ctc technology & energy

engineering & business consulting

Mobile Broadband Networks Can Manage Congestion While Abiding By Open Internet Principles

Prepared for the New America Foundation's

Open Technology Institute – Wireless Future Project

November 13, 2014

TABLE OF CONTENTS

FIGURES

TABLES

1. EXECUTIVE SUMMARY

As the FCC considers whether and how to extend its Open Internet rules to mobile broadband networks, mobile carriers have asserted that mobile networks present unique traffic management challenges that would make network neutrality, particularly a non-discrimination rule, technically infeasible and incompatible with quality of service (QoS). Mobile carriers emphasize the need to dynamically manage often unpredictable congestion among users who move from cell to cell on a network that is inherently capacity-constrained.

This Report examines these technical claims and concludes that Long Term Evolution (LTE) technology is capable of managing moderate congestion through prioritization protocols that are application-agnostic (e.g., user-directed prioritization) and which can, when faced with severe congestion, prioritize latency-sensitive traffic while avoiding discrimination among like applications, content, or services.

1.1 MOBILE BROADBAND PRIORITIZATION IS NECESSARY AND TECHNICALLY FEASIBLE—BUT IT REQUIRES CLEAR POLICY PRINCIPLES

The Internet is increasingly populated by devices and networks using wireless spectrum licensed by the FCC. Growing use of mobile computing devices (e.g., smartphones, tablets) and broadband applications (primarily streaming video) is driving a rapidly increasing consumer demand for mobile data.

At least in the short term, while we await implementation of more advanced and spectrum-efficient technologies, licensed mobile carrier spectrum can be scarce relative to demand. As designed, carrier networks face periodic, sometimes unpredictable capacity constraints. On mobile broadband networks, as on the roadways, congestion and bottlenecks can develop in particular places (cells/sectors) and at certain times, such as during sporting events or during peak hours in busy urban centers.

Although a significant and growing share of mobile device data traffic is now offloaded via Wi-Fi onto fixed networks, mobile network QoS policies are periodically necessary to manage latency-sensitive applications and functions in congested environments. In severely congested cells and sectors, prioritized applications can continue operating as consumers expect, while “like”[1] applications that are not prioritized or managed may work poorly or stop.

Long Term Evolution (LTE) technology provides a framework with enormous capability and flexibility for the prioritization and management of mobile data traffic. LTE’s QoS techniques can significantly improve the quality of the network users’ experience, especially with latency-sensitive applications such as voice and video communications. However, the implementation of the protocols can produce widely different outcomes for Internet users and edge providers (i.e., third-party application and content providers) depending on choices made by the wireless carrier. For example, carriers can choose to prioritize users or to prioritize uses – and to do either or both dynamically in different situations depending on the degree of congestion.

Left without clear public policy principles governing the implementation of QoS, wireless carriers can choose to use QoS to benefit only their own applications or those of industry partners. Such practices would leave unaffiliated edge providers or small application developers at a significant disadvantage.[2] Unaffiliated applications not prioritized by the carrier might be less reliable than the affiliated applications. For example, unaffiliated entities hosting voice might find their calls breaking up or stopped by congestion, while an affiliated voice application would continue operating.

1.2 THE WIRELESS INDUSTRY’S APPROACH TO PRIORITIZATION GIVES PREFERENCE TO AFFILIATED APPLICATIONS AND POTENTIALLY HARMS OTHER APPLICATIONS

CTIA - The Wireless Association’s 2014 paper *Net Neutrality and Technical Challenges of Mobile Broadband Networks*[3] suggests an approach under which, in our estimation, only applications that are very tightly integrated with a wireless carrier will be able to get QoS—and only those applications will work properly in most circumstances, particularly in the midst of network congestion. CTIA’s paper also suggests that QoS on mobile broadband networks is so complex

[1] In this context, “like” applications have common packet delivery, interactivity, and bandwidth requirements and typically present similar forms of media. Examples include toll-quality voice, video streaming (potentially with separate categories for different resolutions), audio streaming, and interactive video.

[2] See generally Barbara van Schewick, “Network Neutrality and Quality of Service: What a Non-Discrimination Rule Should Look Like,” Stanford Law School, Public Law and Legal Theory Working Paper Series, No. 2459568 (June 2014), forthcoming, Stanford Law Review, Vol. 67, Issue 1 (2015), at 2-3.

[3] Jeffrey H. Reed and Nishith D. Tripathi, *Net Neutrality and Technical Challenges of Mobile Broadband Networks.* Paper, Washington, D.C.: CTIA - The Wireless Association, 2014. http://www.ctia.org/docs/default-source/default-document-library/net-neutrality-and-technical-challenges-of-mobile-broadband-networks-9.pdf and http://apps.fcc.gov/ecfs/document/view?id=60000871627 (accessed Oct. 22, 2014).

that there is no viable approach other than to let wireless carriers determine how and whether to offer it. And the paper provides, as a sole example of QoS, a video conferencing system in which QoS is provided in a framework managed in total by the carrier, without a process for providing QoS to unaffiliated applications.

Contrary to the suggestion of the CTIA paper, this is not the only feasible framework for QoS in a mobile environment. Prioritization techniques can be targeted to the places and times that congestion undermines QoS. Depending on the level of this congestion, the mobile carrier has a variety of options, some of which are application-agnostic, and none of which should require treating like applications differently.

1.3 DIFFERENT LEVELS OF NETWORK CONGESTION DEMAND DIFFERENT APPROACHES TO PRIORITIZATION

As a starting point, it is critical to distinguish the very different challenges that confront any mobile broadband network depending on whether a cell or sector is not congested, moderately congested, or severely congested.

First, at many times and places, the capacity of mobile broadband networks is ***not congested*** and there is little if any need to prioritize certain data traffic over other traffic. The reality today is that despite the powerful and flexible QoS capabilities of LTE technology (described in Section 2), the majority of mobile broadband traffic currently travels without QoS, as "best effort" communications, including video applications such as YouTube and over-the-top voice like Viber and Skype. Even in congested environments, these applications are not prioritized over less sensitive applications such as e-mail, device updates, or file downloads.

Second, at times and places ***when congestion is moderate,***[4] it is suitable to limit mobile broadband network prioritization to a manner that does not discriminate among specific uses of the network or classes of uses. In this case, LTE networks are capable of allocating bandwidth among users who have control over prioritization—for example, by offering users the option to purchase a premium level of service that ensures faster throughput for whatever application the user chooses to use. This type of prioritization has the benefits of 1) enabling users who need or want faster or more consistent service to be able to obtain it, 2) making prioritization open and transparent, and 3) not discriminating among edge providers.

[4] In this context, moderate congestion is defined as congestion that will not impact the ability of a latency-sensitive application to function.

Finally, at times and places of ***severe congestion***, many or all latency-sensitive applications may deteriorate or fail. In these situations, LTE is entirely capable of prioritization that treats similar applications the same way, regardless of the identity of the edge provider or whether the application is affiliated with the wireless carrier.

