Latency in HSPA Data Networks

Siddharth Mohan, Rohit Kapoor, Bibhu Mohanty

Abstract:

Latency and throughput are two critical performance metrics of a communication network. Recently, a lot of attention has been focused on improving throughput (or spectral efficiency) of Wireless Wide Area Networks (WWANs) through the use of physical and MAC layer techniques, such as higher order modulation, MIMO and aggregation of bandwidth (multi-carrier). While some data applications directly benefit from the higher data rates, for many applications high data rates do not translate to improved user experience unless the latency is low. In this paper, we examine the Control plane (C-plane) and User plane (U-plane) latencies in an HSPA data system. We show that significant C-plane latency reduction can be achieved in HSPA by carrying signaling on HS-DSCH and E-DCH channels (as opposed to dedicated channels, which is the practice now). We also compare the C-plane and U-Plane latencies of HSPA and LTE, which have comparable spectral efficiency [1].

1. Introduction

3rd Generation Partnership Project (3GPP) has standardized Code Division Multiple Access (CDMA) based packet-switched air-interfaces for downlink and uplink, called High-Speed Downlink Packet Access (HSDPA) and High-Speed Uplink Packet Access (HSUPA) respectively, together referred to as High Speed Packet Access (HSPA) [2][4]. In today's HSPA deployments, signaling messages are typically carried on dedicated channels. These dedicated channels are typically low rate channels (3.4 kbps), which allows larger spreading factors to be used so that more users can share the cell's resources. Since some signaling messages tend to be fairly large, the use of these low rate dedicated channels leads to call setup latencies that are in the range of a few seconds. Newer HSPA deployments are considering carrying signaling on HSPA channels (i.e., HS-DSCH on the downlink and E-DCH channels on the uplink). Such configurations not only allow better statistical use of power and code resources, but also allow signaling to benefit from higher rate HSPA channels. We will show that significant improvements to call setup latency can be achieved by carrying signaling on HSPA channels.

In parallel with the evolution of HSPA, 3GPP has also standardized an Orthogonal Frequency Division Multiplexing (OFDM) based air-interface, called Long Term Evolution (LTE) [3]. Though the spectral efficiencies of the two systems are comparable, Rel-9 LTE provides higher peak rates (due to its wider 20MHz bandwidth) compared to Rel-8/Rel-9 HSPA, which allows a User Equipment (UE) to receive on 10 MHz bandwidth with carrier aggregation. This paper focuses on the *call setup* and *user plane* latencies of the two systems. *Call setup latency* determines how quickly the user can start to receive service, while *user plane latency* is important since the performance of applications such as web browsing is very sensitive to the Round Trip Time (RTT) to the server.

In this paper, we compare call setup latency of HSPA when signaling is carried on dedicated channels versus HSPA channels. We also compare call setup and user plane latencies of HSPA and LTE. A detailed breakdown of the call setup as well as user plane latencies of HSPA and LTE systems is available in the Appendix.

2. Classification of Latency

Latency can typically be classified as:

User Plane Latency: Loosely speaking, User plane (U-plane) latency is the delay seen by an application in exchanging data with a server. A metric that is commonly used to characterize U-plane latency is the PING delay, i.e., the delay from a UE sending a small PING to the first IP node in the network and receiving the PING response. This PING latency is the definition of U-plane latency used in this paper.

Control plane latency: Control plane (C-plane) latency, also known as call setup latency, is the latency for a User Equipment (UE) to transition to a state where it can send/receive data.

The definition of call setup latency merits further clarification. The 3GPP specification for HSPA allows the UE to camp[1] with low battery consumption in either RRC idle or CELL_PCH (or URA_PCH, which is similar to CELL_PCH for the purposes of this paper) states [4]. In these states, the UE only needs to listen to the network in pre-assigned "on" periods, while the UE can save battery by shutting off its receiver during the rest of the DRX (Discontinuous Reception) cycle. Camping in CELL_PCH state has a number of benefits: the signaling cost and delay to transition the UE from CELL_PCH to a state where the UE can send/receive data (i.e., CELL_FACH) is very small compared to starting from idle state, while the UE's battery consumption in CELL_PCH and idle states is similar (assuming same DRX cycles for both).

Given the two types of camped states (idle and CELL_PCH) allowed by the HSPA specification, we consider two definitions of call setup, one from RRC idle to CELL_DCH (we refer to this as *call setup from idle*) and one from CELL_PCH to CELL_FACH (we refer to this as *call setup from connected*). Legacy HSPA networks typically transitioned the UE to RRC idle state once data transfer was finished, so UEs experienced the latency of *call setup from idle*. However, with the deployment of CELL_PCH, we expect that UEs will typically camp in CELL_PCH, and only very long continuous inactivity (of the order of tens of minutes or more) may cause them to be transitioned to RRC idle. Thus, in future HSPA networks, we expect CELL_PCH (as opposed to idle) to be the default camped state from which the UE will initiate most of its data transactions.

While CELL_PCH is supported in the specifications since R99, an Enhanced version of CELL_PCH was added in Rel 7 [4]. Enhanced CELL_PCH allows the UE to retain a dedicated H-RNTI in CELL_PCH state, which means that the UE is primed to send/receive data.

---

[1] Camping traditionally refers to camping of the UE in Idle mode. In this paper, we apply the term camping to mean user being inactive, which corresponds to Idle, CELL_PCH and URA_PCH states for HSPA, and connected state with long DRX for LTE.

In the case of LTE, the only two RRC states are idle and connected. In connected state, the UE is allowed to camp with a long discontinuous receive (DRX) cycle [5]: for the purposes of this paper, this can be considered the equivalent of the CELL_PCH state in HSPA. Thus, for LTE, the term *call setup from idle* refers to a transition from *RRC idle* to a state where data can be sent/received, and the term *call setup from connected* refers to a transition from *RRC connected with a long DRX cycle* to a state where *data can be sent/received*.

