Quality of Service in 3GPP R12 LTE-Advanced

Najah Abu Ali, United Arab Emirates University
Abd-Elhamid M. Taha, Alfaisal University
Hossam S. Hassanein, Queen's University

ABSTRACT

Growing demand for mobile data traffic is challenging even the capacities of next generation wireless networks. In response, operators worldwide are expanding and updating their deployment. In turn, 3GPP continues to explore ways to empower operators with features for more capable, economic, and energy-efficient networks. Toward 3GPP Release 12, focus has shifted to accommodate the inevitable traffic explosion in both magnitude and traffic types. In this work, we highlight some of the features of LTE-Advanced Release 12 relevant to improving quality of service. Specifically, we focus on solutions explored to enhance network capacity and service delivery in terms of offloading, improved services, and improved congestion control.

INTRODUCTION

As updates are made to existing third generation (3G) networks, and new commercial Long Term Evolution (LTE) networks are deployed worldwide, the demands of mobile data traffic continue to grow in magnitude and variety. Motivating this growth is the rise in penetration for smartphones, in addition to the boom in sensing/actuating and machine-to-machine (M2M) applications. In turn, operators continue to expand the capacities of their networks. The Global Mobile Suppliers Association (GSA) predicts that the number of LTE commercial deployments will reach 284 networks in 87 countries by the end of 2013 [1]. Specified by Third Generation Partnership Project (3GPP) Releases 8 and 9, LTE realizes substantial gains in network capacity relative to its predecessor. LTE will pave the road for LTE-Advanced networks. Described by 3GPP Releases 10 and beyond, LTE-Advanced networks are promising at least 1 Gb/s in the downlink for stationary to low-mobility users.

Projections for mobile data traffic, however, are challenging even these next generation network capacities. In 2012, global mobile data traffic was 70 percent larger than that of 2011, rising to an average of 885 Pbytes/mo. By 2017, this average is expected to rise to 11.16 Ebytes — a 13-fold increase that is to be generated by around 40 billion devices [2, 3]. Of these, 563,481 Tbytes will be for M2M traffic. Therefore, early deployments of LTE-Advanced (frozen March 2011) already began in late 2012 and early 2013, even though LTE-Advanced handsets have yet to hit the market.

Various enhancements for LTE-Advanced have already been made in Release 11 (frozen December 2012). These include additional carrier types for carrier aggregation, enhanced downlink control channel, the introduction of coordinated multipoint transmission, further enhanced intercell interference, overload control for machine type communication (MTC), in addition to other improvements in signaling and energy saving techniques.

Currently, 3GPP is working toward Release 12 (to be finalized by June–September 2014). The objectives specifically target the expected traffic explosion, in addition to other continued concerns for energy savings, cost efficiency, support for diverse applications and traffic types, higher user experience, and backhaul enhancement [4].

In this article, we highlight some of the main efforts in 3GPP Release 12 addressing quality of service (QoS). We review measures taken to enhance network capacity at the various layers, but especially in terms of offloading to wireless LANs (WLANs) and small cells. We also look at enhancements to relevant network services such as evolved multimedia broadcast/multicast service (eMBMS) and MTC, in addition to newfound support for peer-to-peer traffic. We also look at solutions considered to address congestion control in LTE-Advanced.

A NOTE ON READING 3GPP STANDARDS

LTE-Advanced is described in Release 10 and beyond. For the latest in any release, the reader should consult http://www.3gpp.org/ftp/specs/html-info/status-report.htm. A review of 3GPP releases and their freeze dates can be found at http://www.3gpp.org/Releases. A description of releases, in addition to the ongoing study and

 0163-6804/13/$25.00

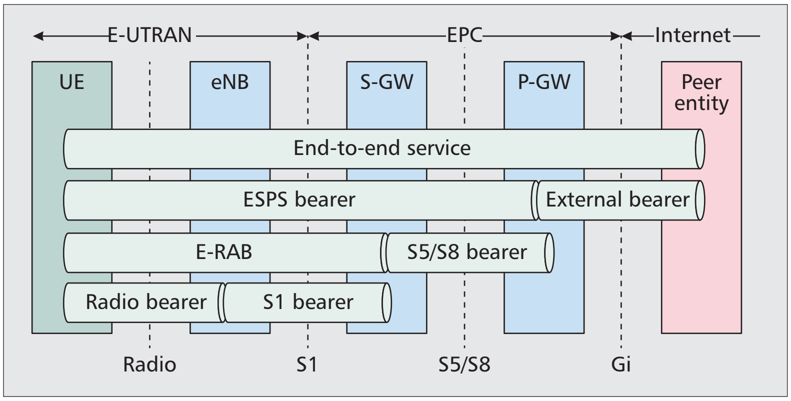

Figure 1. *An elaboration on the different bearers employed in end-to-end service delivery in LTE and LTE-Advanced.*

