7/27/2010

HARISHVADADA.WORDPRESS.COM   LTE-WiMAX Blog

QOS over 4G networks | Harish Vadada

Cellular network operators across the world have seen an explosive growth of mobile broadband usage. Traffic volume per subscriber is increasing daily; in particular, with the introduction of flat-rate tariffs and more advanced mobile devices. Operators are moving from a single-service offering in the packet-switched domain to a multi-service offering by adding Value added services (VAS) that are also provided across the mobile broadband access. Examples of such services are multimedia telephony and mobile-TV. These services have different performance requirements, for example, in terms of required bit rates and packet delays. Solving these performance issues through over-provisioning typically is uneconomical due to the relatively high cost for transmission capacity in cellular access networks which includes radio spectrum and backhaul from the base stations.

4G broadband wireless technologies such as IEEE 802.16e/m and Third Generation Partnership Project (3GPP) - Long Term Evolution (LTE) have been designed with different QoS (Quality of Service) frameworks and means to enable delivery of the evolving Internet applications. QoS specifically for evolving Internet applications is a fundamental requirement to provide satisfactory service delivery to users and also to manage network resources

Quality of Experience (QOE) for wireless networks has always been dependant on more variable Radio links and factors like RSSI, RxQUAL, BER, Interference, mobility and number of simultaneous sessions. A wireless link generally offers less bandwidth than a wired connection. Throughput, latency/delay and transmission errors vary much more widely over a wireless connection because of constantly changing radio signal conditions and extensive digital radio processing. Standard internet protocols, designed for use over a more stable wire-based connection, are not well-suited to handle these variations. The same wideband radio channel must be shared among many user sessions and each user session typically involves many different types of data streams and protocols. The traditional QoS framework is no longer efficient and/or sufficient to support these new mobile Internet applications with good or required user experience. Traffic shaping and QOS will play a big role in how services are delivered in the evolved 4G systems, with IP as a backbone to their deliverability and will be done via Deep Packet Inspection(DPI) mechanisms that will exist at the control nodes of the network.

QoS refers to the ability (or probability) of the network to provide a desired level of service for selected traffic on the network.

- Service levels are specified in terms of **throughput**, **latency** (delay), **jitter** (delay variation) and **packet errors or loss**.
- Different service levels are specified for different types or streams of traffic.
- To provide QoS, the network identifies or "classifies" different types or streams of traffic and processes these traffic classes differently to achieve (or attempt to achieve) the desired service level for each traffic class.
- The effectiveness of any QoS scheme can be measured based on its ability to achieve the desired service levels for a typical combination of traffic classes ("traffic profile").

Let us discuss the QOS/queuing algorithms for LTE and WiMAX (802.16e/m).

QOS over LTE networks:
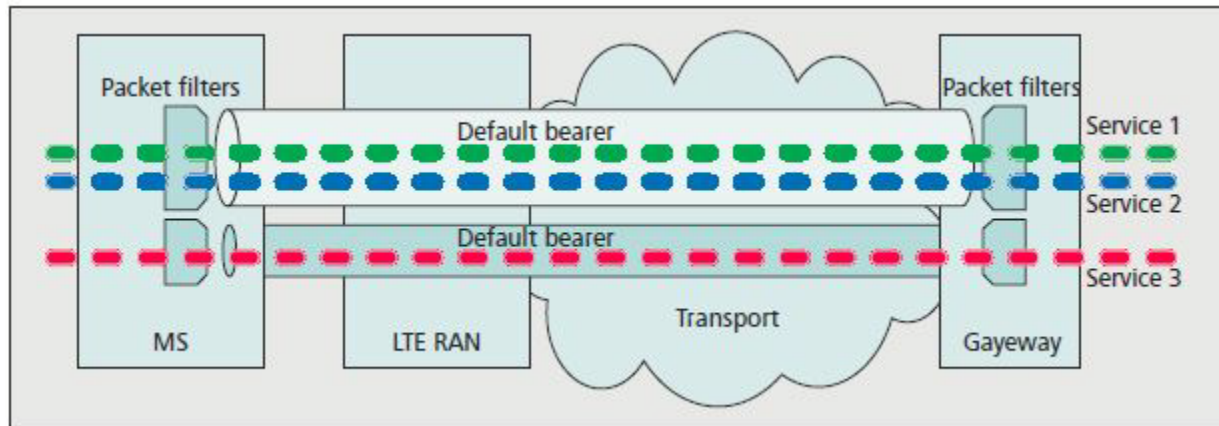
The QoS level of granularity in the LTE evolved packet system (EPS) is bearer, which is a packet flow established between the packet data network gateway (PDN-GW) and the user terminal (UE or MS). The traffic running between a particular client application and a service can be differentiated into separate service data flows (SDFs).

