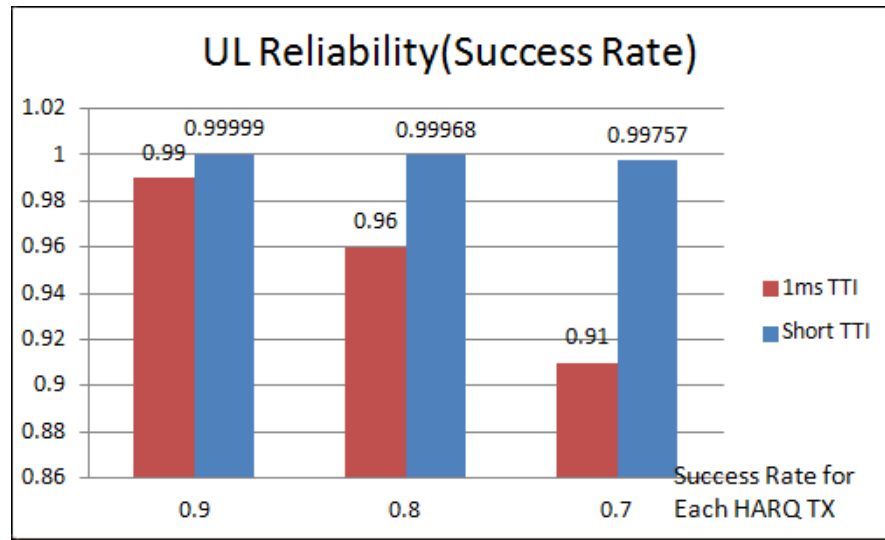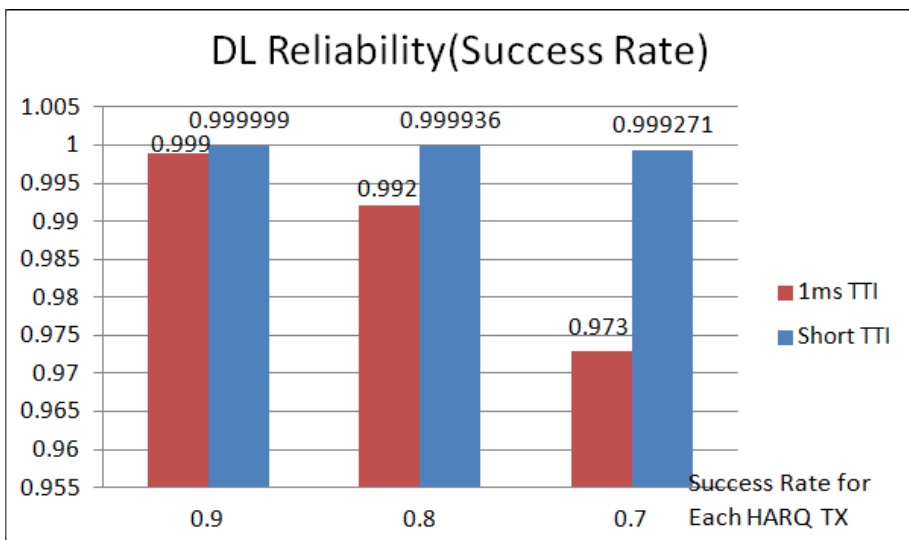
TCP throughput improvement via shorter TTI

The traffic models of the simulation above are referred to [6][7]

Observation: Lower latency enabled by shorter TTI can accelerate the feedback of TCP ACK, the increase of the TCP Window and the slide of TCP Window, and accordingly the increase of the TCP throughput