3. Overview of Results

In this section, we summarize the key results of this paper. We present the following comparisons:

Comparison 1: C*all setup from Idle*

C-Plane Scenario	Signaling on Dedicated Channels for HSPA (ms)	Signaling on HSPA Channels (ms)	LTE (ms)
MO	1717	546	188
MT	1920	749	365

Table 1: Latencies of Call setup from idle for HSPA (when carrying signaling on Dedicated channels or HSPA channels) and LTE

Comparison 2: *Call setup from connected*

C-Plane Scenario	CELL_PCH to CELL_FACH for HSPA (ms)	Enhanced CELL_PCH to CELL_FACH for HSPA (ms)	Long DRX to Connected for LTE (ms)
MO	82	14	13
MT	242	174	173

Table 2: Latencies of Call setup from connected for HSPA (when starting from CELL_PCH or Enhanced CELL_PCH) and LTE

Comparison 3: *U-Plane latency*

U-Plane Process	HSPA (ms)	LTE (ms)
PING latency	38	32

Table 3: U-Plane latencies of HSPA and LTE

A detailed analysis of how the numbers in this section are derived, including call flows and assumptions, is shown in the Appendix.

Conclusion

In this paper, we analyzed the latency of *call setup from idle* for a legacy WCDMA network where signaling is carried on dedicated channels. We showed that a significant (**~1.2 sec**) gain in latency of *call setup from idle* can be achieved by carrying signaling on HSPA channels. We showed that LTE further reduces the latency of *call setup from idle* latency by ~360 ms compared to an HSPA system where signaling is carried on HSPA channels. One of the key reasons that contribute to this gain for LTE is that the Radio Bearer Setup message is assumed to be a synchronized message for HSPA, whereas the corresponding message in LTE is unsynchronized. Unsynchronized Radio Bearer Setup is supported by the HSPA specification, but we did not consider it in this paper since it is not commonly deployed.

We also compared the latencies of *call setup from connected* between HSPA and LTE systems and showed them to be similar, if the Enhanced CELL_PCH state, introduced in Release 7 of the HSPA specification, is considered. If starting from legacy CELL_PCH, HSPA latency is 69-75 ms more than that of LTE. We expect that UEs will typically camp in *connected camped* state (i.e., CELL_PCH for HSPA and connected with long DRX cycle for LTE), and thus latency of *call setup from connected* is likely to drive the user's perception of the *responsiveness* of the system when the UE is camped.

Finally, we compared the U-plane latency (or PING request-response latency) of HSPA and LTE systems. We showed that, with similar H-ARQ operating points and similar processing latencies of network nodes, the U-plane latencies for HSPA and LTE are similar.

Appendix

A1. Transport Network Technology and Architecture for HSPA and LTE systems

Figure 1 shows the network architecture for UMTS and the Evolved Packet System (EPS) flat architecture for LTE. In the EPS flat architecture, there is user plane direct tunneling between the Serving Gate way (S-GW) and the eNodeB. For EPS, RNC functions are partly in the eNodeB and partly in the the MME, SGSN functions are in the MME, GGSN functions are in the S-GW and the PDN-GW.

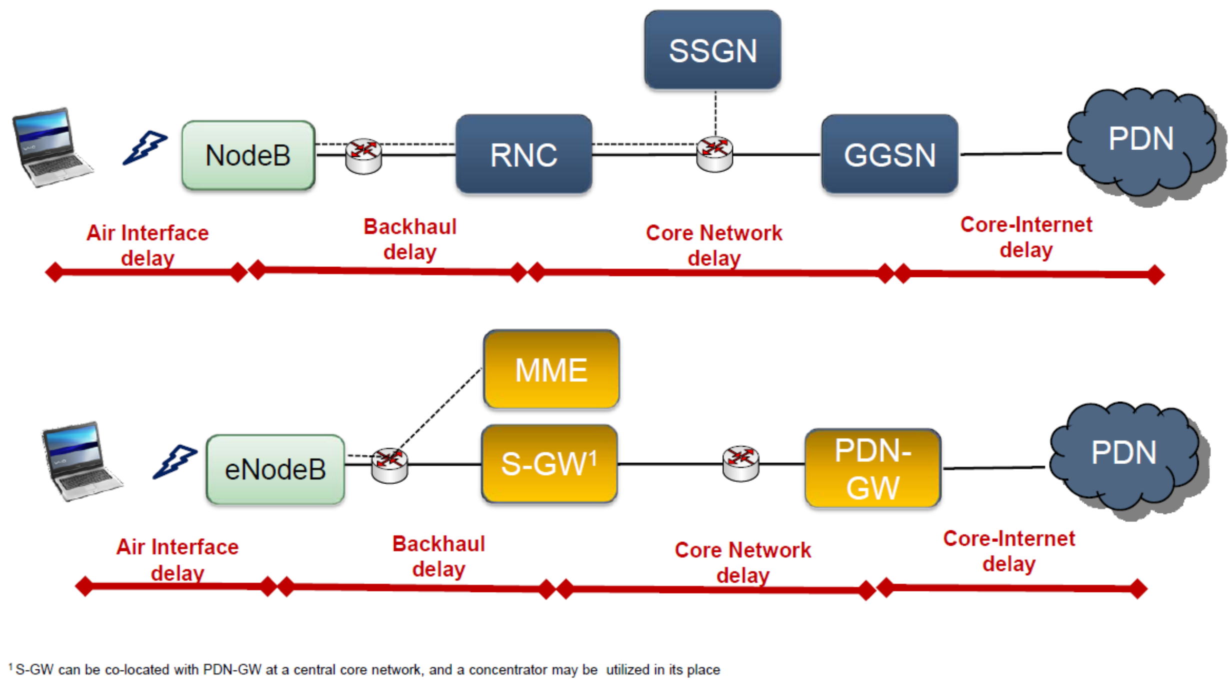

Figure 1: Transport Network Architecture for HSPA and LTE

A2. WCDMA Latency of Call Setup from Idle when Signaling is carried on Dedicated Channels

In this section, we present a breakdown of the typical latency of *call setup from idle* in legacy 3GPP WCDMA networks. This section focuses on the latency when signaling is carried over dedicated channels, whereas Section A3 compares this with the case where signaling is carried over HSPA channels (i.e., HS-DSCH and E-DCH channels for downlink and uplink respectively). We consider both mobile originated (MO) as well as mobile terminated (MT) cases.