work items, can be found at http://www.3gpp.org/ftp/Specs/html-info/FeatureListFrameSet.htm and http://www.3gpp.com/ftp/Information/WORK_PLAN/Description_Releases/. These descriptions include the one on Release 12, which was relied on when preparing this work. A good place to start on all access-relevant issues is TS 36.300 [5], including the overall network architecture for both access and core, in addition to an elaboration on end-to-end QoS management. High-level QoS requirements for LTE-Advanced and Evolved Packet System (EPS) can be found in TS 22.278. A significant report on aspects of traffic explosion can be found in TR 22.805.

PRELIMINARIES

Prior to discussing QoS in LTE-Advanced Release 12, a review of certain preliminaries on definitions and traffic handling is due.

PARAMETERS

Satisfying the QoS requirements of the various applications and services entails quantifying these requirements in terms of parameters that identify target performance levels. Such parameters include throughput, delay, jitter, and packet loss [5]. 3GPP identifies the following major quantitative parameters.

Throughput: Characterized through the guaranteed bit rate, maximum bit rate, and aggregate maximum bit rate

- Guaranteed bit rate (GBR): Allocated fixed network resources that do not change after bearer establishment or modification. This is hence a guaranteed service data flow.
- Maximum bit rate (MBR): Limits the bit rate that can be expected to be provided to GBR bearer, and is enforced by the network shaper to restrict the traffic to its MBR agreement.
- Aggregate maximum bit rate (AMBR): Used for non-GBR flows, and has two types, access point name-AMBR (APN-AMBR) and user equipment-AMBR (UE-AMBR). The APN-AMBR is a subscription parameter stored at the home subscriber server (HSS) per APN. The HSS defines a QoS class identifier (QCI) for each packet data network (PDN) (identifiable by an individual PDN identifier) and an APN-AMBR for each ARP. The APN-AMBR parameter refers to the maximum bit rate that can be consumed by all non-GBR bearers and all PDN connections of this APN. This parameter is enforced by the PDN gateway (P-GW) in the downlink and by both UE and the P-GW in the uplink. The UE-AMBR parameter, on the other hand, refers to the MBR allowed for all non-GBR bearer aggregates for the respective UE. This parameter is enforced in both the downlink and the uplink.

Delay: 3GPP defines nine categories for delay, with 50 ms being the tightest and 300 ms the slackest. The latter value is used for delay-tolerant applications.

Packet loss: Defined as the packet error loss Rate, similar to the packet delay budget in having nine categories with 10^{-6} being best and 10^{-2} being the worst.

Priority: Specified by the allocation/retention priority (ARP) parameter, which is used to indicate the priority of both allocation and retention of the service data flow. The ARP dictates whether a bearer establishment/modification request can be accepted or rejected in the event of conflicts in demand for network resources. At times of exceptional network resources limitations, such as handover, ARP can be used by the eNodeB to drop a flow with a lower ARP to free up capacity. ARP, however, has no effect on the network treatment received by the flow once the flow is successfully established.

Note that GBR and MBR are defined per bearer, while the AMBR parameters are defined per group of bearers. All throughput parameters have two components, one for downlink and another for uplink.

END-TO-END QOS

Figure 1 elaborates on the end-to-end view of data flows in LTE and LTE-Advanced. The figure identifies the three network levels: Evolved-Universal Mobile Telecommunications System (UMTS) terrestial radio access network (E-UTRAN), Evolved Packet Core (EPC), and the Internet. Within the E-UTRAN, there is the evolved Node B (eNB) serving the UE. Within the EPC there are the serving gateway (S-GW) and the P-GW, in addition to other entities such as the mobility management entity (MME). Finally, the figure illustrates the S1 interface between the eNB and the EPC, and the S5/S8 interface between the S-GW and the P-GW.

