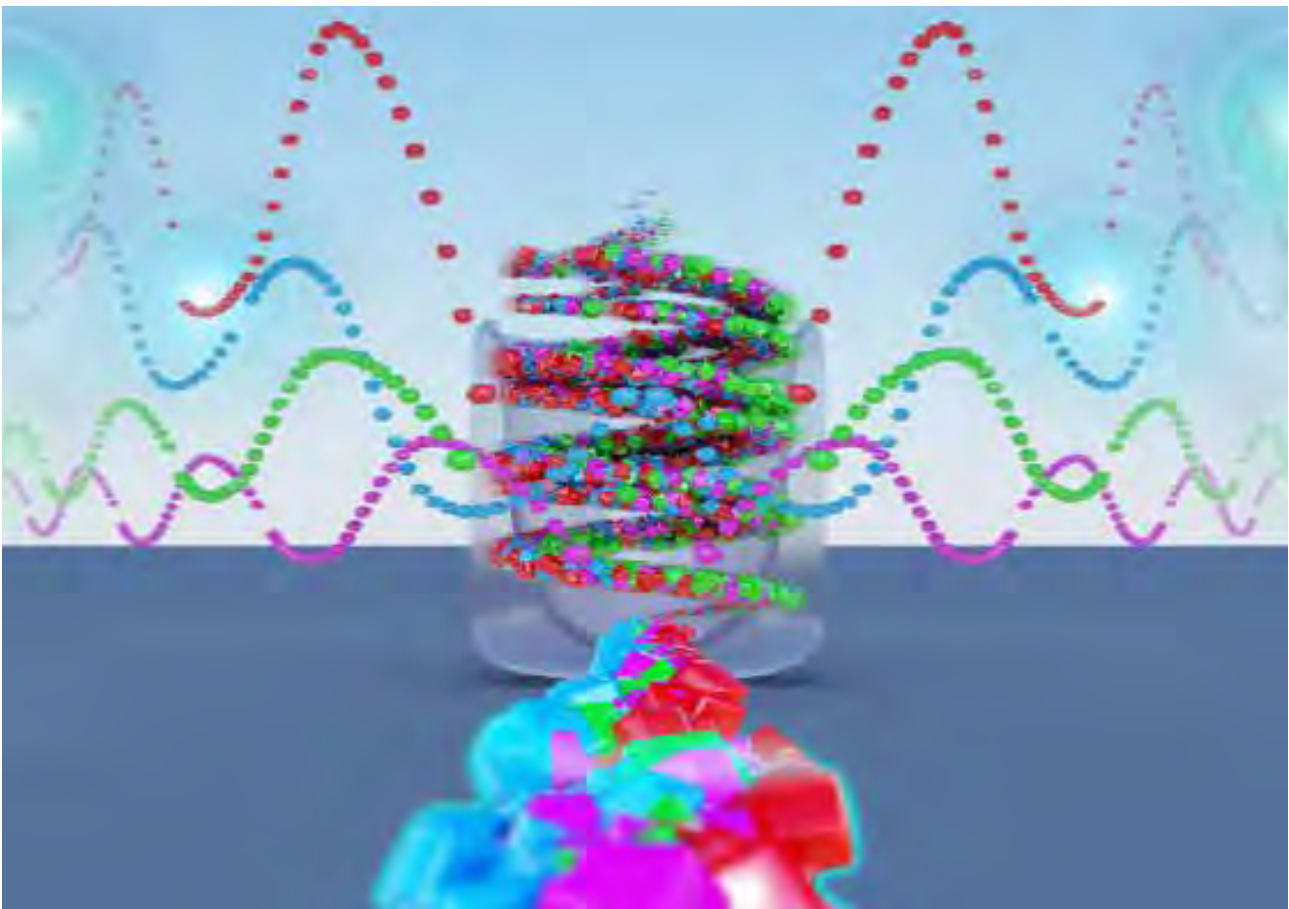


LTE MAC Scheduler & Radio Bearer QoS



Abstract

Long Term Evolution (LTE) uses a packet-switched, all-IP architecture which takes advantage of the latest packet switching and wireless technologies to deliver increased capacity and reduced cost-per-bit. The lack of a circuit-switched domain means that voice and other real-time services are multiplexed with non-real time services over the air interface, radio access and core networks. The full benefits of LTE will only be realised by networks which dynamically accommodate a mix of real and non-real time services. An end-to-end, class-based Quality-of-Service (QoS) architecture has been defined for LTE in order to support such a mix of services.