Mobile Originated Call Setup from Idle

The call flow for a Mobile Originated (MO) *call setup from idle* for a packet switched (PS) call is shown in Figure 2 [4]. Some of the key procedures involved in call setup are:

1. *RRC Connection Establishment Procedure:* The purpose of this procedure is to transition the UE from RRC idle mode to RRC connected state. The messages involved are RRC Connection Request initiated by the UE, RRC Connection Setup sent by the RNC in response to the RRC Connection Request and RRC Connection Setup Complete sent by the UE.
2. *GPRS Attach Procedure:* The purpose of this procedure is to make the UE's presence known to the packet switched core network. The messages involved in this procedure are the Initial Direct Transfer sent by the UE to the core network and the Security Mode Command initiated by the core network to activate ciphering and integrity protection. The GMM Attach Accept sent to the UE completes the attach procedure.
3. *PDP Context Activation Procedure:* The purpose of this procedure is to assign an IP address to the UE and create a data path from the packet network to the UE. The steps involved are Activate PDP Context sent by the UE to the GGSN (not shown inFigure 2), radio link setup between the RNC and the NodeB, setting up of radio bearers between the RNC and the UE, and Activate PDP Context Accept sent to the UE.

Table 4 shows the breakdown of the latency for the steps shown in Figure 2. The numbers in Table 4 assume that signalling radio bearers are mapped to 3.4 kbps dedicated channels with 40ms TTI. The sizes of the messages exchanged in the MO call flow are shown in Table 4. Other assumptions, including message processing latencies of various nodes, are shown in Section A6. Note that the call setup latency could change if the message processing latencies of nodes and/or the backhaul (RNC-Node B) latency are different than what is assumed.

From Table 4, the average total delay incurred in a MO PS *call setup from idle* is **~ 1.7sec**, when signaling is carried on 3.4 kbps dedicated channels.