3GPP distinguishes between bearers at the different network levels. A radio bearer, for example, is the over-the-air connection between the UE and the RAN, while the S1 bearer is that between the eNB and the EPC network entities.

Bearers can also be classified into default and dedicated. A default bearer is initiated and established at startup to carry all traffic toward a destination. The default bearer is a non-GBR bearer and does not provide bit rate guarantees. A dedicated bearer, on the other hand, can be either a GBR or non-GBR bearer. If GBR, it can specify the bit ratethat is guaranteed, packet delay, and packet loss error rate. Each dedicated bearer is characterized by a traffic flow template

(TFT) detailing the bearer's QoS parameters. An uplink TFT is used to map the UE uplink traffic to specific QoS parameters, with the mapping carried out at both the eNodeB and the UE. Mapping for the downlink TFT is carried out at the S-GW or P-GW. Table 1 gives an example of a traffic classification based on the QoS parameters defined in the LTE QoS framework. Each class is identified by a scalar number, which is the QCI. A QCI identifies a group of QoS parameters describing the packet forwarding treatment in terms of priority, allowable delay, and packet error rate.

SERVICES

PEER-TO-PEER CONTENT DISTRIBUTION

The enhanced capacities of 3GPP networks have greatly encouraged the use of peer-to-peer (P2P) applications. Similar to e/MBMS, there is great motivation for enabling P2P services in LTE-Advanced due to their great potential in reducing the demand for backbone and backhaul traffic. In TR 22.906 (R11) [6], 3GPP studied P2P content distribution in IP multimedia subsystem (IMS)-based connections for both real-time and non-real-time traffic, and have indicated that IMS needs to be enhanced to support use cases such as content on demand, live streaming, and file downloading. A feasibility study of the support can be found in TR 23.844 [7] (architectural aspects) and TR 33.844 (security aspects). Considerations for voice P2P traffic were not made in this study.

In terms of requirements, TR 23.844 identified that a network abstraction is required for a P2P content delivery system (CDS). This includes an understanding of the different delivery options available for the network and the UE, their access type, offloading capability, current load, and path cost information. It also includes peer information such as (network) location, peer proximity, battery status, and number of clients served by peers. Support of QoS has naturally been mandated, in addition to the use of mechanisms that guarantee protections of copyrighted content.

Architecturally, TR 23.844 introduces three functional entities to the IMS network: a content source server (CSS), content cache server (CCS), and tracker application s erver (Tracker AS). The CSS is a content resource server that executes resource segmenting, and may execute resource encoding and transcoding, while the CCS caches content and/or content segments as dictated by the Tracker AS. The Tracker AS performs several tasks, including maintaining the list of peers, according to their activities; information associated with each Peer ID (access type, access conditions, workload, etc.); maintaining an index of contents and/or content segments and where the contents and/or content segments might be cached; directing communication and transfers between CSS and CCS.

The report offers three possible alternatives for supporting a P2P CDS system in IMS. The first alternative, shown in Fig. 2, considers when the signaling between the user peer and other peers does not traverse the IMS core. The second alternative considers when signaling between the user peer and other peers does traverse the IMS core. In this case, the PCRF controls all peers' communication. As for the third alternative, neither the user peer and Tracker AS signaling, nor the user peer and other peer signaling traverse the IMS core; rather, the user peer interfaces directly with the tracker AS.

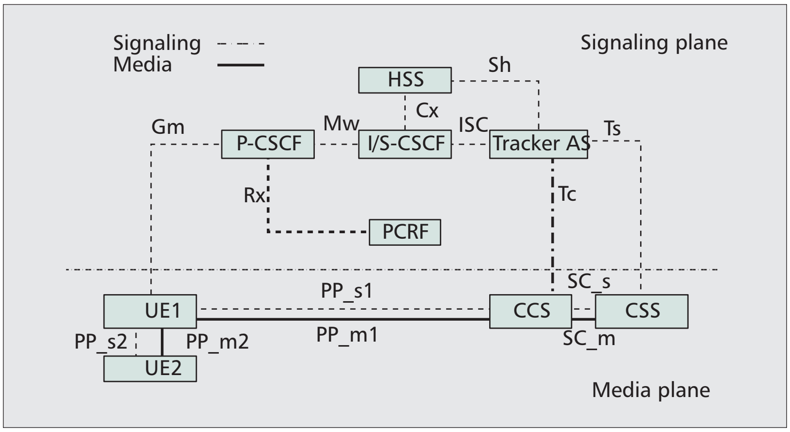

Figure 2. *Overview of the first alternative architecture for the IMS P2P CDS [7].*

In joining a CDS, the UE first associates with the Tracker AS to request initial and successive peer lists before connecting to the CSS/CCS. QoS control can be achieved in one of three ways.