The air interface and backhaul are likely to remain key constraints in future LTE networks, making the base station a cornerstone of any end-to-end QoS mechanism. The air interface presents a particular challenge for QoS given the time-varying nature of the wireless channel and differing user device capabilities. This paper considers the challenges facing Radio Bearer QoS in LTE and how Roke Manor Research can help network equipment providers and operators realise the benefits via a QoS-aware MAC Scheduler.

Introduction

Long Term Evolution (LTE), as defined in 3GPP Release 8, is the next step in the evolution of cellular mobile communication towards 4G. It will provide the additional capacity and lower cost-per-bit needed to sustain the exponential growth in mobile data witnessed over the last few years.

LTE uses an all-IP, purely packet-switched network architecture to support all services. This is in contrast to GSM/GPRS and 3G/HSPA networks which incorporate both circuit-switched and packet-switched domains in order to support different service types – voice, video, web browsing, mobile apps, etc. The resulting LTE network architecture is flatter and simpler than its 2G and 3G predecessors, with all user-plane traffic multiplexed over a common, packet-switched network.

Early LTE networks will support a best-effort mobile data service, similar to the mobile broadband currently provided by 3G/HSPA. This will enable mobile network operators to continue to grow data-driven revenues. However, such a model does little to recognise the true value and cost associated with mobile data.

In contrast, LTE networks with end-to-end QoS enable operators to:

1. Distinguish and prioritise value-added services over basic access. Revenue per bit (\$/bit) for services such as voice is many times that for basic mobile broadband access.
2. Differentiate access, providing incremental revenue opportunities. Tiered access plans provide customers with a choice for their mobile broadband service, and enables operators to monetise the heavy users.
3. Support the reliable delivery of value-added, real-time services such as voice. Without guaranteed QoS, voice and other real-time services would be subject to the uncertainties affecting current 3G/HSPA mobile broadband.

As well as enabling new (revenue generating) consumer services, end-to-end QoS is essential for a range of other applications looking to benefit from the capacity and cost benefits offered by LTE. These include:

- Public safety – wideband successor to TETRA.
- Railways – potential successor to GSM-R.
- Machine-to-machine applications, e.g. Smart Grids.

Both the public safety and railway domains have strict requirements for reliable, secure voice communication.

One approach to supporting a diverse mix of real and non-real time services with the appropriate end-to-end QoS is to over-provision the LTE network to avoid congestion completely. However, over-provisioning the entire network is rarely practical, economical or even possible. As a result, bottlenecks in the network are inevitable.

“ LTE network bottlenecks inevitable ”

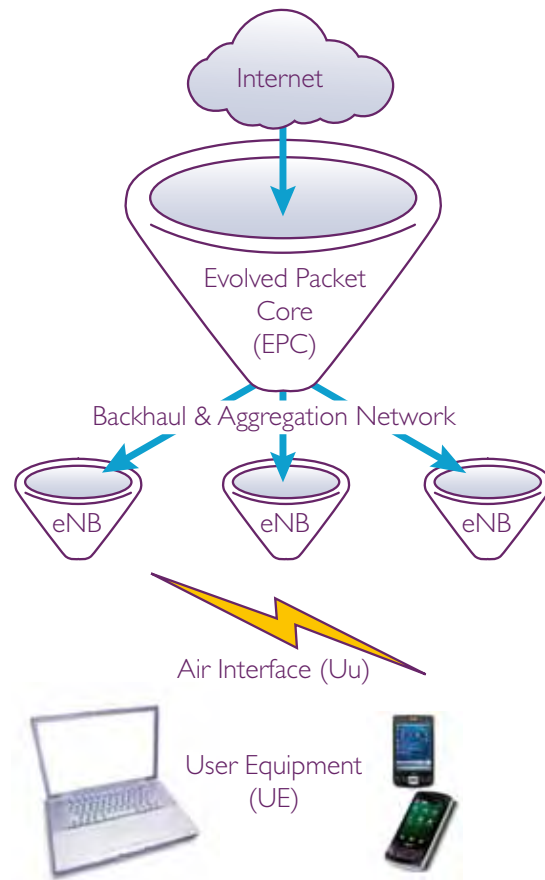


Figure 1: LTE Downlink Bottlenecks

Figure 1 depicts the end-to-end data flow in an LTE network, from Internet to user. This figure highlights two potential bottlenecks:

- The backhaul and aggregation networks which support the S1 interface between the Evolved Packet Core (EPC) and base station (eNodeB/eNB). Backhaul limitations are particularly acute in existing 3G/HSPA networks. Suitable provisioning of “Greenfield” LTE networks, e.g. with fibre to the cell site, should result in fewer backhaul bottlenecks.
- The air or Uu interface between the eNodeB and User Equipment (UE). With limited bandwidth available, interference from adjacent cells and the time-varying nature of a NLOS wireless channel, the air interface will always be a constrained resource for LTE.

It is assumed that a combination of provisioning and traffic shaping will prevent congestion in the EPC. Deep Packet Inspection based traffic shaping is already used extensively in 3G/HSPA core networks in support of “fair use” policies.

“ QoS needs support at the edge ”

Roke Manor Research (Roke) is supporting network equipment providers and operators in realising QoS-enabled LTE networks and services. Roke’s MAC scheduling technology and expertise supports the provision of QoS over the radio link between network and user.

QoS

Quality of Service (QoS) involves the delivery of data whilst meeting a combination of latency, jitter, error rate and maximum/guaranteed bit rate requirements.

As mentioned earlier, traffic shaping in the EPC is an efficient solution to core network congestion and for enforcing “fair use” policies. However, traffic shaping in the core can be rather a blunt instrument in terms of end-to-end QoS as it lacks information about the network edge, the air interface and backhaul network, which are often the main bottlenecks. Traffic shaping functions therefore need to be included in the network edge in order to support the QoS benefits identified.

QoS-aware scheduling over the LTE air interface is integral to end-to-end QoS and is provided by the eNodeB MAC Scheduler.

QoS For VoIP

Voice over IP (VoIP) is a good example of a value-added, real-time service which requires QoS support. Voice quality can be measured using the Mean Opinion Score (MOS) which assigns a metric between 1 (lowest quality) and 5 (highest quality), see Table 1.

Figure 2 shows the effect of packet loss and end-to-end delay (latency) on voice quality for the AMR codec typically used in 3G. It can be seen that voice quality falls off rapidly for both delay and packet error rate. Assuming we want to support toll quality voice, we need to control the packet loss and delay in order to achieve a MOS≈4. Figure 2 indicates that, for the AMR codec, a maximum end-to-end delay of approximately 200ms and packet error rate of 1% (10^{-2}) should achieve toll or near-toll quality voice.

NOTE: LTE actually specifies a maximum delay of 100ms for conversational voice. This only applies to the delivery of IP packets in the LTE network. The AMR codec will introduce additional delay.

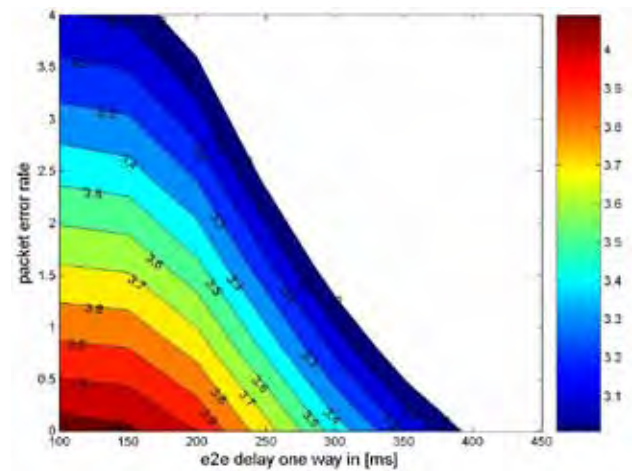


Figure 2: Mean Opinion Score (MOS) for NB-AMR

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good / Toll Quality	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Table 1: Mean Opinion Score (MOS) Definitions