1.4 A NEUTRAL FRAMEWORK WILL TREAT LIKE APPLICATIONS IN A LIKE FASHION

1.4.1 CASE 1: MODERATE CONGESTION—APPLICATION-AGNOSTIC MANAGEMENT

As the FCC declared in its 2010 Open Internet Report & Order, "[u]se-agnostic discrimination (sometimes referred to as application-agnostic discrimination) is consistent with Internet openness, because it does not interfere with end users' choices about which content, applications, services, or devices to use. Nor does it distort competition among edge providers."[5] The Commission suggested "end-user control" (i.e., user-directed prioritization) as a reasonable tool to manage network capacity constraints.[6]

We agree that QoS can be limited to application-agnostic discrimination when congestion is moderate—that is, when congestion will not impact the ability of a latency-sensitive application to function. This Report describes how LTE networks are tailor-made to adopt this technique as an alternative to prioritizing (or throttling) particular applications, or classes of applications, relative to others.

1.4.2 CASE 2: SEVERE CONGESTION—"LIKE MANAGEMENT OF LIKE APPLICATIONS"

During periods of severe congestion, the Commission may consider the prioritization of certain categories of applications (particularly delay-sensitive applications such as voice or video telephony) to be reasonable network management.

---

[5] Report and Order, "In the Matter of Preserving the Open Internet," Para. 73 ("Use-Agnostic Discrimination"), Federal Communications Commission, GN Docket No. 09-191, FCC 10-201, Dec. 23, 2010. https://apps.fcc.gov/edocs_public/attachmatch/FCC-10-201A1.pdf (accessed Nov. 4, 2014).

[6] Report and Order, "In the Matter of Preserving the Open Internet," Para. 71 ("End-User Control"), Federal Communications Commission, GN Docket No. 09-191, FCC 10-201, Dec. 23, 2010. https://apps.fcc.gov/edocs_public/attachmatch/FCC-10-201A1.pdf (accessed Nov. 4, 2014). The Commission stated: "Maximizing end-user control is a policy goal Congress recognized in Section 230(b) of the Communications Act, and end-user choice and control are touchstones in evaluating the reasonableness of discrimination. [citing van Schewick]…[E]nabling end users to choose among different broadband offerings based on such factors as assured data rates and reliability, or to select quality-of-service enhancements on their own connections for traffic of their choosing, would be unlikely to violate the no unreasonable discrimination rule, provided the broadband provider's offerings were fully disclosed and were not harmful to competition or end users."

In this neutral framework, an unaffiliated application can be provided QoS; “like” applications are treated in a “like” fashion; and the LTE network provides QoS to latency-sensitive and/or high-bandwidth applications that benefit from it or require it to run consistently and effectively in congested environments.

1.5 A NEUTRAL-MANAGEMENT APPROACH CAN BE IMPLEMENTED NOW

The 2010 Open Internet Order stated that an application-agnostic QoS approach would be ideal. An application-agnostic model that prioritized certain users over others would be feasible through the LTE framework—but would also have some tradeoffs, including prioritization of applications that are not latency-sensitive and do not benefit from the prioritization—at the expense of non-prioritized users on their latency-sensitive applications. For conditions of moderate congestion, we propose an approach in which wireless carriers offer users one or more enhanced levels of prioritization and service with transparently reported performance metrics. This model will offer customers some improvements in performance in conditions of moderate congestion (see Sections 2.4 and 3.1).

For conditions of severe congestion, when latency-sensitive applications are impaired or no longer work, we propose another neutral approach that also functions within the architecture of LTE and within the structure and philosophy of the Internet. Indeed, this alternative is neither novel nor radical; rather, the approach proposed here is feasible, available today, and falls within the capabilities provided by 3GPP LTE standards (see Section 3.2 *et seq.*).

Our example approach (or a similar one) can be implemented now using standards-compliant technologies as the needed processes are developed in detail.

This approach entails the following steps:

1) Industry standards bodies or another industrywide process approved by the FCC create generic QoS profiles related to latency sensitivity or other attributes that need similar QoS treatment, and make them open to all like applications, such as toll-quality voice and video communications.

2) Wireless carriers define the type of network management each profile will receive, understanding that the management may be dynamic and complex, but that all like applications within the profile will receive the same treatment.

3) The FCC or standards bodies create a streamlined process through which edge providers can identify their content and applications to the wireless carriers for treatment according to a QoS profile.

4) The FCC or standards bodies create a process, such as periodic audit of active QoS rules, to transparently verify that the defined management structure is being implemented consistently.

5) The FCC or standards bodies revisit the profiles regularly, and revisit the need for QoS and prioritization as spectrum efficiency increases and other technological improvements enter the marketplace.

The coming standards evolution will provide for an even more streamlined and automated process of application identification in the next few years and will further broaden the wireless carriers' options for delivering QoS equally to edge provider applications and carrier-delivered applications.[7] It is important to note, too, that many of the further streamlined approaches in the standards roadmap are already available today from network technology companies that provide systems to wireless carriers; the approaches exist but are not yet part of the 3GPP standard.[8]

1.6 LTE TECHNOLOGIES AND STANDARDS FACILITATE NON-DISCRIMINATORY POLICY RULES FOR MOBILE DATA TRAFFIC MANAGEMENT

CTIA contends that mobile networks are so significantly different from fixed networks that they cannot be subject to the same, or even comparable, regulation.

While it is true that there are marked differences between the technical functionality and capability of fixed and wireless networks, there exist technological options for managing mobile networks that would allow equitable QoS treatment for like applications. LTE technologies and 3GPP LTE standards facilitate—and certainly do not preclude—non-discriminatory policy rules for mobile data traffic management. Network management strategies should take into consideration the capabilities and limitations of each type of network, and use the appropriate means within the constraints of the associated network technology and the capabilities

[7] 3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Policy and Charging Control (PCC); Reference points (Release 11), 3GPP TS 29.212 V11.13.0 (2014-06), p. 69-84. http://www.3gpp.org/DynaReport/29212.htm (accessed Nov. 4, 2014).
[8] Sandvine, "Network Policy Control and the Migration to LTE," p. 21. https://www.sandvine.com/downloads/general/whitepapers/network-policy-control-and-the-migration-to-lte.pdf, (accessed Nov. 4, 2014). *See also:* Cisco, "Cisco ASR 5000 Series Application Detection and Control Administration Guide," Version 14.0, 2014. http://www.cisco.com/c/dam/en/us/td/docs/wireless/asr_5000/14_0/OL-27200_ADC_Admin.pdf (accessed Nov. 4, 2014).

outlined in technical standards to enact policies that enable a neutral experience for the users of like applications on the network.

2. TECHNOLOGICAL OPTIONS FOR NETWORK MANAGEMENT

Network management refers to the process of operating and administering a network through a variety of mechanisms to manage network traffic and congestion. It can be thought of as a complex series of knobs that can be adjusted manually or that can operate automatically based on pre-determined variables and criteria or real-time needs, enabling control over the flow and direction of data.