SDFs mapped to the same bearer receive a common QoS treatment (e.g., scheduling policy, queue management policy, rate shaping policy, radio link control (RLC) configuration). A bearer is assigned a scalar value referred to as a QoS class identifier (QCI), which specifies the class to which the bearer belongs. QCI refers to a set of packet forwarding treatments (e.g., scheduling weights, admission thresholds, queue management thresholds, and link-layer protocol configuration) preconfigured by the operator for each network element. The class-based method improves the scalability of the LTE QoS framework. The bearer management and control in LTE follows the network-initiated QoS control paradigm, and the network initiated establishment, modification, and deletion of the bearers.

LTE bearers:
Guaranteed bit rate (GBR): Dedicated network resources related to a GBR value associated with the bearer are permanently allocated when a bearer becomes established or modified.
 Non-guaranteed bit rate (non-GBR): A service utilizing a non-GBR bearer may experience A non-GBR bearer is referred to as the default bearer, which is also used to establish IP connectivity, similar to the initial Service Flow in WiMAX. Any additional bearer(s) is referred to as a dedicated bearer and can be GBR or non-GBR.

In LTE the mapping of SDFs to a dedicated bearer is classified by IP five-tuple based packet filter either provisioned in PCRF or defined by the application layer signalling. However, the default bearer typically uses a match all packet filter; any SDF that does not match any of the existing dedicated bearer packet filters is mapped onto the default bearer. Therefore, if a dedicated the default bearer.

LTE specifies a number of standardized QCI values with standardized characteristics, which are preconfigured for the network elements. This ensures multivendor deployments and roaming. The mapping of standardized QCI values to standardized characteristics is captured below diagram. Besides QCI, the following are QoS attributes associated with the LTE bearer:

QCI	Resource type	Priority	Packet delay budget	Packet error loss rate	Example services
1	GBR	2	100 ms	10^{-2}	Conversational voice
2	GBR	4	150 ms	10^{-3}	Conversational video (live streaming)
3	GBR	3	50 ms	10^{-3}	Real time gaming
4	GBR	5	300 ms	10^{-6}	Non-conversational video (buffered streaming)
5	Non-GBR	1	100 ms	10^{-3}	IMS signaling
6	Non-GBR	6	300 ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7	Non-GBR	7	100 ms	10^{-6}	Voice, Video (live streaming), Interactive gaming
8	Non-GBR	8	300ms	10^{-3}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9	Non-GBR	9	300ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)

QCI: A scalar representing a set of packet forwarding treatments (e.g., scheduling weights, admission thresholds, queue management thresholds, and link layer protocol configuration).

Allocation and retention priority (ARP): A parameter used by call admission control and overload control for control plane treatment of a bearer. The call admission control uses the ARP to decide whether a bearer establishment or modification request is to be accepted or rejected. Also, the overload control uses the ARP to decide which bearer to release during overload situations.

Maximum bit rate (MBR): The maximum sustained traffic rate the bearer may not exceed; only valid for GBR bearers.

GBR: The minimum reserved traffic rate the network guarantees; only valid for GBR bearers

Aggregate MBR (AMBR): The total amount of bit rate of a group of non-GBR bearers In 3GPP Release 8 the MBR must be equal to the GBR, but for future 3GPP releases an MBR can be greater than a GBR.The AMBR can help an operator to differentiate between its subscribers by assigning higher values of AMBR to its higher-priority customers compared to lower-priority ones.