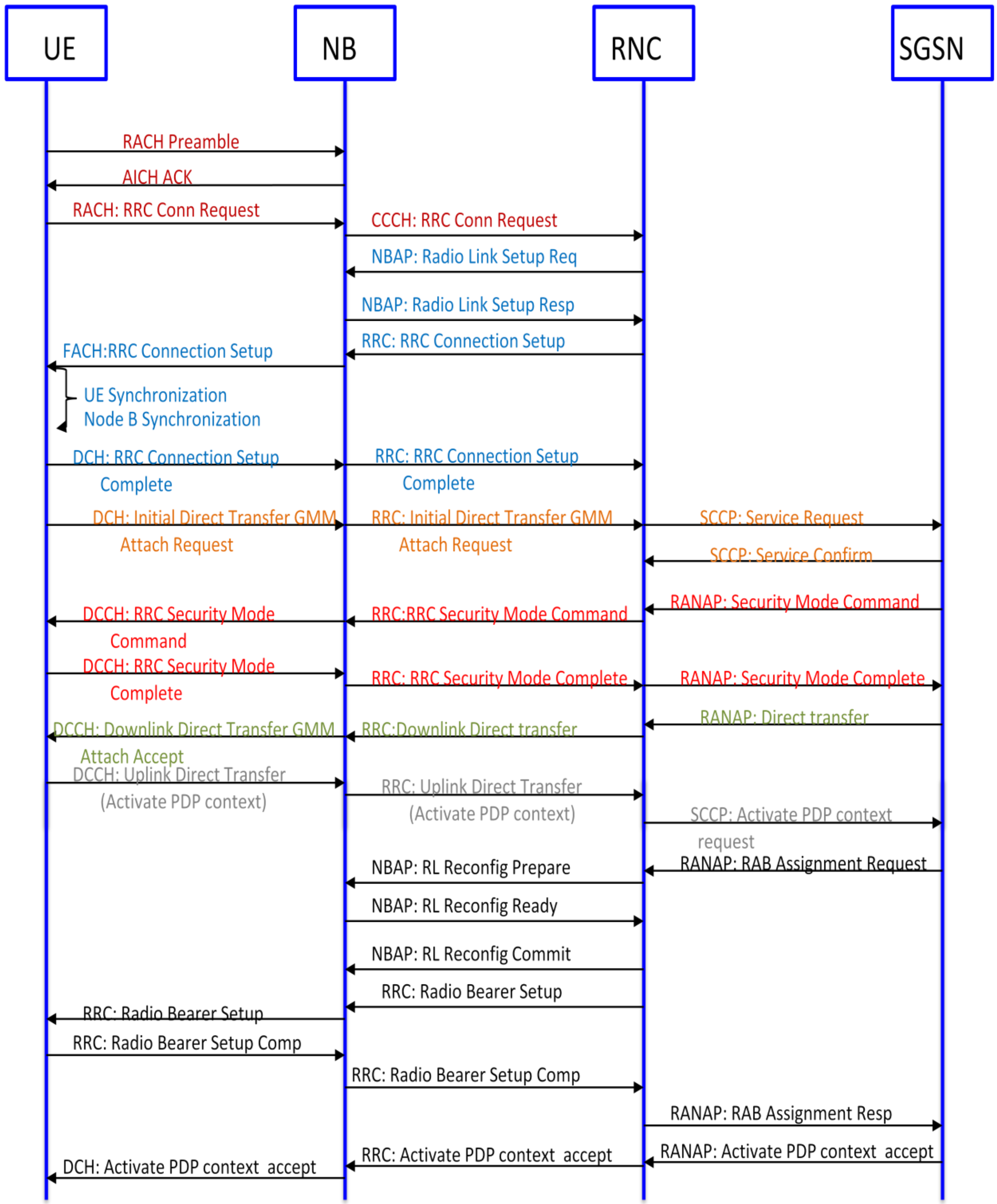

Figure 2: MO PS Call Flow for Call Setup from Idle for WCDMA

Process	Delay (ms)	Process Latency (ms)
1. Random Access		**72**
SIB 7 Uplink interference parameter delay (Assuming 80ms cycle for SIB7 broadcast)	40	
Preamble		
PRACH period	0.67	
PRACH transmit time	1.33	
PRACH processing at NodeB	4	
AICH Transmit time	2	
AICH response processing	2	
Data		
RRC connection request	10	
Node B Phy processing	2	
Node B - RNC Delay and RNC processing	10	
2. RRC Connection Setup Procedure		**174**
Radio Link Setup	17	
RRC Connection Setup DL Tx*	40	
RNC to NodeB delay for DL	5	
UE Physical layer and RRC layer processing delay for DL message	8	
NodeB Synchronization	20	
UE Packet preparation	4	
RRC Connection Setup Complete UL Tx*	80	
NodeB and RNC Processing and delay (Happens in parallel when UE is transmitting Attach Request)	0	
3. Initial Direct Transfer (GMM Attach Request)		**99**
GMM Attach Request UL Tx*	80	
NodeB, RNC and SGSN Processing	12	
NodeB, RNC and SGSN delay	7	
4. Security Mode Command		**191**
SGSN to RNC, RNC to NodeB delay	6	
UE Phy layer and RRC layer processing delay for DL msg	8	
RRC Security mode command DL Tx*	80	
UE Packet preparation	4	
RRC Security mode complete UL Tx*	80	
NodeB and RNC Processing	7	
NodeB to RNC, RNC to SGSN delay	6	
5. Downlink direct transfer (GMM Attach Accept)		**100**
RNC to NodeB delay for DL	5	
UE Physical layer and RRC layer processing delay for DL message	8	
GMM Attach Accept DL Tx *	80	
RNC to SGSN and SGSN to RNC Delay	2	
SGSN Processing	5	
6. Uplink Direct Transfer		**130**
UE Packet preparation	4	
Activate PDP Context Request UL Tx*	80	
NodeB Phy and RNC Processing	7	
NodeB, RNC and SGSN delay	7	
Radio Link Prepare/Ready/Commit	27	
SGSN Processing	5	
7. RAB Assignment		**951**
SGSN-RNC Delay	1	
Radio Bearer Setup DL Tx*	800	
UE Packet preparation	4	
Radio Bearer Setup Complete UL Tx*	40	
NodeB Phy and RNC Processing	7	
NodeB to RNC, RNC to SGSN delay	6	
RNC to NodeB delay for DL	5	
UE Physical layer and RRC processing delay for DL message	8	
Activate PDP context accept DL Tx*	80	
Total MO PS latency		**1717**

*Refer Section A3

Table 4: MO PS Latency for Call Setup from Idle when Signaling is Carried on 3.4 kbps Dedicated Channels

Mobile Terminated Call Setup from Idle

For Mobile Terminated (MT) *call setup from idle*, the call flow is similar to Figure 2, with the addition that the UE receives a paging message before it sends a RRC Connection Request message. The paging procedure adds an additional delay (shown in Table 5) for the MT call compared to the MO call. For MT PS *call setup from idle*, the average total delay is **~ 1.9** sec.

Process	Delay(ms)	Comments/Explainations
Expected Delay due to Paging Cycle of 320ms	160	Average delay is mean of paging cycle
GGSN processing	5	Processing delays across network elements
SGSN processing	5	
SGSN-RNC delay	1	
RNC processing	5	
RNC-Node B delay	5	
Node B processing	2	
Paging Delay	20	Delay in transmitting the paging indicator and paging channels on the downlink
Total Page delay	**203**	

Table 5 : Latency involved in Paging the UE

A3. WCDMA/HSPA Latency Improvements for Call Setup from Idle due to carrying Signaling on HSPA channels

In this section, we focus on latency of *call setup from idle* when signaling is carried on HSPA channels. Table 6 compares, for the various messages exchanged in the call flow, the air interface latency involved when carrying signaling over HSPA versus 3.4 kbps dedicated channels. The delay of the messages shown in Table 6 is obtained by dividing the size of the message by the channel rate, i.e., 3.4 kbps for dedicated channel. For signaling carried over HSPA channels, 2ms TTI is assumed for E-DCH. Other assumptions, including processing latencies of various nodes and average number of H-ARQ transmissions for HS-DSCH and E-DCH, are shown in Section A6. The following explains the latency numbers for two of the messages from Table 6 (the latency reduction for other messages follows a similar argument):

1. RRC Connection Setup (NodeB to UE):
   The size of RRC connection Setup is typically ~100 to 120 bytes. When sent over FACH with a SCCPCH channel rate of 30Kbps*, the transmission delay is 30ms to 40ms. When sent over FACH mapped to HS-DSCH channels (as in Enhanced CELL_FACH, which is a Release 7 feature), RRC Connection Setup can be transmitted using 2ms TTI. Selecting a conservative 320 bit payload to cover the cell edge (due to absence of CQI/ACK in CELL_FACH), the transmission delay is 6ms.
2. Radio Bearer Setup (NodeB to UE):
   The typical size of the Radio Bearer Setup message is ~120bytes. This requires ~300ms to be transmitted over a 3.4kbps dedicated channel. Moreover, since Radio Bearer Setup is typically a synchronized message (i.e., it includes an Activation Time), the Activation Time needs to account for RLC retransmission(s) as well as other network delays. Thus, typically a conservative activation time

* All messages except RRC connection setup are carried over the dedicated channel. The RRC connection setup message is carried over the FACH channel.

of 500ms is selected, which leads to a latency of ~800ms, when carried over dedicated channels. When transmitted over HS-DSCH channels, the transmission time of this message would be 3.2ms (assuming 1.1 average H-ARQ transmissions and 1 ms waiting time for TTI). Moreover, the Activation Time in this case does not need to be as conservative as that for dedicated channel, since a smaller time needs to be budgeted for RLC retransmission(s). Thus, the total latency for this message, assuming a 200ms Activation Time, is 203.2 ms. It may be noted here that WCDMA/HSPA also supports an Unsyncronized Radio Bearer Setup message (with Activation Time of "now"), which has the potential to further reduce the C-plane latency. Unsynchronized Radio Bearer Setup has not been considered in this paper.