The purpose of network management is to ensure that, in the face of congestion, users are able to obtain access to the network, and that applications will run consistently and meet consumer expectations in most cases. Any network management strategy will eventually fail in the event of saturation by very heavy congestion or interference. However, in the face of moderate congestion, well-executed management can allow the most critical network uses to continue and can prioritize the most sensitive applications so that they can continue operating, while less sensitive applications can be delayed with no impact or minimal impact to the users.

Reasonable network management includes both consistent treatment of applications and transparency. These are essential for keeping the network functional. Consistent treatment of applications entails treating all “like” applications the same, regardless of their origin or developer.

This section of the Report discusses the means of QoS in an LTE environment.

LTE was designed from the outset to provide capability to manage the network in a flexible way. The LTE design and standards assume that growth in network usage necessitates ways to manage capacity as a complement to techniques for growing capacity, including use of additional spectrum, small cells, advanced multiple-input multiple-output (MIMO) technologies, other antenna and spectrum enhancements, and handoff to unlicensed Wi-Fi networks. Network management is especially needed in places and times when networks become congested, such as in crowded areas, during special events or emergencies, or with the emergence of new applications.

In the LTE environment, all decisions regarding QoS are made and managed by the wireless carrier. Prioritization within the LTE network is entirely under the carrier’s control, and the carrier’s prioritization settings take precedence, even if an application developer or outside network applies prioritization in the form of class of service tags or another identifier on communications to or from an application, or if an outside Internet Service Provider (ISP) or connecting network otherwise prioritizes traffic.

2.1 THE COMPONENTS OF AN LTE NETWORK

Figure 1 (see below) illustrates the relevant components of the LTE network in the current LTE environment adopted by most wireless carriers (which includes systems in compliance with 3GPP Release 9 or later). The explanation is deliberately simplified and focuses on the main parts of the network and the components needed to set up and manage QoS.

On the left is the user equipment (UE), which can be a smartphone, tablet computer, router, or any other user-operated device. The next block is the radio access network (RAN), comprising the antennas and cell sites (also known as eNodeB) and the interconnection between the sites, usually over wireline fiber optic cable. The next three rows of blocks, in the middle, include several components that are part of the LTE evolved packet core (EPC), which are usually situated in a central site managed by the wireless carrier—a facility that can manage thousands of cell sites and hundreds of thousands of user devices. Finally, on the right is a server operated by an edge provider, managing and hosting an application or content. The edge provider may be a large company, such as Google or Netflix, but may also be a small technology company with a new application—or a group of individuals independently hosting video content, using off-the-shelf technology and not necessarily providing an application to the user.

Within the EPC in the center of the architecture are the Policy and Charging Rules Function (PCRF), the Packet Gateway (PGW) (including the Policy and Charging Enforcement Function (PCEF)), the Home Subscriber Server (HSS), the Mobility Management Entity (MME), and the Serving Gateway (SGW).

Given policy rules provided by the wireless carrier, the PCRF stores all information related to QoS and is responsible for initiating the setup of dedicated bearers—virtual channels with designated QoS for QoS-specified communications—for each user. Bearers are described in Section 2.2, and examples of rules are provided in Section 2.3.

The PCRF is typically programmed by the wireless carrier, but can also be configured to receive policy rules from an edge provider through the Rx interface (a trusted interface).[9] The PCRF also receives charging information from the network and can provide charging information back to

[9] 3rd Generation Partnership Project (3GPP); Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 9), 3GPP TS 23.203 V9.13.0 (2013-09). http://www.3gpp.org/DynaReport/23203.htm (accessed Nov. 4, 2014).

the edge provider. In this way, the PCRF is capable of fulfilling its role for applications operated and managed by the wireless carrier or for applications provided over-the-top by edge providers.

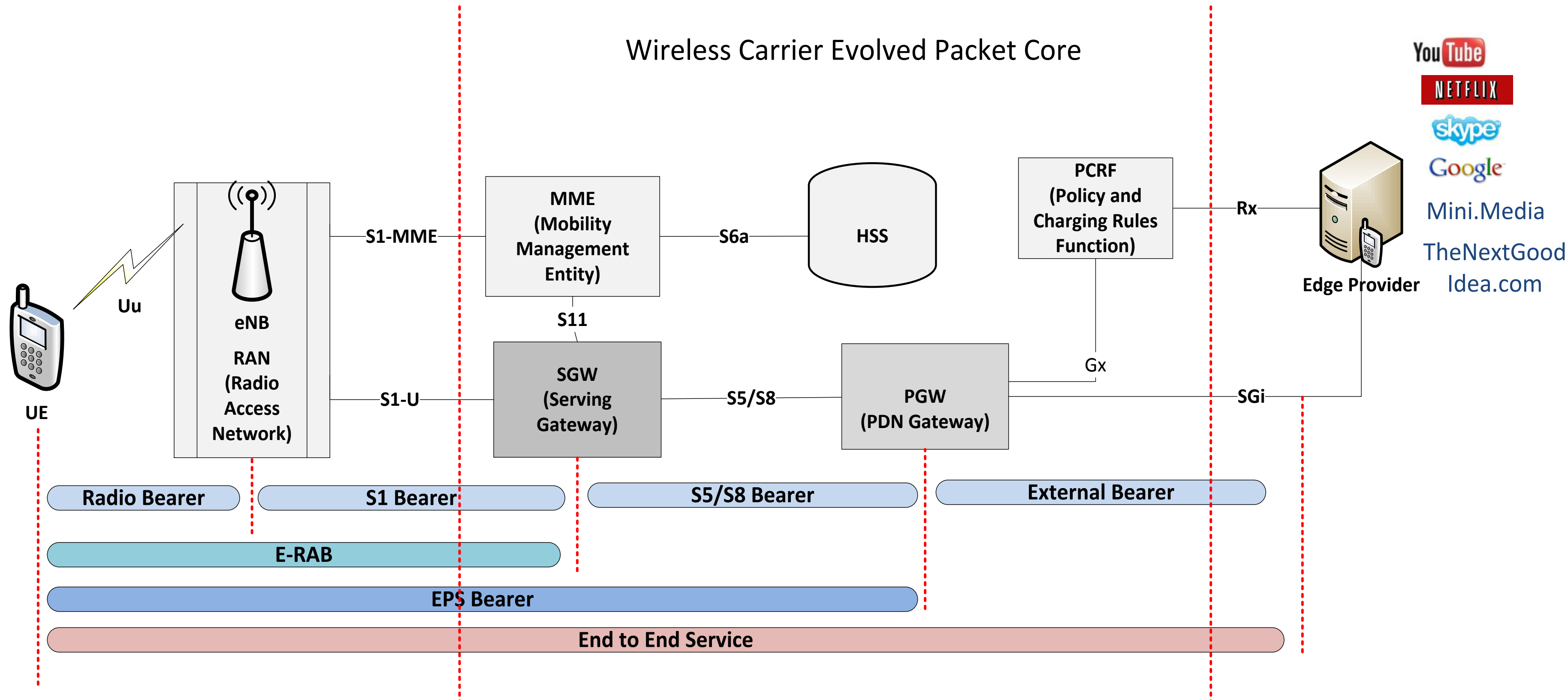
The PCEF implements and enforces the rules stored and initiated by the PCRF. It is part of the PGW, which is the termination point of outside networks (including the Internet). The PGW is also able to identify packets entering the network based on application or other criteria. Traffic Detection Function (TDF) and Application Detection and Control (ADC) functionality are currently available in many implementations of the PGW, and are also available as a standalone function between the PGW and the outside networks.[10] TDF and ADC technologies are expected to be standardized within 3GPP in the next two years. TDF/ADC can identify packets and also classify them into predetermined QoS profiles.