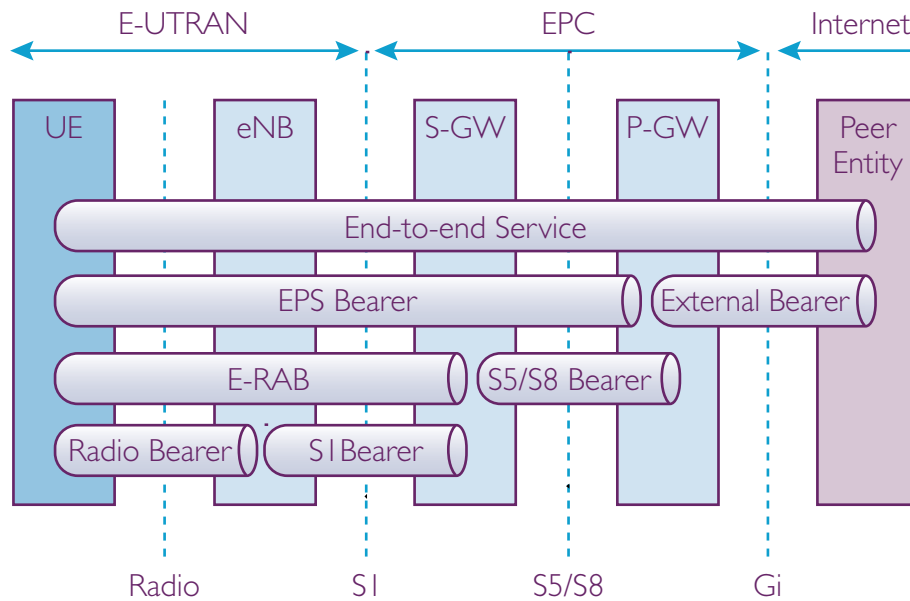


Figure 3: LTE Bearer Architecture. Source: 3GPP TS 36.300

QoS in LTE

An end-to-end, class-based QoS architecture has been defined for LTE. This mechanism is based on the concept of data flows and bearers, as shown in Figure 3. Data flows are mapped to bearers, with three individual bearers (Radio, SI and S5/S8) combining to provide the end-to-end QoS support via the EPS Bearer.

NOTE: Full end-to-end QoS requires External Bearer support, however this is beyond the scope of the LTE standards.

The class-based QoS approach adopted for LTE is significantly simpler than the QoS mechanisms defined for 3G/HSPA. It enables bearers to be mapped to a limited number of discrete classes. Network nodes can be pre-configured for these classes, limiting the amount of QoS information needing to be dynamically signalled.

Table 2 summarizes the QoS information for each EPS Bearer. This needs to be implemented by each constituent bearer (Radio, SI and S5/S8) which combine to form the EPS Bearer.

LTE also supports QoS for EPS Bearer aggregates, at both a user and a network level, as detailed in Table 3. The UE-AMBR is particularly useful as it prevents a UE from monopolising cell capacity by setting up multiple non-GBR bearers.

Parameter	Description
QoS Class Identifier (QCI)	Scalar which indicates a specific priority, maximum delay and packet error rate – all of which are pre-configured in the network node. The index also indicates whether the bearer has a Guaranteed Bit Rate (GBR) or not (non-GBR). The actual bit rate is signalled separately.
Allocation and Retention Policy (ARP)	Used in prioritisation and pre-emption decisions with respect to bearers.
Guaranteed Bit Rate (GBR)	Bit rate that can be expected to be provided by a bearer. Not applicable for non-GBR bearers.
Maximum Bit Rate (MBR)	In 3GPP Release 8, MBR = GBR.

Table 2: EPS Bearer QoS Profile

Parameter	Description
APN-AMBR	Aggregate maximum bit rate across all Non-GBR bearers and across all PDN connections of the same APN. Enforced by P-GW for downlink.
UE-AMBR	Aggregate maximum bit rate across all Non-GBR bearers of a UE. Enforced by eNodeB for both uplink and downlink.

Table 3: Aggregate EPS Bearer QoS Parameters

Table 4 shows the nine standardised QCI profiles defined by 3GPP [1]. These are targeted at common services types – conversational voice and video, streaming video, gaming, IMS signalling and differentiated access. Additional QCI profiles may be supported by network equipment providers and used by operators.