- The UE may establish a dedicated bearer with the QoS policy from the policy charging and rules function (PCRF), the packet filters generated from the local and remote addresses, and port numbers from the peer list.
- It is also possible for the operator to assign a particular APN for content delivery. In this case, a user interested in using the CDSs would have the data received on a bearer established based on the APN's QoS policy.
- The third alternative involves a fourth P2P entity called the media proxy between the UE and the CCS. The proxy reduces the load resulting from the excessive number of IP filters generated from the large number of caches. It interacts with the Tracker AS to perform the required address/port mapping and load distribution.

MACHINE TYPE COMMUNICATION

MTC will assume a considerable share of the projected traffic increase in future networks. This motivates 3GPP's efforts on enabling operator networks to satisfy the requirements of MTC applications, while providing satisfactory service to both human and machine applications. Various issues have already been addressed in Releases 10 and 11, including overload and congestion control, low-priority access, downlink throttling, addressing space, device triggers, and defining interfaces between MTC servers and the mobile network.

Efforts in Release 12 [8] continue to be motivated by maintaining both human communication and MTC. Enhancements such as optimizations for collocated MTC devices, loca-

QCI	Type	Packet delay budget	Packet error loss rate	Example services
1	GBR	80 ms	10^{-2}	Conversational voice
2		130 ms	10^{-3}	Conversational video (live streaming)
3		30 ms	10^{-3}	Real-time gaming
4		280 ms	10^{-6}	Non-conversational video (buffered streaming)
5	Non-GBR	80 ms	10^{-6}	IMS signaling
6		280 ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., web, -mail, chat, FTP, P2P file sharing, progressive video)
7		80 ms	10^{-3}	Voice, Video (live streaming) Interactive gaming
8		280 ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., web, email, chat, FTP, P2P file sharing)

Table 1. *An example QCI to ARP mapping (TS36912-b00).*

tion tracking, and network selection are being introduced. In the following, we consider three major enhancements that either have been introduced or are currently under consideration: traffic categorization, dedicated cores, and device-to-device (D2D) communications.

Traffic Categorization — TS 22.368 [8] distinguishes specific modes of MTC, including small data transmissions, and time controlled and group-based communication. For small data transmissions, measures are taken to reduce communication requirements in terms of signaling overhead and network resources, as well as minimizing the delay for reallocation. For such devices to transmit, they can be either attached to or detached from the network, although for charging purposes the system must recognize communications instances.

Machines communicating in a time controlled manner serve applications that can tolerate sending or receiving data only during defined time intervals, and thus require reduced signaling. Time controlled devices may communicate outside these intervals and be charged.

For group communications, the standard does not specify how "grouping" should be done, leaving the matter up to operators and vendors. The standard does note that features assigned to some machines within a group may not necessarily be assigned to others. This includes QoS policy features, where an operator can specify a so-called group-based policing MTC feature.

Architecture for Device-to-Device MTC — A prominent enhancement to MTC in Release 12 is the introduction of an architecture for D2D communication. There are several modes in which D2D can take place. Devices can communicate over 3GPP networks or through an MTC server. Grouped devices can also communicate directly with other devices within the group, but not beyond. Note that in grouped communication, only one device, the MTC gateway, connects to the 3GPP networks in its radius. Local connectivity within the group can be in IEEE 802.16, Zigbee, Bluetooth, and so on.

Dedicated M2M Core — As the number of machines utilizing 3GPP networks increases, an existing concern is that MTC may affect human type traffic at the network core. Efforts have been invested in realizing a dedicated network core to divert the MTC traffic, an architectural feature that would become particularly advantageous in instances of congestion. This separation allows for further enhancements, including shared cores between several M2M service operators, or having multiple cores within the same network.