No:	OTA message	Size	Air Interface delay when Signaling on Dedicated Channels (ms)	Air Interface delay when Signaling on HSPA (ms)
1	RRC Connection Setup	107	40	6
2	RRC Connection Setup Complete	25	80	10
3	Initial direct transfer – GMM Attach Request	21	80	10
4	Security mode command	25	80	3.2
5	Security mode complete	25	80	10
6	Downlink direct transfer – GMM Attach Accept	10	80	3.2
7	Uplink direct transfer- Activate PDP context	25	80	10
8	Radio Bearer Setup	126	800 (including Activation Time of 500 ms)	203.2 (including Activation Time of 200 ms)
9	Radio Bearer Setup Complete	10	40	10
10	Activate PDP context Accept	14	80	3.2
	Total Delay		**1440**	**269**

Table 6: Air Interface Latency comparison when carrying Signaling over Dedicated and HSPA channels

Table 7 shows a comparison of the latencies of *call setup from idle* for HSPA when signaling is carried on dedicated and HSPA channels. Refer to Table 10 for a detailed breakdown of HSPA latency numbers for a 546 ms MO *call setup from idle*. From Table 6 and Table 7, we can see that when signaling is carried on HSPA channels, the transmission of signaling messages over the air interface contributes a small percentage ((69 ms =269ms – activation time of 200ms) out of a total of 546 ms) of this latency and the rest is contributed by other factors such as backhaul, core network, Activation Time and SIB reading. MT call latency is obtained by adding the paging delay (203ms from Table 5) to the MO delay.

C-Plane Process	Signaling on Dedicated Channels (ms)	Signaling on HSPA Channels (ms)
MO Delay of Call Setup from Idle	1717	546
MT Delay of Call Setup from Idle	1920	749

Table 7: Comparison of Latency of Call Setup from Idle when Signaling is carried on Dedicated and HSPA channels

A4. HSPA Latency of Call Setup from Connected

For the case of *call setup from connected*, we assume that the UE camps in CELL_PCH state. The HSPA specifications support *legacy CELL_PCH* as well as *Enhanced CELL_PCH* (introduced in Rel 7). Enhanced CELL_PCH allows the UE to retain a dedicated H-RNTI in CELL_PCH state, which means that the UE is primed to send/receive data. In comparison, legacy CELL_PCH requires the UE to exchange a Cell Update/Cell Update Confirm with the network before uplink/downlink data can be exchanged. While Enhanced CELL_PCH offers latency benefits over legacy CELL_PCH, networks are only now starting to deploy legacy CELL_PCH. Thus, we analyze the call setup latency considering both Enhanced and legacy CELL_PCH. Note that apart from the case when the user starts a data session after a long inactivity period, the user will likely be in CELL_FACH state. From this state, the latency for *call setup from connected* is similar to that from the Enhanced CELL_PCH state (for similar DRX cycle in CELL_FACH.

Latency of Call Setup from Connected for Enhanced CELL_PCH

We assume that the state to which the UE transitions where it can send/receive data is CELL_FACH (specifically Downlink and Uplink Enhanced CELL_FACH). As mentioned before, Downlink and Uplink Enhanced CELL_FACH (defined in Rel 7 and Rel 8 respectively) allow the use of HS-DSCH and E-DCH channels, thus enabling high-speed data transfer in CELL_FACH state.

Figure 3 shows the procedure for an MO *call setup from connected* from Enhanced CELL_PCH to Enhanced Uplink CELL_FACH: the two steps required before data can be transmitted are: (a) sending RACH preamble and receiving acknowledgment on the Access Indicator Channel (AICH) and (b) the sending of a configurable number of pilot-only preambles. Though the minimum duration for which pilot-only preambles need to be sent is 2 TTIs (or 4 msec), for a more typical setting, the network may configure a slightly larger number (example 5) of pilot-only preambles. Thus, we expect the delay in this procedure to be in the range of 14-20 msec. A breakdown of this latency is shown in Table 8.

Note that as part of Uplink Enhanced CELL_FACH, a default uplink grant is carried in the SIBs, so that the UE can start transmitting without requiring the network to assign an uplink grant. Moreover, Uplink Enhanced CELL_FACH allows a default value of Uplink Interference to be sent in SIB5, and a UE supporting Uplink Enhanced CELL_FACH does not need to read the dynamic value of Uplink Interference from SIB7 before starting transmission.

For an MT *call setup from connected*, the only delay involved is in waiting for the "on" TTI in the paging cycle. Since the UE is in Enhanced CELL_PCH (with dedicated H-RNTI), it can directly receive data in the "on" TTIs during the paging cycle, without needing to go through the Cell Update procedure. The delay involved in this procedure on the average is half the length of the paging cycle, i.e., 160 msec (Table 5). Note that in Enhanced CELL_PCH, the UE still monitors the Paging Indicator Channel (PICH) to determine whether to decode HS-DSCH during the "on" TTIs. On receiving data, the UE autonomously transitions to CELL_FACH.

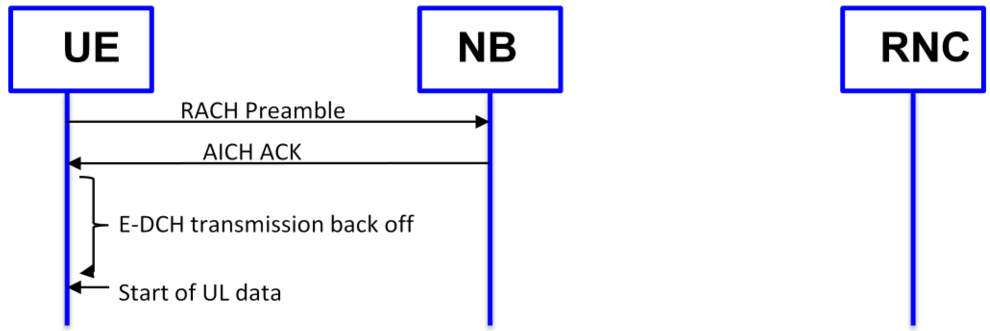

Figure 3: Call Setup from Connected for Enhanced Cell PCH callflow

Call Setup from Connected for Enhanced Cell PCH	Delay(ms)	Total Delay (ms)
PRACH period	0.67	14-20
PRACH transmit time	1.33	
PRACH processing at NodeB	4	
AICH Transmit time	2	
AICH response processing	2	
E-DCH Transmission backoff	4-10 (2-5 pilot premables)	

Table 8: Breakdown of Latency of MO Call Setup from Connected for Enhanced cell PCH

Latency of Call Setup from Connected for Legacy CELL_PCH

Comparing to camping in Enhanced CELL_PCH, camping in legacy CELL_PCH incurs the additional latency of the UE sending a Cell Update message to the RNC, and receiving a Cell Update Confirm from the RNC. This is expected to incur an additional latency of 68.2 msec (when signaling is carried on HS channels), bringing the *MO call setup from connected* to 82.2-88.2 msec and the *MT call setup from connected* to 248 msec (MT latency = MO latency + half the length of the paging cycle, i.e., 160ms). The breakdown of this additional latency is shown in Table 9.