QCI	Resource Type	Priority	Packet Delay Budget	Packet Error Loss Rate	Example Services	
1	GBR	2	100 ms	10^{-2}	Conversational Voice	
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)	
3		3	50 ms	10^{-3}	Real Time Gaming	
4		5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)	
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling	
6		6	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc).	
7		7	100 ms	10^{-3}	Voice, Video (Live Streaming) Interactive Gaming	
8		8	8	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc).
9			9			

Table 4: Standardised QCI Characteristics [source TS 23.203]

LTE MAC Scheduler & Radio Bearer QoS

The air interface, as highlighted in the Introduction, is likely to be one of main bottlenecks and contributors to delay, jitter and packet loss in the LTE network. Therefore, Radio Bearer QoS is key to achieving end-to-end QoS in LTE. The MAC Scheduler [2] within the eNodeB is responsible for scheduling the LTE air interface and supporting Radio Bearer QoS.

GBR bearers are provisioned, in the sense that their bandwidth requirements are checked against the current cell utilisation before allowing, or disallowing, the connection to be formed. This can include the case where the acceptance of a higher priority user can cause the call of a lower priority user to be pre-empted.

Non-GBR bearers have no guaranteed allocation of resources and hence provide a best-effort service. Through the priority mechanism and latency requirements indicated by QCI, some of these best-effort services are more equal than others.

The allocation of QCI to Radio Bearer is signalled to the eNodeB in the (S1) Bearer Setup messages. In the MAC Scheduler, the QCI is mapped to pre-configured information about priority, packet delay and packet loss.

NOTE: The numbers given in Table 4 are for the EPS Bearer. The values configured in the MAC Scheduler are for the Radio Bearer and chosen so that the Radio Bearer contributes to the overall required QoS.

A QoS-aware MAC Scheduler aims to distribute the available air interface resources to the UEs within the cell such that their (competing) bearer-level QoS requirements are met. The MAC Scheduler must also attempt to make the most efficient use of the air interface to maximise the cell and system capacity.

In LTE, data transmissions are scheduled over the air interface every subframe or TTI (1ms). This granularity of scheduling is intended to facilitate low-latency data transfer. A more detailed description of the lower-level MAC resource scheduling can be found in a separate Roke white paper "LTE MAC Scheduler & Radio Resource Scheduling" [3].

Scheduling

Downlink scheduling is performed solely within the eNodeB – the MAC sublayer decides what data is to be sent (the scheduling part) and sends it along with allocation information which enables the intended UE to receive the data.

Uplink scheduling results in grants being signalled to UEs, indicating what resources the UE can use to transfer data to the eNodeB. The UE is responsible for determining which data is transmitted within the granted resources. Therefore, a resource allocation granted in response to a request caused by one bearer may be used to transfer data from another higher priority bearer, belonging to the same UE, that has subsequently become available for transmission.

QoS-Aware Scheduling

A QoS-aware MAC Scheduler enables an eNodeB to support a larger number of users with an appropriate level of service as opposed to a best-effort scheduler which can either (a) support a smaller number of users reliably for the required levels of service or (b) support the same number of users but without any reliable level of service. A QoS-aware MAC Scheduler therefore contributes to the twin objectives of maximising network resource utilisation and maximising user Quality of Experience (QoE).

Figure 4 illustrates the operation of a QoS-aware MAC Scheduler in the downlink direction. Data for multiple Radio Bearers is queued in the RLC sublayer, with the MAC Scheduler receiving Buffer Status updates as new data arrives. The MAC Scheduler determines the downlink allocation of resources for each subframe. These allocations are signalled to the MAC sublayer which constructs the UE-specific Transport Blocks (TB), where each TB contains data from one or more Radio Bearers.

NOTE: UEs may have multiple active data Radio Bearers, as shown for UE #2 in Figure 4.