EVOLVED MULTIMEDIA BROADCAST MULTICAST SERVICES

Multicasting and broadcasting are part of the integral services in 3GPP networks, and play a key role in enabling video, audio, and data broadcasts in LTE-Advanced. Increasing interest has been made in employing LTE/-Advanced eMBMS to enhance network resource utilization and enhance overall user multimedia experience. More recently, applications in public safety and MTC are also relying on eMBMS.

eMBMS was first defined for LTE in R9. Since then, additional features and enhancements were made releases R10 and R11 that include utilizing Address Resolution Protocol (ARP) for prioritizing MBMS, reducing signaling requirements, supporting dynamic adaptive streaming of HTTP (DASH), and supporting service continuity during interfrequency and intercell handovers. Further enhancements are expected in R12, including more efficient restoration (failure recovery) mechanisms from failure types that were not addressed in previous releases.

The logical architecture of eMBMS in LTE/-Advanced involves a multicell/multicast coordination entity (MCE) between the eNB and the MME, in addition to an MBMS gateway (MBMS-GW) [5]. In turn, the MBMS-GW operates between the eNB and a broadcast multicast service center (BMSC). An MCE oversees the admission control, bearer management, resumption, and suspension of an MBMS session in a single multicast-broadcast single-frequency network (MBFSN). (An MBMS service area comprises several MBFSNs). Meanwhile, an MBMS-GW utilizes IP multicast in order to send and broadcast MBMS packets to each eNB involved in the transmission service. It also performs session control (start/update/stop) toward the LTE-Advanced network through the MME.

In previous releases, 3GPP has introduced restoration procedures for CS and PS services resulting from various failures in the EPC. However, it lacks effective restoration mechanisms to minimize the impact of entity (MCE, MBMS,-GW, BMSC) or path failures between any of the connected entities. Mechanisms to overcome such failures are currently under investigation.

Other enhancements currently under investigation include enhanced support for DASH and support for emergency handling MBMS sessions. More notably, the possibility of detecting the rise and fall of surging high-demand requests by operators (e.g., breaking news or firmware update) and reducing MBMS channel signaling requirements are also being explored.

User Plane Congestion Control in 3GPP RANs

TR 22.805 offers a detailed study on the effect of smartphones and data services penetration on network activities, especially when considering new modes of data transfer such as social networks and push traffic. As demand exceeds network capacity, congestion in the user plane is inevitable. Accordingly, 3GPP is currently exploring enhancements in which this congestion can be controlled.

The QCI-based QoS architecture described above enables some basic policy enforcement. For instance, QCI/ARP pairs can be allocated to different bearers (default or dedicated). As an example, QCI 8 can be assigned to the default bearer of high-priority subscribers, while QCI 9 is assigned to subscribers with lower priority.[1] Application data can then be mapped to the different bearers. QoS policy associated with the bearer's QCI/ARP can then be forced on the aggregated data of the different applications within that bearer.

Such an approach, however, is limited in several aspects. For example, for certain traffic types it can only be applicable to traffic in the downlink. It also becomes operationally exhaustive in instances of mobility and roaming. Meanwhile, the approach lacks sufficient granularity if there is interest in handling congestion for traffic of a specific application.

Instances of congestion cannot be transient/short, and should not be confused with well utilized network links. For a radio access network (RAN), congestion in the user plane can take place in one of three spots:

- Over the air
- At the eNB (i.e., when user traffic exceeds the backhaul capacity of the eNB)
- Between the RAN and the core (i.e., when the RAN's traffic exceeds the backhaul capacity of the network)

Overall requirements for user plane congestion management are noted in TS 22.101. Requirements include that a RAN can detect the onset and abatement of user plane congestion, recognize congested cells, and enforce congestion control policies in a timely manner and at a low signaling overhead. Three mechanisms are noted in implementing congestion control: prioritizing traffic, reducing traffic, and limiting traffic. Traffic prioritization is expected at three levels: user/subscriber, application, and traffic type. Reducing traffic refers to the possibility of adapting multimedia applications or reducing signaling requirements for optimizable protocols when congestion is noted. Finally, limiting traffic refers to the capability of limiting traffic from either operator controlled and/or third party services based on the RAN user plane congestion for a UE device.

The above noted requirements indicate interest in congestion control mechanisms with tuneable granularity, where at the coarser end the network should be able to prioritize subscribers, and at the finest end the network should be able to identify traffic types independent of the application. For example, with the varied nature of social network traffic, the interest extends to enable the networking to control video traffic within the overall social network traffic. Such granularity does not exist in a 3GPP network. Between these two ends, the expected congestion control mechanisms will be able to isolate the traffic for different applications.