It is important to note that TDF/ADC is not a critical component of this model, but can be an enhancement. TDF/ADC can automate identification of latency-sensitive communications and applications, but this can also be done by other means discussed in Section 3.4.

The SGW is the termination point of the RAN in the core—that is, it is the point in the core that connects out to the antennas and cell sites (eNodeB). The SGW directs traffic to and from one or more eNodeBs and the user devices.

[10] Sandvine, “Network Policy Control and the Migration to LTE,” p. 21.

Figure 1: Architecture of Selected LTE Components Needed for QoS Management



The MME is responsible for managing the part of the network from the core to the user devices, inclusive of the eNodeB sites. It has many roles, including keeping track of the location of active or idle user devices, determining the paths of bearers within the mobile network, and configuring QoS for those bearers.

The MME and PCRF are part of what is called the "control plane"—the part of the network that makes all of the connections happen. The "user plane" is the part of the network where the user data travels. In

Figure 1, the control plane is the top half, and the user plane is the bottom half.

To summarize the component QoS roles in brief, the control plane stores the rules and translates those rules into bearers. The user traffic on the network travels through bearers set up in the user plane.

There is feedback between the user and control planes as users and edge providers request and terminate resources, congestion mounts and subsides, and the wireless carrier tracks network usage by bearer, users, and other parameters.

2.2 HOW QOS IS PROVIDED OVER AN LTE NETWORK

QoS for bearers can be implemented in a static or dynamic way. Static implementation uses set rules. Dynamic implementation can be associated with a range of categories that can emerge on the network, such as congestion, activation of devices, interference, and rogue devices, among others. Dynamic is not a different set of rules, but is simply the dynamic application of static rules.

As discussed, bearers are virtual channels from one point to another on the network. The bearer is virtual, in that it does not take up a fixed amount of spectrum[11] and only uses network capacity when the user device or network sends information through it. Each bearer also has QoS settings. A user device can connect through multiple bearers.

The default bearer is configured at first connection of the device. It typically uses "best effort" QoS. If the device needs to operate an application or service with a particular specified QoS,

[11] LTE networks assign capacity in combinations of channel spectrum and time slots that are dynamically reassigned as needed. The capacity used by a non-Guaranteed Bit Rate (non-GBR) bearer will expand and contract within this frequency and time space. A GBR bearer will take up fixed capacity while it is in use.

such as a toll-grade voice application, the network can set up a "dedicated" bearer to support that QoS level.

An Evolved Packet Switched (EPS) bearer is a type of bearer (which could be default or dedicated) that traverses the network from the user device to the Packet Gateway (the PGW, where the wireless carrier meets the Internet and other outside networks).

QoS bearer service architecture is defined in the 3GPP TS36.300 Section 13.1;[12] bearer service architecture is shown in the bottom portion of

Figure 1. Service data mapped to the same EPS bearer all receives the same bearer-level QoS treatment (e.g., scheduling policy, queue management policy, rate shaping policy).

As currently configured, LTE options for providing QoS at the bearer level include:

1. QoS Class Identifier (QCI)
2. Allocation and Retention Priority (ARP)
3. Guaranteed (minimum) Bit Rate (GBR)
4. Aggregate Maximum Bit Rate (AMBR)

Each of these is described in the sections below.

2.2.1 QOS CLASS IDENTIFIER

QoS Class Identifier (QCI) levels as defined by standards are provided in

Table 1. QCI was designed to put different types of applications and traffic into QoS categories suited to groups of applications, and to specify latency and packet loss parameters for the groups. QCI categorization makes it possible to organize the way network components manage QoS; allows manufacturers to optimize equipment to work well for those categories; and makes it possible for wireless carriers to avoid having to customize QoS in a very granular and different way for specific applications within the group of application. The wireless carrier configures

[12] 3rd Generation Partnership Project (3GPP); Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 9), 3GPP TS 36.300 V9.10.0 (2012-12), Section 13. http://www.3gpp.org/dynareport/36300.htm (accessed Nov. 4, 2014).

eNodeB equipment so that the network as a whole manages packets according to the QCI guidelines in

Table 1. Nodes are configured to more granular parameters, including "scheduling weights, admissions thresholds, queue management thresholds, [and] link-layer control configuration."[13]

Table 1: Priority, Delay, and Loss Parameters According to QCI

QCI	Guarantee	Priority	Delay budget	Loss rate	Example Application
1	GBR	2	100 ms	1e-2	VoIP
2	GBR	4	150 ms	1e-3	Video call
3	GBR	3	50 ms	1e-3	Real time gaming
4	GBR	5	300 ms	1e-6	Streaming
5	Non-GBR	1	100 ms	1e-6	IMS signaling
6	Non-GBR	6	300 ms	1e-6	Streaming, TCP
7	Non-GBR	7	100 ms	1e-3	Interactive gaming
8	Non-GBR	8	300 ms	1e-6	Streaming, TCP
9	Non-GBR	9	300 ms	1e-6	

An example of QCI is QCI=4, geared to streaming. It provides both a guaranteed minimum bit rate (so that quality does not fall below a certain level) and a maximum latency (more tolerant than voice latency, geared toward buffering within streaming applications). It is based on the behavior of common video streaming applications, and developers can gear media applications that they create for this level of delivery over networks.

2.2.2 ALLOCATION AND RETENTION PRIORITY

Allocation and Retention Priority (ARP) identifies the priority of a bearer in the event that congestion forces the network to drop users in a sector or decline access to a user entering a sector. Bearers are assigned an ARP between 1 and 15. A value of 1 means highest priority

[13] U.S. Department of Commerce – Boulder Laboratories, Public Safety Communications Research Program, "Priority, Preemption, QoS," Presentation, p. 7. http://www.pscr.gov/projects/broadband/700mhz_demo_net/meetings/inaug_stakeholder_mtg_042010/day_2/5_Priority_Preemption_and_QoS-final.pdf (accessed Oct. 16, 2014).

whereas 15 means no priority. The ARP level determines whether a bearer can be denied connection in congestion, or whether a bearer is vulnerable to being dropped to manage congestion. ARP is also used for admission control to a cell sector.