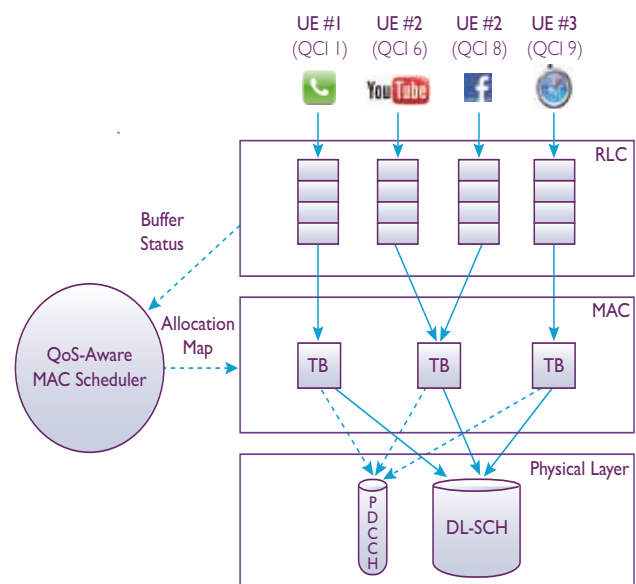


Figure 4: LTE Downlink QoS-Aware Scheduling

Whilst GBR and non-GBR bearers share common QoS requirements such as delay bounds and packet loss rate, the Scheduler needs to consider:

- GBR bearers need to be provided with a guaranteed throughput upto a committed guaranteed bit rate, irrespective of the radio channel conditions for the UEs concerned.
- Non-GBR bearers typically need to be provided with a service which meets certain fairness criteria. This is described in more detail below. Additionally, UE-AMBR needs to be enforced across multiple non-GBR bearers for the same UE.
- The balance of GBR and non-GBR traffic in each subframe, particularly for bursty traffic being carried by a GBR bearer.

It is generally assumed that the MAC Scheduler should be work conserving, i.e. the MAC Scheduler should continue to allocate resources as long as data is available to be sent, until all resources are used (or there is no more data to send). How this should be applied for excess GBR traffic is not clear: Should excess GBR traffic be dropped? Should it be prioritised over non-GBR traffic? One possible approach would be to:

1. Schedule the prioritised GBR traffic.
2. Schedule the prioritised non-GBR traffic, using the chosen fairness criteria.
3. Schedule the remaining non-GBR traffic.
4. Schedule the excess GBR traffic.

The QoS-aware MAC Scheduler would follow steps 1-4 each subframe until all data was scheduled or all resources were used.

Fairness is an important criteria for non-GBR bearers as it has a significant impact on both QoE and system capacity. A number of well established approaches to fairness may be used:

- Max-Min Fair: Aims to maximise the minimum rate for all bearers and therefore provides the “fairest” allocation of the three approaches given here. May be implemented using a round-robin algorithm.
- Proportionally Fair: Aims to maximise the overall rate for all bearers using a logarithmic utility function. This favours UEs with better channel conditions which are able to transfer data with a higher spectral efficiency. However, it still ensures that UEs with poor channel conditions, e.g. at the cell edge, still get some service.

Proportionally Fair scheduling strikes a balance between maximising system capacity and providing QoE and as such is widely used in 3G/HSPA. May be implemented using a WFQ algorithm.

- Max Throughput: Aims to maximise the overall rate for all bearers. This approach is skewed completely towards the most spectrally efficient UEs. Therefore, it provides the highest system capacity but poor QoE for cell edge users. May be implemented using a strict priority algorithm.

As indicated above, the choice of scheduling algorithm to use will depend on the fairness criteria sought. Selection of both the fairness criteria and scheduling algorithm is left as an implementation decision, i.e. it is not specified in the 3GPP. The following points should be considered when selecting scheduling algorithms:

- How the algorithm performs in terms of maximising the utility function (i.e. achieving the optimal allocation) for the chosen fairness criteria.
- How computationally efficient the algorithm is. NOTE: First come, first served (FCFS) and strict priority algorithms may be very efficient but poor in terms of yielding an optimal resource allocation.

Fairness-based utility functions alone will not provide QoS for non-GBR Radio Bearers. A wider view of utility is needed in order to accommodate latency. For example, an optimally fair scheduling algorithm may provide the required bearer throughput, but if the latency/jitter is too high for the required service/application then utility is reduced.