In addition to the variable granularity requirement, TR 23.705 [10] identifies further key issues in developing a congestion management solution. These include the architectural design of the RAN congestion mitigation system, its measures, and its scope; congestion awareness mechanisms and their relevant signaling and actions; offering differentiated treatment for non-deducible service data flows; video delivery control; and uplink traffic prioritization.

Given the access/core separation maintained in LTE-Advanced, it is expected that congestion control mechanisms will be implemented in both the core network and the RAN. The specific manner in which the RAN predicts or detects a RAN user plane congestion, as well as the manner in which the congestion is recognized to be abated, is left to vendors and operators. The in-progress 23.705 provides a high-level description of congestion management, which essentially breaks down to the following:

- Congestion prediction/detection at the RAN
- Congestion indication to the core network
- Decision on mitigation measures at the core network
- Mitigation mechanisms enacted at the core network
- If applicable, mitigation information/decisions relayed to the RAN

Overall requirements for user plane congestion management are noted in TS 22.101. Requirements include that a RAN can detect the onset and abatement of user plane congestion, recognize congested cells, and enforce congestion control policies in a timely manner and at a low signaling overhead.

[1] *Example noted in 3GPP TR 23.705 V0.3.0.*

It is the support for PMIP that expands the interoperability to networks employing other non-3GPP RATs but with IP-based cores. This support also allows for extending the IMS PCC to non-3GPP networks, as well as the required charging and QoS control policies.

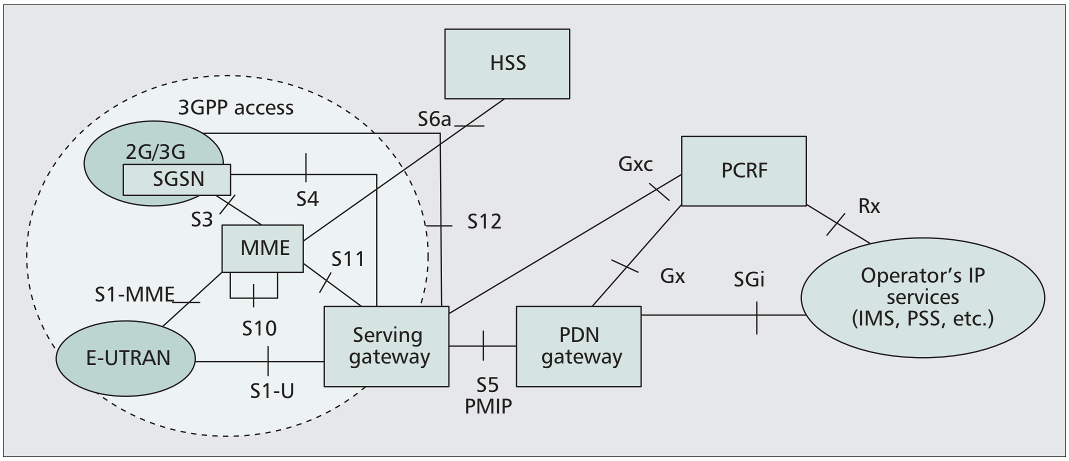

Figure 3. *Non-roaming architecture for 3GPP access wtihin EPS using PMIP-based S5 [11].*

To enable the RAN to indicate congestion to the core, the report suggests utilizing a RAN congestion control information (RCI) extension field in the general packet radio service (GPRS) Tunneling Protocol (GTP)-U header. (GTP is the tunneling protocol utilized on the S1 interface between the eNB and the core). At minimum, the RCI would identify the level of the RAN user plane congestion in addition to the location of the congested RAN (e.g., using the cell ID).

To enable the differentiation between traffic for different applications or different traffic types, TR 23.705 also introduces the notion of a flow priority indicator (FPI), which is to be used to identify the priority of the packet compared to other packets mapped in the same bearer. It is noted that the FPI is not intended to replace QCI, and that no conflicts are to occur between the two indicators. Rather, the FPI complements the use of QCI. Both indicators will be used to differentiate flows of the same UE device, and flows of different UE devices. However, in instances of congestion where the packet delay budget can be satisfied by a service data flow aggregate with the same QCI/ARP, the scheduler will attempt to satisfy the packet delay budget for packets with higher FPI.