Call Setup from Connected for Legacy Cell PCH	Delay(ms)	Total Delay (ms)
Latency for Enhanced Cell PCH	14-20	82.2 to 88.2
Packet preparation for Cell Update	4	
Cell Update Uplink Transmission	10	
NodeB Phy Processing	2	
NodeB to RNC delay	5	
RNC processing delay	5	
RNC to NodeB delay for DL	5	
Cell Update confirm DL Tx	3.2	
UE Physical layer processing delay for decoding DL message	3	
UE RRC processing delay for DL message	5	
UE Packet preparation for PCRC	4	
PCRC Uplink Transmission	10	
NodeB Phy Processing	2	
NodeB to RNC delay	5	
RNC processing delay	5	

Table 9: Breakdown of Latency of MO Call Setup from Connected for Legacy Cell PCH

A5. Latency Comparison of HSPA and LTE

A5.1 Latency Comparison of Call Setup from Idle

In this section, we show the call flow of MO *call setup from idle* for an LTE system (Figure 4). We identify some key differences between the call setup procedures for LTE and HSPA. We also compare the latency of *call setup from idle* of an HSPA system (where signaling is carried on HSPA channels) with that of an LTE system. For the comparison, we assume similar H-ARQ operating points and similar processing latencies of network nodes for HSPA and LTE.

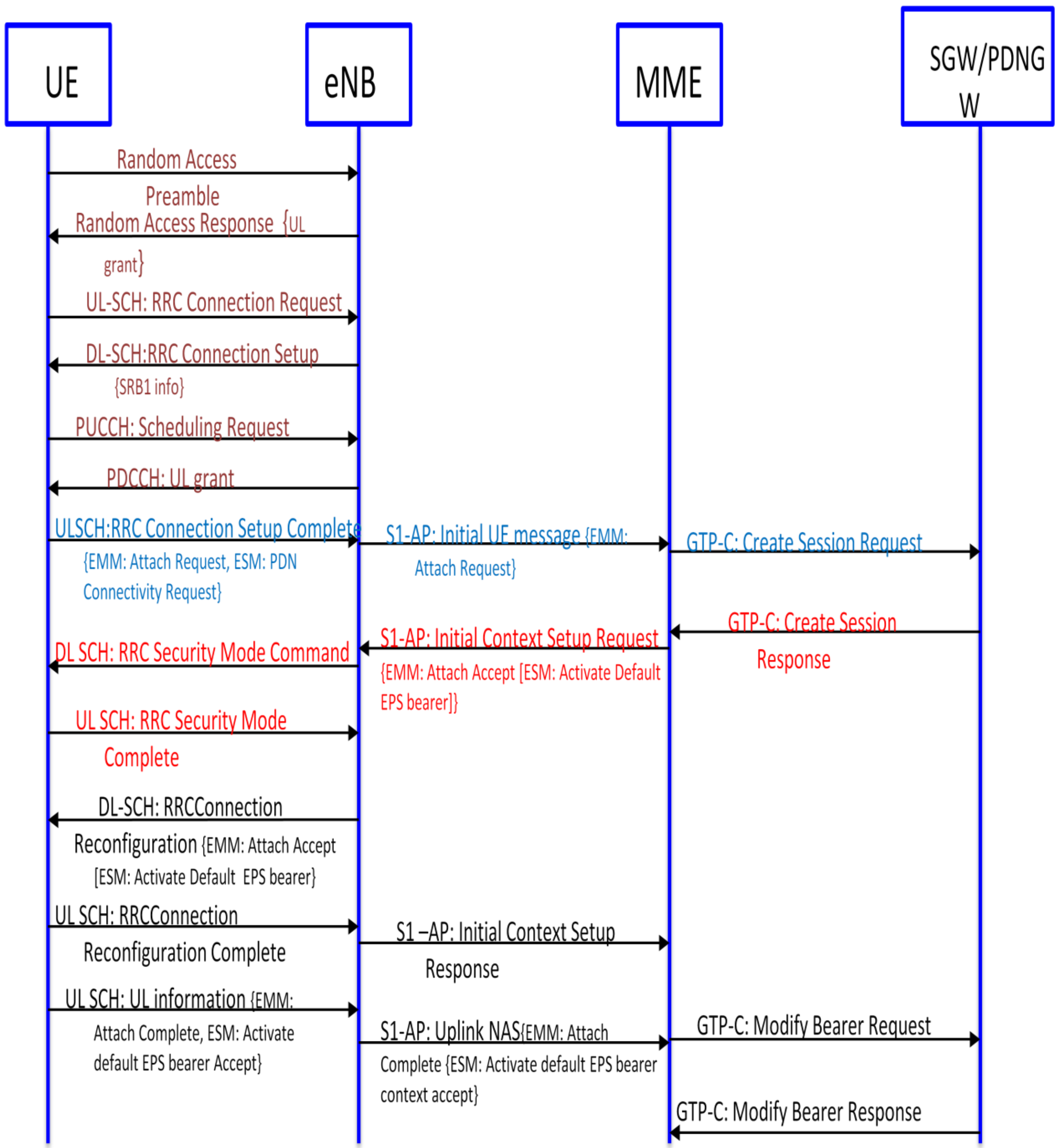

Figure 4: Call flow for MO Call Setup from Idle for LTE

Comparing Figure 4 (LTE) with Figure 2 (HSPA), the key differences are:

1. In LTE, the Attach Request (a NAS PDU) is piggybacked on the RRC Connection Setup Complete message. In HSPA, the Attach Request is sent seperately after RRC Connection Setup Complete.
2. In LTE, the UE needs to get an uplink grant before it can send the RRC Connection Setup Complete message.
3. The Activate PDP Context message is not sent in LTE since an IP address is assigned to the UE in response to Attach Request (this is possible due to LTE being a PS only system). In addition, a default radio bearer (default EPS bearer) is setup in response to the Attach Request and the UE can start transmitting data after sending RRC Connection Reconfiguration Complete. However, the MME needs to wait for the Attach Complete message sent by the UE to arrive before it can start forwarding downlink data.
4. The RRC Connection Reconfiguration Complete message in LTE does not have an Activation Time, while the corresponding message (i.e., Radio Bearer Setup) in HSPA is typically sent with an Activation Time.

MO Call

HSPA Procedure	HSPA delay (ms)	LTE Procedure	LTE delay (ms)
--SIB reading time	40	--SIB reading time	40
--Random Access Preamble	11	--Random Access Preamble	9
--RRC connection request --RRC connection Setup	78	--RRC connection request --RRC connection Setup --Scheduling Request --UL grant	19.8
--RRC connection setup complete --GMM Attach Request --Activate PDP context Request	63	--RRC Connection Setup Complete (Attach Request) -- S1 connection establishment	37.9
--Security Mode Command	44.2	--Security Mode Command	32.7
--GMM Attach Accept --Radio Bearer Setup --Activate PDP context Accept	311.4	--Attach Accept	48.7
Total	**546**		**188**

Table 10: Comparison of MO Call Setup Latency from Idle between HSPA and LTE

Table 10 shows that LTE reduces the MO call setup latency from idle by ~360 ms compared to HSPA (when signaling is carried on HSPA channels).

MT Call

Procedure	HSPA (ms)	LTE (ms)
MO latency (fromTable 10)	546	188
Paging latency (from Table 5 for HSPA and Table 15 for LTE)	203	177
Total MT latency	**749**	**365**

Table 11: Comparison of MT Call Setup Latency from idle between HSPA and LTE

For MT *call setup from idle*, the UE needs to be paged in addition to the procedures detailed in Table 10. Table 11 shows that LTE has a lower MT call setup latency from idle of ~385ms compared to HSPA.

The following are the key reasons for the lower call setup latency for LTE compared to HSPA:

1. The use of synchronized Radio Bearer Setup for HSPA, while the corresponding message (RRC Connection Reconfiguration) for LTE is unsynchronized. This leads to an additional latency of ~200 ms for HSPA. Note that, unsynchronized Radio Bearer Setup is also supported in the specifications for HSPA, but is not considered in this paper due to the fact that most of the current deployments use synchronized setup.
2. NodeB power control synchronization for HSPA after entering CELL_DCH: this leads to an additional latency of 20ms for HSPA.
3. Messages exchanged between NodeB and RNC for radio link setup for HSPA. This leads to an additional latency of ~50ms for HSPA.
4. All RRC signaling messages experience the backhaul delay for HSPA, whereas in LTE, RRC messages terminate at the eNodeB. This leads to an additional latency of ~50ms for HSPA.

A5.2 Latency Comparison of Call Setup from Connected

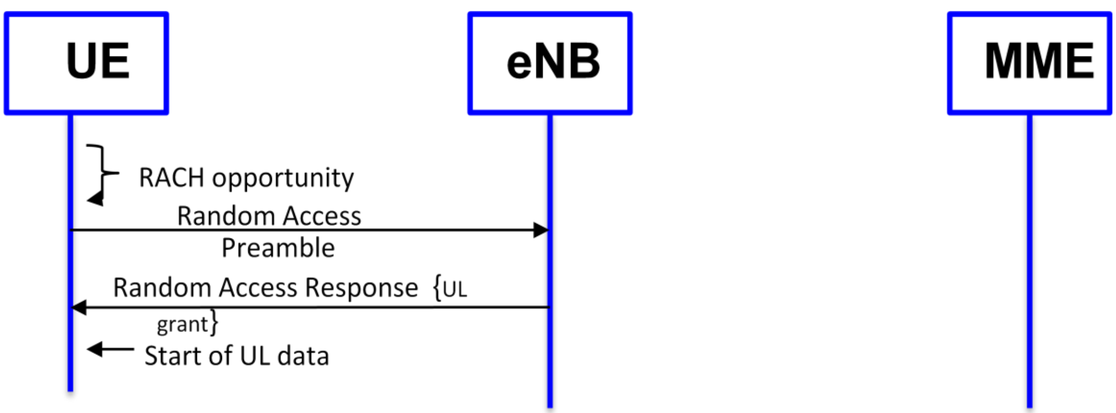

Figure 5: MO Call Setup from Connected for LTE

Figure 5 shows the procedure for *MO call setup from connected* for LTE, i.e, the steps required to reach a state where data can be transmitted starting from RRC connected with long DRX cycle. The delay in this procedure (as shown inTable 12) is 13 ms.

For MT *call setup from connected*, the only delay involved is in waiting for the "on" TTI in the DRX cycle. The delay involved in this procedure on the average is half the length of the DRX cycle, i.e., 160 msec (since a DRX cycle of 320 msec is assumed in this paper).

Call Setup from Connected	Delay (ms)	Total Delay (ms)
PRACH period, transmit time and processing time	4	13
RA response scheduling delay and transmission (ms)	2	
RA response processing	3	
Waiting for PRACH opportunity	4	

Table 12: Latency of Call Setup from Connected for LTE

Comparing the latency of *call setup from connected* for HSPA and LTE (i.e., Sections A4and A5.2 respectively), we find that:

- the latency for HSPA and LTE is similar when the starting state in HSPA is Enhanced CELL_PCH, and
- the latency for HSPA is ~69-75 ms larger than LTE when the starting state in HSPA is legacy CELL_PCH.