Transmission reliability increase via shorter TTI

- With shorter TTI, more transmission opportunities are possible within the delay budget for a particular QoS type.
- If the delay budget for the transmission is less than 25ms, the possible maximum HARQ transmission number for DL and UL with different TTI/RTT are assumed in the table
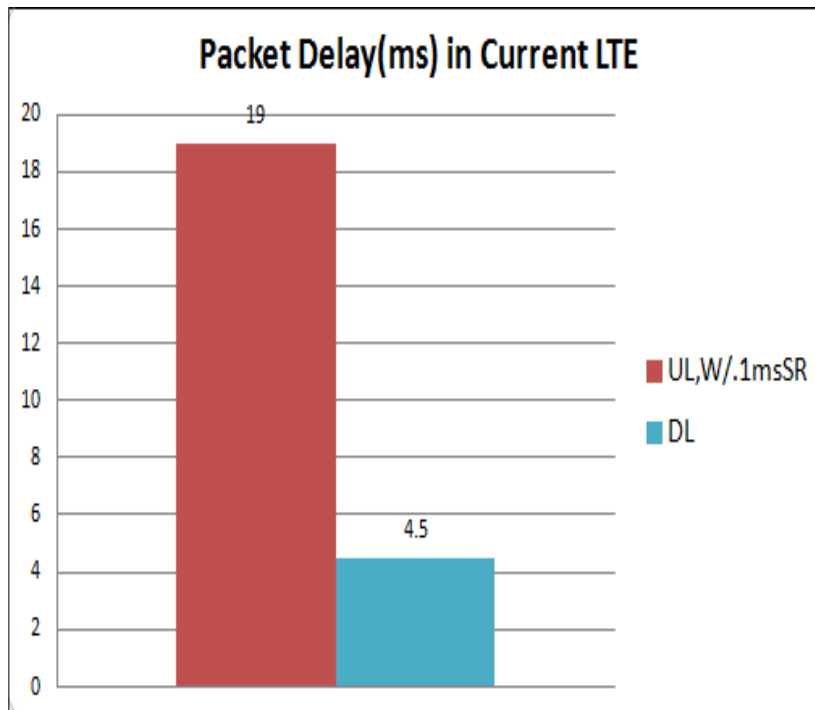
DL/UL	1ms TTI (8ms RTT)	0.5ms TTI (4ms RTT)
DL Max. HARQ Tx. Num.	3	6
UL Max. HARQ Tx. Num.	2	5

DL Reliability(Success Rate)

UL Reliability(Success Rate)

Observation: more transmission opportunities enabled by shorter TTI can increase transmission reliability
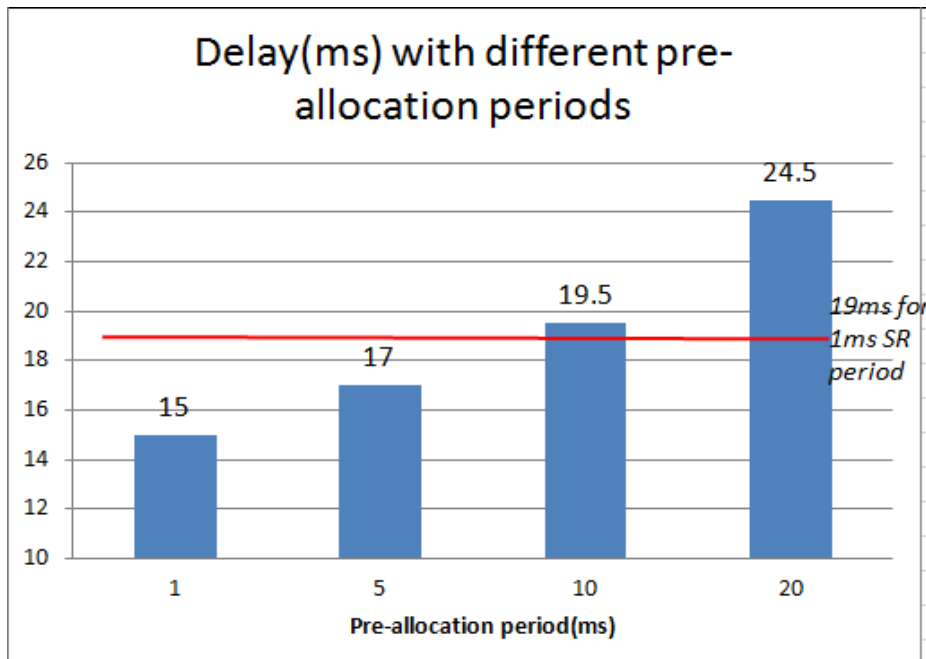
Latency reduction with UL enhancements