A final note on the issue of prioritizing uplink traffic: 3GPP's interest is in enabling both per-flow and per-user prioritization. Recent demand indicates that users may potentially utilize considerably large data rates in the uplink, as is the case with P2P applications, video conference, and gaming. Systems for congestion control should therefore consider the possibilities of paired and/or unpaired uplink and downlink links. For example, a bidirectional voice call would be assigned similar priority in both directions.

Interworking LTE-Advanced Networks with WLAN

The use of WLANs for offloading traffic required three main issues to be resolved:
- Enabling handover from 3GPP networks to WLANs and vice versa
- Extending EPS charging and control policies
- Enabling measures to reliably predict QoS performance in WLANS

A seamless 3GPP-WLAN handover is one where the UE is allowed to hold simultaneous connections with both 3GPP networks and WLANs with both networks connected to the same core. In Release 10, seamless handovers were defined for flows, enabling the UE to distribute the flows between the RATs. Release 10 also defined non-seamless handover where the UE can route specific IP flows via the WLAN without traversing the EPS. The selection of these flows would be based on user preferences. Focus in Release 11 was on enhancing policy and QoS in WLAN access, in addition to support of trusted WLAN procedures.

To allow for interworking EPS with non-3GPP networks (WiFi, WiMAX, etc.), the S5/S8 interface between the S-GW and P-GW, shown in Fig. 3, has been defined in two flavors, GTP and proxy mobile IP (PMIP).[2] It is the support for PMIP that expands the interoperability to networks employing other non-3GPP radio access technologies (RATs) but with IP-based cores. This support also allows for extending the IMS PCC to non-3GPP networks, as well as the required charging and QoS control policies.

The model for QoS applied to PMIP does not use bearer IDs for user plane packets; rather, it utilizes packet filters at the core's PCRF using QoS parameters (QCI, ARP, MBR, GBR). Bearer connectivity for the UE is as follows: The EPS bearers concatenate the radio bearer and the S1 bearer. The bearer for PDN connectivity concatenates the EPS bearer and the IP connectivity between the S-GW and the P-GW.

There are currently several ongoing efforts to enhance offloading to WLANs in 3GPP networks. For example, WLAN network selection for a roaming 3GPP UE device or machine is currently being evaluated [12]. Several considerations relevant to QoS are taken into account, including satisfying connection requirements, load-based and context/venue-based networking selection, enabling simultaneous connections (i.e., connected to a 3GPP network and a WLAN at the same time), policy interaction between

[2] *S5 for non-roaming users; S8 for roaming ones.*

home and visited networks, in addition to load balancing.

Optimizing 3GPP-WLAN offloading is also currently under consideration [13]. The issues addressed include revising the 3GPP network selection function (the access network discovery and selection function); mitigating undesired bearer handling, like loss of bearers or significant degradation in bearer QoS; and mitigating ping-pong offloading.

Small Cells

Small cells entail base stations with reduced coverage offering higher and more reliable data rates. In addition to enhancing user experience, small cells are more economic in deployment and operation, and are expected to offer great reduction in energy requirements for both the user and the network. Understandably, deploying small cells introduces challenges at both the physical and higher layers [14].

Due to mixed-mode deployment (planned and ad hoc) and high temporal and spatial variations in traffic, real-time medium and capacity management become inevitable requirements. Various physical layer advances, such as coordinate multipoint and intercell interference cancellation (ICIC), in addition to notions such as self-optimization become applicable candidates. Frequency allocation between macro and small cells should also allow for both co-channel and orthogonal allocations.

At the higher layers, issues such as signaling overhead and mobility management must be addressed. For example, it has been shown that handover at the small cell level is not as good as it is at the macro level in terms of handover failure and ping-pong effect. Meanwhile, support for femtocell-to-femtocell handover is revisited to emulate X2 functionalities (i.e., inter-eNB interface functionality) with reduced complexity due to interworking delays. Finally, support for wireless backhauls for relay small cells is also being investigated in order to enhance plug-and-play operation and reduce deployment cost [15].