2.2.3 GUARANTEED BIT RATE

In Table 1, bearers with a QCI between 1 and 4 are Guaranteed Bit Rate (GBR) bearers. These are associated with a minimum bit rate, which is provided as long as there is sufficient capacity within the network. GBR is for applications that cannot tolerate interruptions.

2.2.4 AGGREGATE MAXIMUM BIT RATE

Aggregate Maximum Bit Rate (AMBR) specifies the maximum upload and download speed associated with the bearer. There are two types of AMBR designation.

User Equipment AMBR (UE-AMBR) limits the aggregate bit rate that can be provided across all non-GBR bearers of a user device—effectively limiting the speed that a user device can have for all services, not including services for which the wireless carrier might have created GBR bearers (for example, toll-quality voice).

Access Point Name AMBR (APN-AMBR) limits the aggregate bit rate that can be provided across all non-GBR bearers on the same packet data network (PDN). This can be used to limit the capacity in part of the network to be allocated to a particular type of application, a particular application, or a particular source or destination (e.g., website).

2.3 ASSIGNING QOS TO TRAFFIC

If a wireless carrier wishes to assign a type of QoS (QCI, ARP, GBR, AMBR) to a particular type of traffic (e.g., toll-quality voice, many types of video streaming), one widely available and used option is to place those rules in the PCRF, which stores and authorizes the rules as needed.[14]

The PCRF is at the center of the QoS framework. The PCRF takes part in initial user device attachment to the network and in the establishment of additional resources, if necessary to support the use of applications by the user device application session. The PCRF decision-

[14] As defined in the 3rd Generation Partnership Project (3GPP); Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 9), 3GPP TS 23.401 V9.15.0 (2013-03), Sec 5.4. http://www.3gpp.org/DynaReport/23401.htm (accessed Nov. 4, 2014).

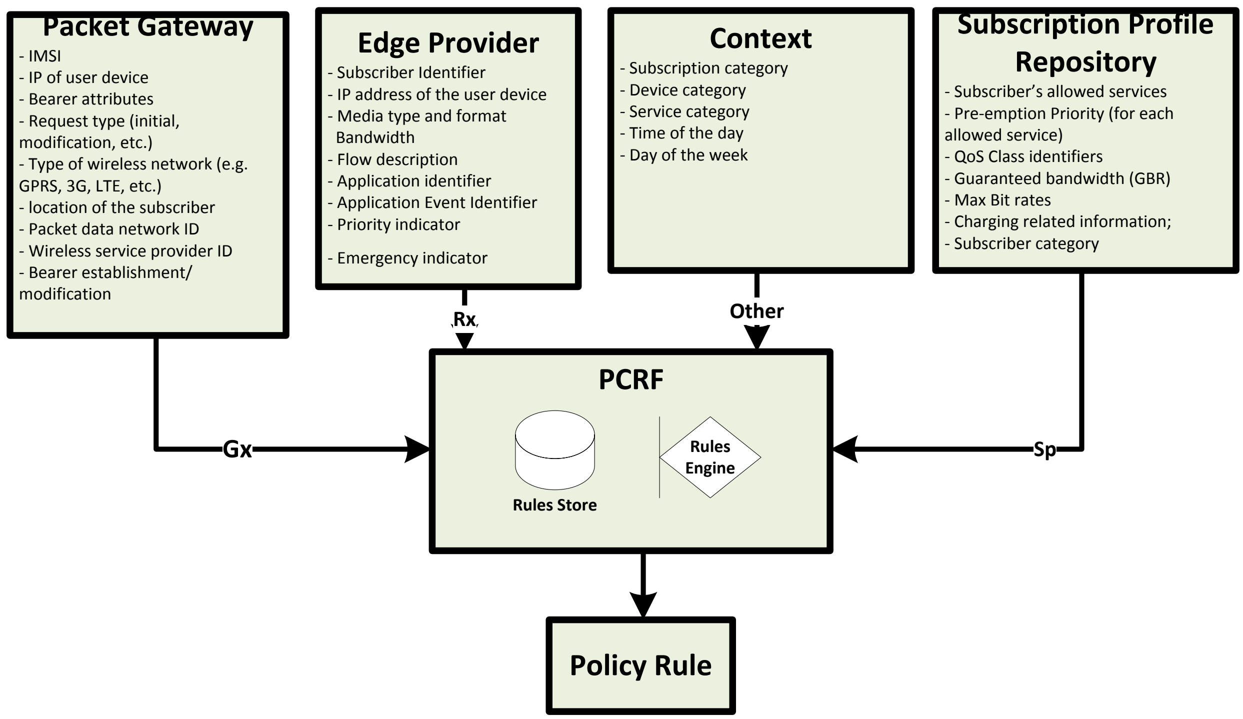
making (rules) engine stores the operator rules and combines them with real-time network information, as well as subscriber and service information. It synthesizes all of this information to make a decision on whether a particular policy change request can be authorized.

An application, whether it is hosted by an edge provider or by the wireless carrier, can send a QoS service request directly to the wireless carrier's PCRF using the standardized Rx interface—potentially to add or modify a policy rule pertaining to the application. In this case the PCRF could combine the application request with per-subscriber information to formulate a policy decision that specifies the QoS level to be granted. (The Subscription Profile Repository, or SPR, contains all of the subscriber subscription related information that the PCRF needs to authorize a QoS request. The PCRF communicates with the SPR to obtain UE subscription information.)

Alternatively, the edge provider could send the request for QoS treatment to the wireless carrier through a separate registration process, with the wireless carrier remaining in full control of policy rule creation and modification. This would require implementation of application detection and control (ADC) functionality, which enables the wireless carrier to implement a variety of rules based on network conditions (e.g., congestion, time, and day of the week). Wireless carriers can manage like applications using the same set of rules so as to maintain the like end user experience.

Figure 2 shows the information the PCRF receives or requests from various entities to make a QoS policy decision.

Figure 2: PCRF Analyzing Requests and Network Data and Provisioning a QoS Policy Rule



The PCRF can configure a policy rule on the PGW or on the user device as a static or dynamic policy.