A5.3 User Plane Latency comparison

In this section, we evaluate the U-plane latencies for HSPA and LTE for a 32 byte PING packet. We assume similar H-ARQ operating points for HSPA and LTE (1.1 average number of transmissions for downlink and 1.5 average number of transmissions for uplink) and the use of scheduled grants for HSUPA (note that with the use of non-scheduled grants, the U-plane delay of HSPA can be further reduced). From Table 13, we see that, with similar H-ARQ and backhaul assumptions, the U-plane latencies for HSPA and LTE are similar. We explain some of the latencies from Table 13 below:

1. *UL Resource Assignment* latency is the latency involved in assignment and transmission of grants for transmission of data on the uplink.
2. *UL Data Transmission* latency is the latency involved in packet preparation, waiting for TTI boundary and transmission of data on the uplink.
3. *Network Node Processing* latency is the latency involved in processing at the RNC and the GGSN for HSPA (note that the 1-tunnel architecture standardized in Rel 7 allows the SGSN to be bypassed for user plane traffic) and the SGW and the PGW for LTE.
4. *DL Transmission* latency is the latency involved in packet preparation, waiting for TTI boundary and transmission of data on the downlink.

Procedure	HSPA delay (ms)	LTE delay (ms)
UL Resource Assignment	8	8
UL Data Transmission	8	4
Node B Decoding	3	3
Backhaul delay	5	5
Network node processing (RNC/GGSN for HSPA and SGW/PGW for LTE)	2	2
Backhaul	5	5
DL Transmission	3	2
UE Decoding and processing	4	3
Total U-Plane Latency	**38**	**32**

Table 13: U-Plane Latency Comparison between HSPA system and LTE system

A6. System Assumptions

Table 14, Table 15 and Table 16 show processing delays and other delay assumptions that have been used in the analysis in this paper. Some of these assumptions are obtained from base station/RNC vendors and others are reasonable estimates.

HSPA System Assumptions	ms
RNC processing (ms)	5
Node B Processing (ms)	2
RNC-Node B Delay (ms)	5
SGSN-RNC Delay (ms)	1
SGSN processing (ms)	5
GGSN processing (ms)	5
UE RRC Processing (ms)	5
NB Scheduling Delay (ms)	1
NB Phy Processing (ms)	2
Paging Cycle (ms)	320
Paging Delay (ms)	20
PRACH period (ms)	1.33
PRACH Tx time (ms)	12
PRACH processing NB (ms)	4
AICH Tx time (ms)	2
AICH response processing (ms)	2

HSPA System Assumptions	ms
RACH UL data Tx (ms)	10
DL Synchronization (ms)	0
UL Synchronization (ms)	20
UE Packet Preparation (ms)	4
UE Phy processing (ms)	3
Avg Number UL H-ARQ for E-DCH	1.5
Avg Number DL H-ARQ for HS-PDSCH	1.1
Downlink FACH Transmission	
Bits per transmission	320
Number of repeats	3
Message Size (bytes)	100
Radio Link Setup	
RNC processing (ms)	5
RNC-Node B Delay (ms)	5
Node B Processing (ms)	2
Node B –RNC Delay (ms)	5

Table 14: HSPA System Assumptions

LTE System Assumptions		ms
Paging		**177**
Half of the Paging Cycle (ms)	160	
SGW processing (ms)	5	
S11 delay (ms)	1	
MME processing (ms)	5	
DL S1-MME delay (ms)	5	
eNB processing (ms)	1	
DL Data Transmission		**10.8**
eNB scheduling delay (ms)	1	
DL Data Tx time (ms)	1.8	
UE Phy processing (ms)	3	
UE RRC processing (ms)	5	

LTE System Assumptions		ms
UL Data Transmission		**20.9**
SR period	5	
SR delay (ms)	2.5	
SR tx (ms)	1	
SR processing (ms)	2	
UL scheduling delay (ms)	1	
UL grant Tx (ms)	1	
UL data Tx delay (ms)	3	
UL data Tx (ms)	3.4	
eNB Phy processing (ms)	2	
eNB RRC processing (ms)	5	

Table 15: LTE System Assumptions

Process	Delay (ms)	Process Latency (ms)	Process	Delay (ms)	Process Latency (ms)
1. Random Access		**69.8**	**3. Security Mode Command**		**32.7**
SIB 2 reading time (Assuming 80ms cycle for SIB2 broadcast)	40		NAS security mode command**	10.8	
Preamble Total			NAS security mode command complete*	20.9	
PRACH period, transmit time and processing time	4		eNodeB processing, MME processing	6	
RA response scheduling delay and transmission (ms)	2		UL S1-MME delay, DL S1 - MME delay	10	
RA response processing	3		eNodeB processing	1	
			RRC security mode command**	10.8	
Data Total			RRC security mode command complete*	20.9	
UE transmission delay, transmission time	2				
RRC connection request*	1		**4. Attach Accept**		**48.7**
eNodeB PHY and RRC processing	7		*RRC reconfiguration*		
RRC connection setup**	10.8		RRC connection reconfiguration**	10.8	
			Attach Complete*	20.9	
2. RRC Connection Setup Complete/Attach Request		**37.9**	*Update Bearer*		
RRC Connection Setup Complete*	20.9		eNodeB and MME processing	6	
S1 establishment			UL S1-MME delay, DL S1 - MME delay	10	
eNodeB and MME processing	6		eNodeB processing	1	
UL S1-MME delay, DL S1 - MME delay	10				
eNodeB processing	1		**Total MO delay for LTE**		**188.1**

Table 16: LTE Idle Call Setup MO Latency breakdown

References

[1] 3G Americas – *"Transition to 4G, 3GPP Broadband Evolution to IMT-Advanced"*- http://www.ngmn.org/fileadmin/user_upload/News/Partner_News/3G_Americas_RysavyResearch_HSPA-LTE_Advanced_FINAL.pdf

[2] 3GPP TS 25.214, "*Physical Layer Procedures (FDD)*".

[3] 3GPP TS 36.214, "*Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements*".

[4] 3GPP TS 25.331, "*Radio Resource Control (FDD)*".

[5] 3GPP TS 36.331, "*Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification*".

*Refer Table 15 DL data, **Refer Table 15 UL Data