- In practice, the uplink delay for the packet transmission is larger than downlink delay [10]
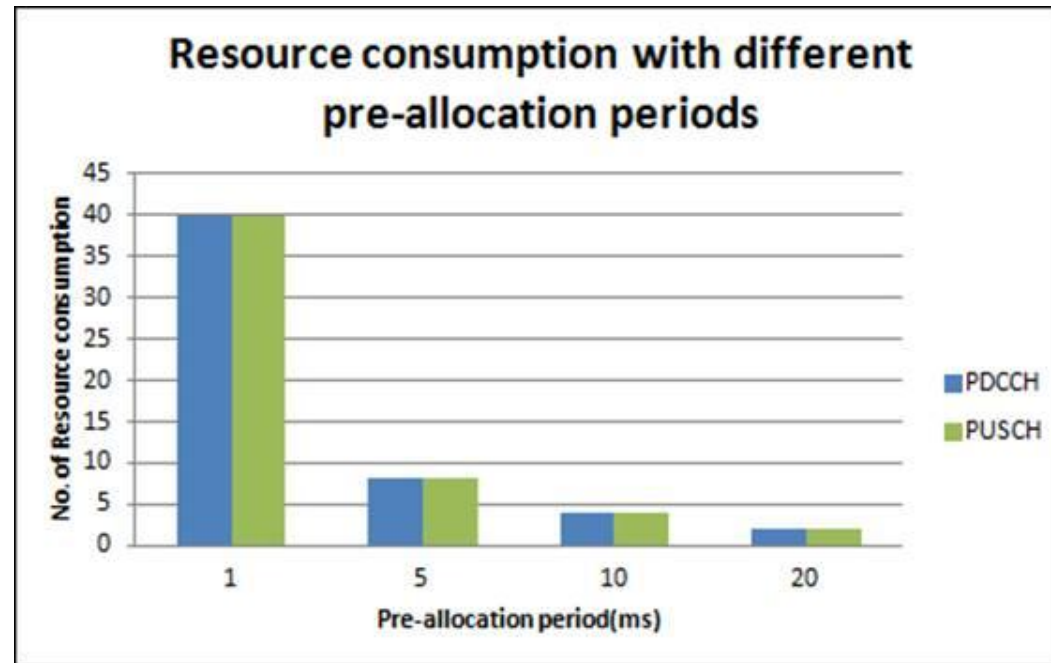- Additional enhancements for uplink needs to be studied other than shorter TTI

Packet Delay(ms) in Current LTE

Legend:
- UL,W/.1msSR
- DL

Values: 19 (UL,W/.1msSR), 4.5 (DL)

The calculation is based on section B.2 of reference [10]

Additional Enhancements for uplink transmission

Delay(ms) with different pre-allocation periods

Assuming the delay is the combination of the transmission time and the process time

Resource consumption with different pre-allocation periods

Assuming the packet inter-arrival time is 40ms and one PDCCH and PUSCH is required for the packet

Observation
- Pre-allocation is better than SR procedure in latency reduction only if the period is small enough, i.e. less than10ms
- The pre-allocation with smaller period consumes more PDCCH resources, PUSCH resources and UE power

Conclusion
- The study of additional enhancements for uplink transmission is needed to support lower latency with higher resource efficiency and power efficiency

Objectives of the study

- Shorter TTI can be studied to provide the latency reduction for both downlink and uplink
  - The start point for shorter TTI study can be 0.5ms (i.e. one Time Slot).
  - The backward compatibility needs to be ensured when supporting the configuration of the new TTI length
- Additional enhancement for uplink latency reduction needs to be studied other than shorter TTI
  - E.g. the schemes that can reduce the uplink transmission delay with higher resource efficiency and power efficiency than the current pre-allocation scheme.
- In addition, other enhancements for latency reduction can also be studied if the allocated time unit fits

Thank you

www.huawei.com

Annex 1: Capacity gain simulation assumption

- Small packets are an important part for packet transmission
- Small packets consume less PRBs
- Padding are un-avoided in a TB, and smaller TB may lead to higher padding proportion

Traffic	Packet Size
On-line Gaming[1]	50B~200B
M2M [4]	20B~200B, 46.4%<50B
Background [5]	UL/DL, 90%<100B
IM[5]	UL/DL, 80%<100B
Gaming[5]	UL<100B, DL, 50%<100B
Interactive content pull[5]	UL, 90%<100B, 80%<50B
TCP ACK	40Bytes
ROHC	Change 40B to about 5B
BSR/RLC ACK	3Bytes

Annex 2: References

- [1] FP7 LOLA /D2.1_Target_Application_Scenarios

- [2] R1-050246

- [3] ICT-317669-METIS/D2.1 Requirement analysis and design approaches for 5G air interface

- [4] TR 45.820

- [5] TR 36.822

- [6] 80216m-08_004r2-IEEE 802.16m Evaluation Methodology Document (EMD)

- [7] TR 36.814

- [8] RP-142026 Motivations for Study on latency reduction techniques

- [9] RP-142025 New SI proposal: Study on Latency reduction techniques for LTE

- [10] TR36.912