Two types of procedures are used to apply policy rules:[15]

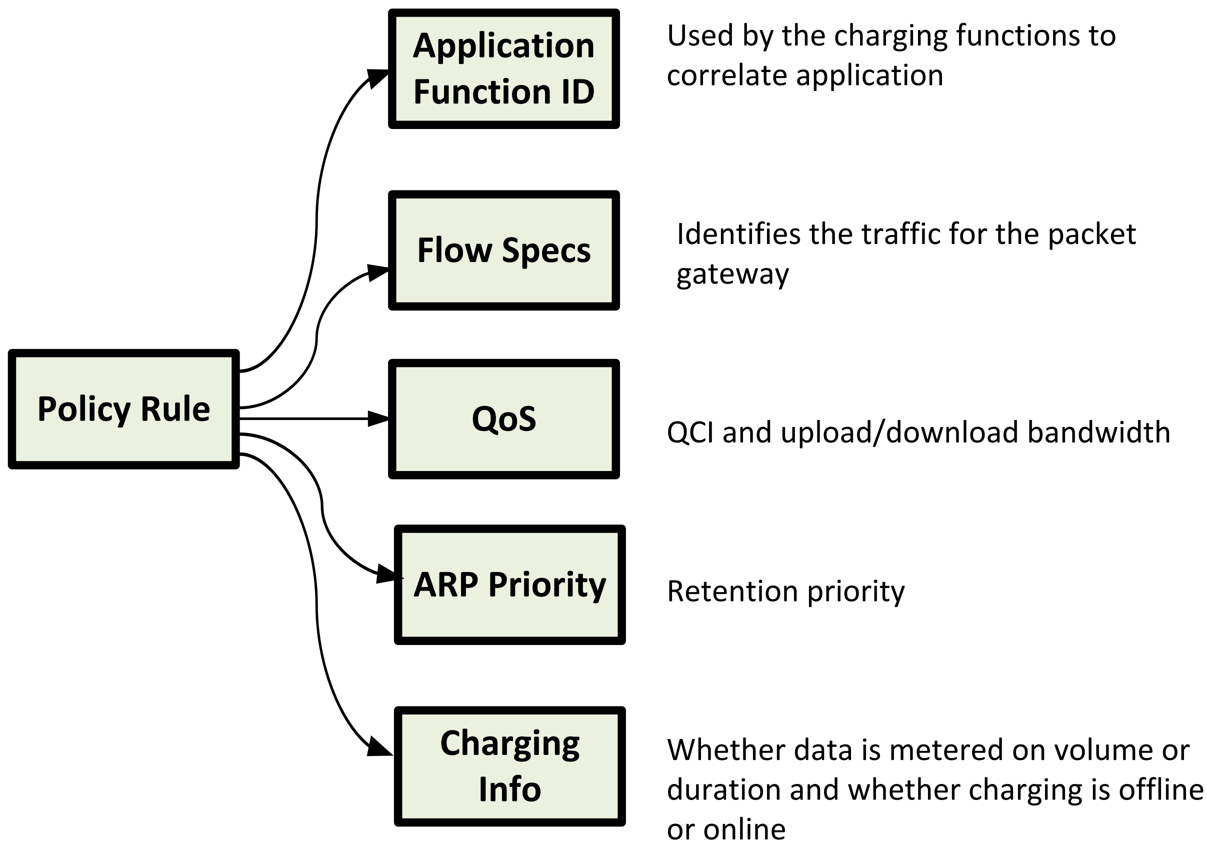
1. PULL Procedure—when a request is made by the network through the PGW. For example, the network may detect an application or IP address that requires special QoS treatment. The user device can also request QoS by requesting a bearer with a specific QoS profile.

2. PUSH Procedure—when the PCRF provisions a rule, for example in response to a request from the application server or in response to a change in the subscription status of a subscriber.

[15] 3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Policy and Charging Control (PCC); Reference points (Release 11), 3GPP TS 29.212 V11.13.0 (2014-06), p. 27. http://www.3gpp.org/DynaReport/29212.htm (accessed Nov. 4, 2014).

Based on the decision model shown above, the PCRF provides the policy rule to the PGW to indicate QoS and charging treatment for a specific user session. Figure 3 shows the high level information that is contained in a policy rule sent to PGW to be used by the PCEF to enforce the rule.

Figure 3: Elements of an LTE Policy Rule



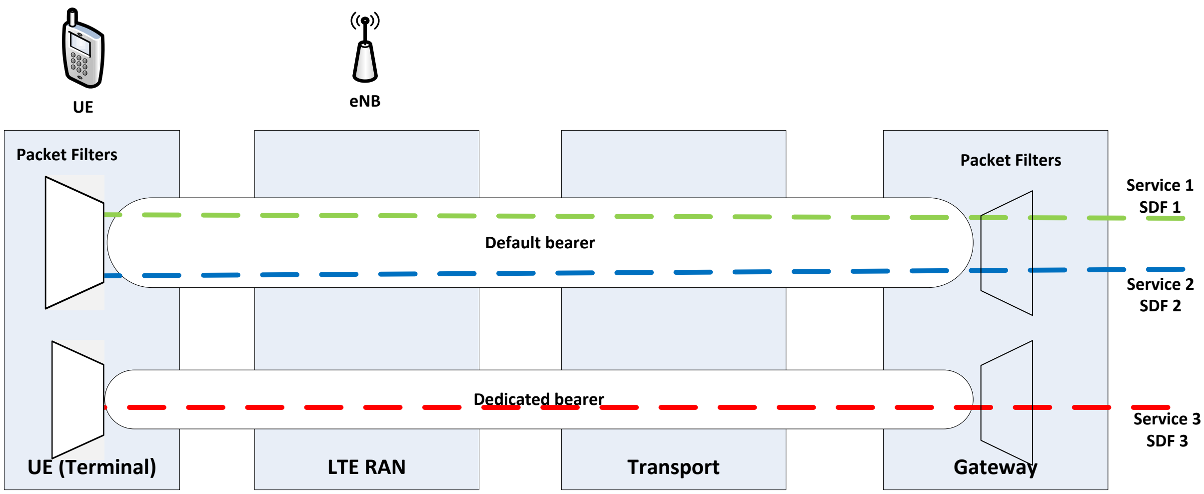
The PCRF forwards rules to data filters also known as traffic flow templates (TFT), located in the appropriate part of the data network. TFT are in place both at the user device (for the uplink direction) and in the PGW (for the downlink direction).

TFTs filter each individual “flow” of data—that is, each set of communications with a source or destination on the network or in the Internet. In LTE, these are called Service Data Flows (SDF).

Based on the TFT, each SDF is placed into a bearer corresponding to the correct QoS treatment. For example, if a bearer is established for the streaming example described in Section 2.2.1, each SDF that fits that category in the TFT is placed in that bearer.

Figure 4 below depicts an example of two bearers in an LTE network. The top bearer is the default bearer set up for each active user device. The bottom bearer is a dedicated bearer set up for a distinct QoS profile. SDF 1 and 2 are sorted into the default bearer for best-effort treatment. SDF 3 is sorted into the dedicated bearer because the rule used by the TFT places it in that bearer for different QoS treatment.

Figure 4: Data Carried in Service Flows Assigned to Bearers Based on QoS Need



2.4 CURRENT QOS ENVIRONMENT AND PRACTICES

Despite all of the QoS capability described above, the majority of wireless broadband traffic currently travels without QoS, as “best effort” communications, including video such as YouTube and over-the-top voice like Viber and Skype. Even in congested environments, these applications are not prioritized over less sensitive applications, such as e-mail, device updates, or file downloads.