LTE air interface scheduler

The LTE air interface scheduler is responsible for dynamically allocating DL and UL air interface resources among the bearers appropriately while maintaining their desired QoS level in both DL and UL directions. In order to make a scheduling decision, the LTE air interface scheduler uses the following information as input:

- Radio conditions at the UE measured at the eNB and/or reported by the UE.
- The state of different bearers, such as uplink buffer status reports (BSR) that are required to provide support for QoS-aware packet scheduling, elapsed time.
- The QoS attributes of bearers and packet forwarding parameters associated with the QCIs.
- The interference situation in the neighboring cells. The LTE scheduler can try to control intercell interference on a slow basis. This improves the QoE associated with the MSs at the cell edge.

QOS over WiMAX networks:
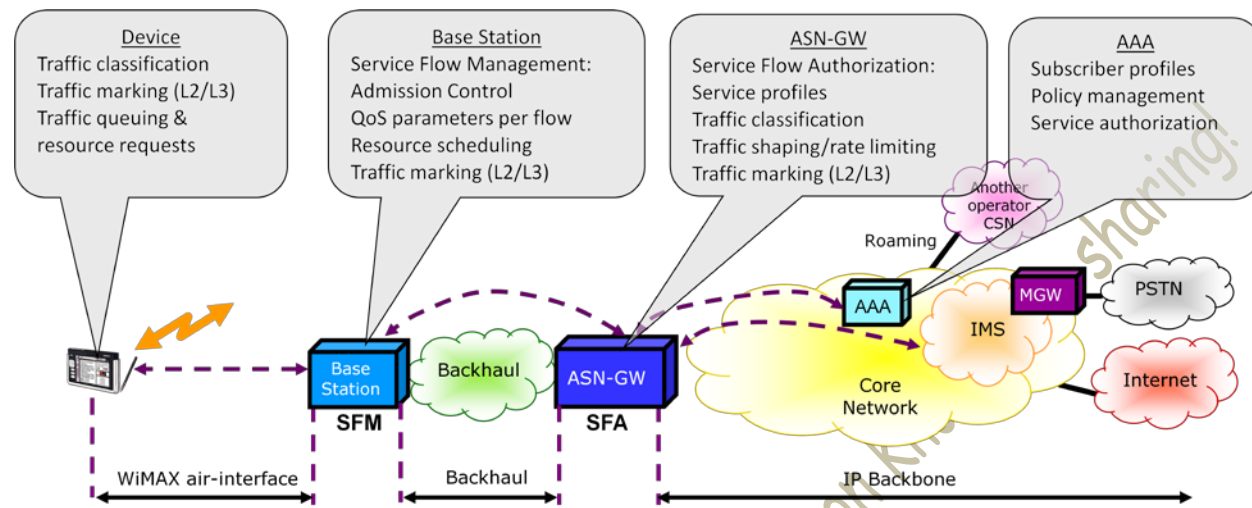
WiMAX employs flow-based QoS - traffic can be classified to different service flows with different QoS parameters. The ASN (Access Service Network) supports admission control & resource scheduling to manage (nonguaranteed) QoS per service flow. The WiMAX ASN also marks traffic to enable other networks/elements (e.g. backhaul network) to provide QoS consistent with the air interface. WiMAX provides QoS by classifying traffic to service flows with different QoS. A service flow (SF) is a unidirectional MAC-layer transport connection with particular QoS parameters.

- The WiMAX network creates at least two (1 DL + 1 UL) service flows (default service flows) for a device when it enters the network.
    - The default service flows are Best Effort and support most traffic.
    - The default service flows are also used for DHCP, DNS, etc.
- Devices may also be pre-provisioned with additional "dedicated" service flows to provide QoS for selected applications.
    - Traffic must be classified to dedicated service flows.
    - A single device can currently support up to 8 active service flows (4 downlink + 4 uplink), including the default service flows.
- WiMAX provides mechanisms for dynamically creating, modifying and deleting dedicated service flows during a subscriber's active session.
    - Requests can be initiated by either the network or device.