Uplink QoS-aware scheduling presents a number of additional challenges. Firstly, the MAC Scheduler has incomplete knowledge of the data available at the UE. Secondly, the MAC Scheduler has reduced control over what data gets transferred over the uplink air interface. Thirdly, the MAC Scheduler has to factor in the additional latency between allocating uplink resources and data being transferred. Given these additional challenges, it is tempting to focus on downlink QoS. However, uplink QoS is just as, if not more important as the uplink air interface represents a limited and therefore vital resource.

“Scheduling algorithms not standardised”

“Uplink QoS presents additional challenges”

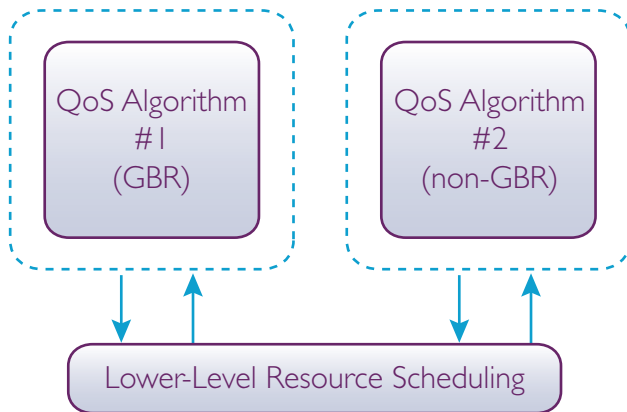


Figure 5: Swappable QoS Scheduling Algorithms

A “one size fits all” implementation of a QoS-aware MAC Scheduler is unlikely to be optimal for all scenarios. Ideally, the MAC Scheduler should support the use of multiple scheduling algorithms which can be adapted or swapped according to the given deployment scenario.

Figure 5 illustrates a QoS-aware MAC Scheduler with separate scheduling algorithms for GBR and non-GBR bearers. A common interface is employed, enabling different scheduling implementations to be inserted.

Summary

LTE has adopted an all-IP, packet-switched architecture which enables the full range of mobile services to be supported more efficiently. LTE networks are expected to offer a best-effort mobile data service initially, with operators applying “fair use” policies similar to those used today for 3G/HSPA. These best-effort services will enable operators to grow their data-driven revenues. However, they do little to recognize and, more importantly, monetize the true value associated with mobile data.

QoS-enabled LTE networks allow operators to provide differentiated access and reliable, real-time services such as voice. End-to-end QoS requires support at the edge of the network, in particular the Radio Bearer between the mobile user and the LTE eNodeB.

The eNodeB MAC Scheduler is responsible for providing Radio Bearer QoS over the LTE air interface. A QoS-aware MAC Scheduler should support GBR bearers (e.g. for voice and conversational video) and non-GBR bearers (e.g. for basic access, video download, gaming) in both the downlink and uplink directions. Uplink QoS is particularly important as this controls traffic entering the LTE network.

The choice of scheduling algorithm is not standardised for LTE, instead it is left as an implementation decision. Selecting and implementing the appropriate scheduling algorithms therefore requires vendors to balance QoS, fairness and implementation complexity.

Roke Manor Research (Roke) is developing solutions for the eNodeB MAC Scheduler which enable network equipment providers and operators to realise the benefits of a QoS-enabled LTE air interface. This technology and expertise enables operators to migrate from a best-effort mobile data service to a differentiated LTE service. A differentiated LTE service:

- Provides consumers with a range of reliable, real-time services and access options.
- Provides developers with a set of tools which facilitate the development of innovative new applications.
- Provides operators with the means to recognise the monetise the value associated with mobile data.

Roke has been an active participant and contributor to the LTE design within the 3GPP RAN working groups since 2005.

References

1. 3GPP TS 23.203: *Policy and charging control architecture*
2. 3GPP TS 36.321: *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification*
3. Roke Manor Research, *LTE MAC Scheduler & Radio Resource Scheduling*.

For further information please contact
Patrick Fuller
Business Sector Consultant

Roke Manor Research Ltd
Roke Manor, Romsey
Hampshire, SO51 0ZN
United Kingdom
Part of the Chemring Group

T +44 (0)1794 833675
F +44 (0)1794 833433
patrick.fuller@roke.co.uk
www.roke.co.uk

00485