However, application-specific network management is growing significantly. One example is voice-over-LTE (VoLTE) as wireless carriers migrate voice calls to LTE from traditional 2G/3G voice infrastructure. Another example is hosted push-to-talk applications carried over LTE, such as Sprint Direct Connect. Different wireless carriers provide these applications in different ways. AT&T, Verizon Wireless, and T-Mobile use an IP Multimedia System (IMS) based approach where the voice application server is tightly integrated with the wireless carrier network. Sprint is currently using a different approach that leverages its 3G networks for (non push-to-talk) voice.

LTE QoS is also a matter of intensive interest in the planning and design of FirstNet, the national wireless public safety broadband network. FirstNet is planning its network for continuity in major public safety incidents, so that incident commanders can prioritize communications, and so that sensitive media applications, including potentially mission-critical voice, can coexist with other network uses. FirstNet is also planning to serve primary and secondary users, and needs to be able to prioritize primary users over secondary users, or to preempt secondary users if necessary.

FirstNet will also need to use commercial carriers' RAN resources for service outside its deployment area, and will need appropriate prioritization for public safety-grade services. In coordination with the development of FirstNet, the Public Safety Communications Research Program (PSCR) of the U.S. Department of Commerce has initiated a working group on QoS for public safety on LTE, has prepared a draft QoS template, and is including QoS in its lab and field tests.[16]

One form of prioritization currently used by wireless carriers is "throttling" of users. Throttling is placing limitations on the speed of users under certain circumstances, such as when the user exceeds a threshold of data usage. Throttling can be done in LTE, for example through a policy rule that establishes a particular UE-AMBR for users that are identified as having exceeded the threshold.[17] Throttling can also be done outside the LTE portions of the network—for example, on the "wired" part of the network between the PGW and the Internet.

An alternative form of prioritization is user-directed. LTE technology permits carriers to offer differentiated tiers of service to subscribers that can include a "premium" service that prioritizes an individual subscriber's traffic in times of congestion. For example, earlier this year the Austrian mobile carrier Drei announced it would offer this user-directed prioritization at varying premium service tiers beginning in June 2015.[18]

Wireless carriers are in the early stages of determining what can and should be done using LTE prioritization, and the industry is in the early stages of deploying the technology. What this means is that prioritization as described in the CTIA paper will become increasingly common.

[16] U.S. Department of Commerce – Boulder Laboratories, Public Safety Communications Research Program, "Priority, Preemption, QoS," p. 7.
[17] Jon Brodkin, "Analysis: Wireless data caps more about profit than congestion: Network congestion is real, but monthly data caps don't prevent it," *Ars Technica*, Oct. 8 2014. http://arstechnica.com/staff/2014/10/analysis-wireless-data-caps-more-about-profit-than-congestion/ (accessed Nov. 4, 2014).
[18] "LTE-Leistungsklassen statt Drosselung?" *LTE-Anbieter.info*, October 17, 2014, http://www.lte-anbieter.info/lte-news/lte-leistungsklassen-statt-drosselung (accessed Nov. 12, 2014)

The outcome may be that the carrier-affiliated application with prioritization will work, while a competing edge provider application will not.

3. NEUTRAL MANAGEMENT WITH LTE TECHNOLOGY

This section of the Report proposes an approach that would serve to:

1. Provide application-agnostic prioritization in situations of **moderate congestion**, with the wireless carrier prioritizing users seeking a premium service.

2. Enable like applications to be treated in a like way for purposes of QoS during periods of **severe congestion**, with the wireless carrier determining the mechanics of prioritization for a given class of applications as defined by standards bodies or the FCC.

This approach and others can be implemented within the framework of the LTE standard and within the capabilities of LTE technology. Indeed, it would be counterproductive to create an environment that diverges unnecessarily from other industry efforts or creates arbitrary minimum levels of service based on wireline service norms. Therefore, the proposed framework described below draws on current practices, the LTE standards, work currently underway to create QoS for mission-critical FirstNet applications, and the development of QoS technologies and standards that can include independent edge providers as well as wireless carriers.[19]

3.1 APPLICATION-AGNOSTIC PRIORITIZATION IN MODERATE CONGESTION

As discussed in Section 2.4, wireless carriers are prioritizing users when they "throttle" capacity for users who have exceeded capacity limits. This type of prioritization is already commonly implemented and widely understood. Application agnostic prioritization can also be implemented in a more affirmative, user-directed fashion, such as by offering subscribers the option to pay a premium for prioritization at times of congestion, as noted in the previous section.

One approach for application-agnostic prioritization is for the wireless carrier to identify the users who have registered for prioritization within the SDR. The wireless carrier would create

---

[19] Informa, "OTT communication: Monetization opportunities for operators and OTT players," webinar. http://www.informatandm.com/ott-communication-monetization-form/, (accessed Nov. 4, 2014).

policy rules that would activate the prioritization under specified conditions of congestion (Figure 2). For example, the rules could assign a different QCI to the premium traffic (QCI=8 instead of the "best effort" QCI=9) and could assign an AMBR priority to drop the non-premium traffic before the premium traffic.

In any case, application-agnostic prioritization could be done in a static or dynamic manner, with the details available to the public in transparently-assigned rules. (Transparency is discussed in more detail in Section 4). The PCRF would be responsible for identifying when prioritization should happen in response to the rules and set up dedicated bearers for the prioritized communications.

We note again that in this scenario, all communications by the prioritized users receive the same priority. There is no differentiation between applications or uses of the network or among edge providers.

3.2 PRIORITIZATION OF LIKE APPLICATIONS WITH LIKE PRIORITIZATION IN SEVERE CONGESTION

As discussed previously, in severe congestion, latency-sensitive applications become impaired or stop working. We outline below a technical approach for like prioritization of like application under those conditions.

Wireless carriers have begun to use application-specific QoS profiles and will use them more frequently as the utilization of networks increase, as congestion increases, as latency-sensitive and bandwidth-rich applications grow in popularity, and as wireless carriers develop partnerships with select edge providers.

As described above (page 5), industry standards bodies, or other industrywide processes approved by the FCC, should create generic QoS profiles for categories of applications that are consistent (and transparent). These generic profiles should be developed and implemented once the wireless carriers start to offer application-specific QoS for a particular type of application—for example, voice. Since we understand wireless carriers are already offering VoLTE services and operating application-specific QoS for it, voice should clearly be established as a generic profile. With voice established as a QoS profile established, the QoS settings used for the wireless carrier VoLTE service should also be available for use by any edge provider voice applications in a like manner.

We recommend that the QoS profiles be established by standards organizations or the FCC and reviewed by the FCC for consistency with Open Internet principles. *We note that the standards* *organizations are only defining profiles (categories) and not the actual QoS settings within the LTE network.* For example, the standards organizations might establish generic QoS profiles for: “1. Voice; 2. Interactive Video (mobile handheld device); 3. Interactive video (tablet); . . . ”, and provide a high level description of what applications fit that category.