WiMAX Service Flows:

IEEE 802.16e supports 5 SF types:

Service flow Type	MRTR	MSTR	Max Latency	Max Jitter	Traffic Priority	Targeted Traffic
UGS		x	x	x		Constant bit rate (CBR) services, TDM services
ertps	x	x	x	x	x	VoIP with silence suppression or activity detection
rtps	x	x	x		x	Streaming audio and video
nrtps	x	x			x	File transfers
BE		x			x	Web browsing, email

Unsolicited grant service (UGS): Supports real-time traffic with fixed-size data packets on a periodic basis

Real-time polling service (rtPS): Supports real-time traffic with variable-size data packets on a periodic basis

Extended rtPS (ertPS): Supports real-time traffic that generates variable-size data packets on a periodic basis with a sequence of active and silence intervals

Non-real-time polling service (nrtPS): Supports delay-tolerant traffic that requires a minimum reserved rate

Best effort (BE) service: Supports regular data services

WiMAX air interface scheduler

The SF framework provides QoS granularity and inter-SF isolation over the air interface. The air interface scheduler is responsible for enforcing QoS by assigning DL and UL physical (PHY) layer resource blocks among SFs. This mechanism is called bandwidth allocation. A scheduling decision is determined based on appropriate SFs' QoS state variables, like buffer lengths, elapsed packet delay, SFs' QoS requirements such as MRTR and maximum latency, and radio frequency (RF) conditions of different MSs.

In general:

- SFs with shorter maximum latency or SFs with higher MRTR receive higher priorities in the scheduling decision.
- SFs with late packets or long buffer lengths also, receive higher priorities in the scheduling decision.
- MSs with better RF conditions receive higher priorities by the scheduler in order to improve overall sector throughput. However, an operator can adjust fairness to ensure MSs in poor RF conditions receive reasonable QoS.
- The air interface scheduler may differentiate between traffic flows within an SF by packet priority levels such as DSCP values (intra-SF). Also, it may further utilize the traffic priority attribute of SFs to differentiate between traffic associated with SFs of the same type (inter-SF).

WiMAX and LTE Comparison:

There are more components and functionalities in an end-to-end network providing QoS than the air interface QoS features discussed above, such as policy control and charging (PCC) functions in QoS provisioning. Here, we focus on a comparison of the QoS framework between LTE and IEEE 802.16e/IEEE 802.16m at the air interface:

QoS transport unit: The basic QoS transport unit in the IEEE 802.16e/IEEE 802.16m system is an SF, which is a unidirectional flow of packets either UL from the MS/AMS or DL packets from the BS/ABS [6]. The basic QoS transport in LTE is a bearer between UE and the PDNGW. All packets mapped to the same bearer receive the same treatment.

QoS scheduling types: There are six scheduling service types in IEEE 802.16m including UGS, ertPS, rtPS, nrtPS, and BE from IEEE 802.16e and the newly defined aGP service. LTE supports GBR and non-GBR bearers. The GBR bearer will be provided by the network with a guaranteed service rate, and its mechanism is like rtPS; the non-GBR has no such requirement and performs like BE in IEEE 802.16e/IEEE 802.16m.

QoS parameters per transport unit: Depending on the SF type, IEEE 802.16e/IEEE 802.16m can control maximum packet delay and jitter, maximum sustained traffic rate (MSTR), minimum reserved traffic rate (MRTR), and traffic priority. LTE MBR and GBR are similar to IEEE 802.16e/IEEE 802.16m MSTR and MRTR, respectively. However, MBR and GBR are only attributes of GBR bearers, while in IEEE 802.16e/IEEE 802.16m even a BE SF can be rate limited using its MSTR. Also, with 3GPP Release 8, GBR and MBR are set equal, while IEEE 802.16e/IEEE 802.16m allows the operator to select independent values for MSTR and MRTR. On the other hand, LTE AMBR allows the operator to rate cap the total non-GBR bearers of a subscriber.

QoS handling in the control plane: The SF QoS parameters are signaled in IEEE 802.16e/IEEE 802.16m via DSx/AAI-DSx messages. In LTE the QCI and associated nine standardized characteristics are not signaled on any interface. Network initiated or client initiated QoS are both supported in IEEE 802.16e/IEEE 802.16m systems. Therefore, both operator managed service and unmanaged service can be supported. The flexible architecture gives the mobile client opportunities for differentiation. LTE only supports network initiated QoS control.

QoS user plane treatment: The ARP parameter in LTE provides the following flexibilities to the operator:

- Accept or reject establishment or modification of bearers during the call admission control decision based on not only the requested bandwidth, available bandwidth, or number of established bearers, but also the priority of the bearer
- Selectively tear down bearers based on their priorities during an overload situation