Conclusion

3GPP Release 12 promises to deliver many enhancements to LTE-Advanced and other 3GPP networks, increasing their ability to cope with the ever increasing demand for mobile data services. In this work, we highlight QoS-related matters in currently ongoing considerations. Other items of interest include enhancements to the physical layer, including the introduction of a new carrier type, the use of beamforming multiple-input multiple-output, and new bands defined for use in carrier aggregation. In addition, work must be done to more clearly understand the impact of energy saving measures on network operation, especially in terms of QoS. Also, a powerful new feature, proximity-based services, is expected to produce substantial improvements in various network functionalities, from cell selection to machine localization to other position-dependent operations.

References

[1] GSA Global LTE Market Update, http://goo.gl/6sMa9.
[2] Cisco Virtual Network Index (VNI), http://goo.gl/tfzlU
[3] D. Evans, "How the Next Evolution of the Internet Is Changing Everything," Cisco White Paper, http://goo.gl/mzk4n.
[4] T. Nakamura, "3GPP TSG-RAN Activities toward Further Enhancements for LTE/LTE-Advanced," ITU-R WP5D Workshop Geneva, Switzerland, 16 July, 2012.
[5] 3GPP TS 36.300 v11.5.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description."
[6] 3GPP TR 22.906 v11.0.0, "Study on IMS Based Peer-to-Peer Content Distribution Services."
[7] 3GPP TR 23.844 v12.0.0, "Feasibility Study on IP Multimedia Subsystem (IMS) Based Peer-to-Peer Content Distribution Services."
[8] 3GPP TR 22.888 v12.0.0, "Study on Enhancements for Machine-Type Communications (MTC)."
[9] 3GPP TS 22.368 v12.2.0, "Service Requirements for Machine-Type Communications (MTC)."
[10] 3GPP TR 23.705 v0.3.0, "System Enhancements for User Plane Congestion Management."
[11] 3GPP TS 23.402 v12.0.0, "Architecture Enhancements for Non-3GPP Accesses."
[12] 3GPP TR 23.865 v0.6.0, "WLAN network Selection for 3GPP Terminals."
[13] 3GPP TR 23.890 v0.4.0, "Offloading to WLAN in 3GPP RAT Mobility."
[14] 3GPP TR 36.932 v12.1.0, "Scenarios and Requirements for Small Cell Enhancements for E-UTRA and E-UTRAN."
[15] 3GPP TR 36.836 v2.0.1, "Mobile Relay for E-UTRA."

Biographies

Najah Abu Ali [M] (najah@uaeu.ac.ae) is currently an associate professor in the Faculty of Information Technology at the University of the United Arab Emirates. She got her Ph.D. in computer networks from the Department of Electrical Engineering, Queen's University, Kingston, Canada. Her research interests include analytical and measurement-based performance evaluation of wired and wireless computer networks, radio resource management in broadband wireless networks, and design of analytical and reliable schemes to provide QoS in wireless sensor networks. She has published her work in several prestigious journals and conferences, and delivered several tutorials in her fields of research at high-ranked conferences. She is a principal investigator/co-PI of several major research projects that have been funded by several agencies such as QNRF and NRF/University of the United Arab Emirates. She also co-authored a book detailing the functionalities and services of 4G and beyond technologies.

Abd-Elhamid M. Taha [SM] (ataha@alfaisal.edu) is currently an assistant professor at Alfaisal University, Kingdom of Saudi Arabia, and has written and lectured extensively on broadband wireless networks, focusing on radio resource management techniques. He is the co-author of the book *LTE, LTE-Advanced and WiMAX: Toward IMT-Advanced Networks* (Wiley) and a presenter for several tutorials at flagship IEEE Communications Society events. His other interests include scalability issues in the Internet of Things and modeling challenges in networked cyber physical systems.

Hossam Hassanein [SM] (hossam@cs.queensu.ca) is a leading authority in the areas of broadband, wireless, and mobile network architecture, protocols, control and performance evaluation. His record spans more than 500 publications in journals, conferences, and book chapters, in addition to numerous keynotes and plenary talks in flagship venues. He has received several recognitions and best paper awards at top international conferences. He is the founder and director of the Telecommunications Research (TR) Laboratory at Queen's University School of Computing, with extensive international academic and industrial collaborations. He is the past chair of the IEEE Communication Society Technical Committee on Ad Hoc and Sensor Networks (TC AHSN). He is an IEEE Communications Society Distinguished Speaker (Distinguished Lecturer 2008–2010).

A powerful new feature, proximity-based services, is expected to produce substantial improvements in various network functionalities, from cell selection to machine localization to other position-dependent operations.