Then, once the wireless carrier began to offer QoS for applications in those categories, the carrier would also be required to offer the same QoS treatment to all like applications in those categories, including over-the-top applications and small edge providers unaffiliated with the wireless carrier.

Implementation of our proposed approach by a carrier includes the following steps:

1. The wireless carrier establishes a set of QoS profiles, with each profile supporting a class of applications that may require like QoS treatment (e.g., streaming, interactive video, toll-quality voice, and so on).

2. An edge provider identifies itself to the wireless network as an operator of an application that should be placed in one of the profiles and receive QoS.

3. A user device activates the application, identifies itself to the wireless network and the edge provider through the default bearer connection, and asks to operate the application through the network.

4. The edge provider authenticates the user device to use the application (if required by the edge provider).

5. The wireless carrier sets up QoS between the user device and the Internet according to the QoS characteristics of the profile.

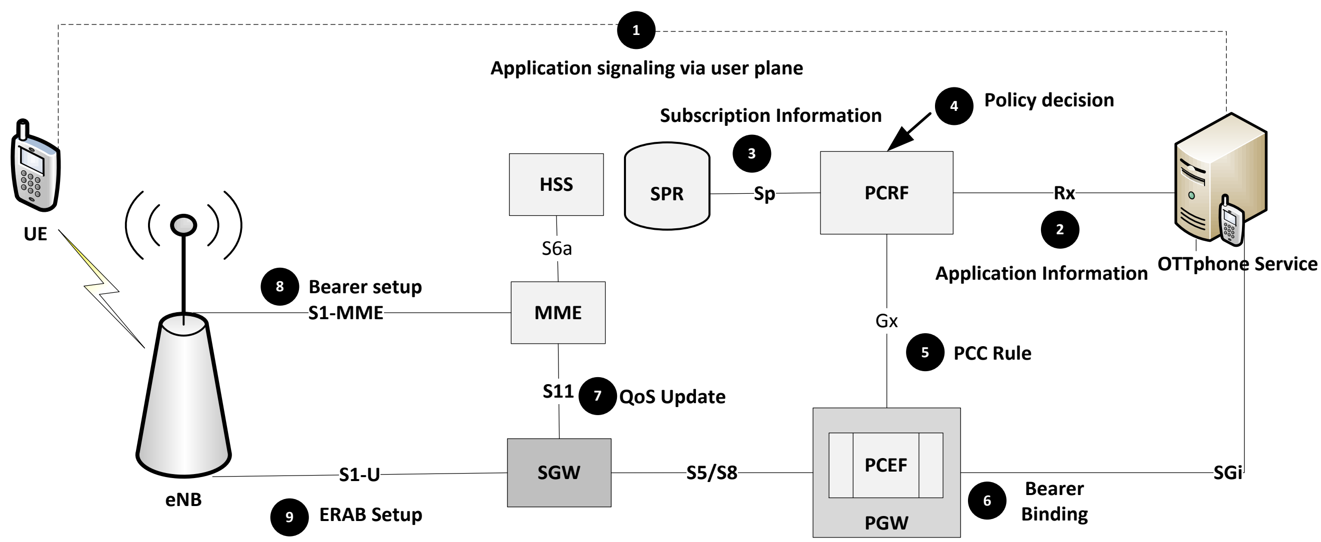
This approach is functionally similar to the example provided in the CTIA paper in the following way: it identifies applications that are designated for QoS based on wireless carrier policy rules, assigns them to a bearer according to those rules, and provides QoS between the user device and the packet gateway.

The approach proposed here, however, can apply to any edge provider with a minimum of coordination with the wireless carrier—and not only to CTIA’s example of a video conferencing technology that is provided by "[a]uxiliary networks [that]…can be viewed as part of the operator's services."[20]

In the language of LTE, our example application is outside the IP Multimedia Subsystem (IMS) framework, and is hosted by an external service provider, with no particular affiliation between the edge provider and the wireless carrier.

The following is a more detailed discussion of the steps outlined above. In this example, the wireless carrier creates a profile for toll-quality voice, an application called "OTTphone" is set up to use the profile, and a user who connects and uses OTTphone obtains the same level of QoS within the wireless network as the toll-quality voice application provided by the wireless carrier (e.g., VoLTE). The steps are shown in Figure 5.

Figure 5: Example QoS Service Provision for Unaffiliated Application on Wireless Network



3.3 SETUP OF QOS PROFILE AND ASSOCIATED NETWORKS

In this step, the wireless carrier establishes a set of QoS profiles, each to support a class of like applications that may require special QoS treatment (e.g., toll-quality voice, video streaming, and so on). These profiles might be based on the delay tolerance of the application or service.

[20] Reed and Tripathi, "Net Neutrality and Technical Challenges of Mobile Broadband Networks," p. 9.

The wireless carrier creates a packet data network (PDN) within its LTE network, a virtual "network within a network" identified with an Access Point Name (APN) associated with toll-quality voice. The toll-quality voice PDN is potentially available to any user device.

Because the specific settings in this model are up to the wireless carrier and not determined by a fixed rule imposed from the outside, they can be dynamic and changed by the wireless carrier to match changing conditions and technology.

For example, the wireless carrier may set the QoS for toll-quality voice with QCI=1, Minimum Guaranteed Bit Rate = 0.080 Mbps, and Allocation and Retention Priority (ARP) = 7, for the first 50 users in the sector. The wireless carrier may also set up entirely dynamic QoS that changes from this QoS to other settings in various conditions, such as areas where less spectrum is available, where there is congestion, or where interference is detected. And the wireless carrier may provide different QoS to the next 50 users trying to use this profile. However, the key thing is that the toll-quality voice profile would stay the same for all application providers using the toll-quality voice profile.

The toll-quality voice QoS settings are created as policy rules and are stored in the PCRF in the wireless carrier network. Individual rules have a specific format. (Elements of a policy rule are shown in Figure 3.) A QoS configuration could include many different rules, each implemented by the PCRF at different times, under different circumstances.

3.4 IDENTIFICATION OF APPLICATION

Identification of the OTTphone application within the LTE network can be accomplished in several ways. One approach would be for the wireless carrier to seek and identify VoIP applications through inspection of data packets—based on source, header information, packet size, or other characteristics of well-known applications in this category. This approach is similar to the way in which carriers identify and filter viruses, spam, and malicious code. However, packet inspection does not always accurately identify an application.

There exists another, potentially more precise option. This would include a brief edge provider registration process, in which the provider of an application like OTTphone provides the wireless carrier with identifying information, such as IP addresses or packet header formats. In this step, an edge provider would identify itself to the wireless network as an operator of an application that should be placed in one of the profiles. This would presumably be an optional process for edge providers deciding not to rely entirely on packet inspection.

It is likely that, as ADC and TDF technologies (Section 2.1) improve, this step may be streamlined or eliminated, with the network instead automatically identifying the latency-sensitive applications and edge providers not needing to register. In addition, packet inspection should always be available as a back-up, because in many cases an online application or service provider that a consumer chooses may not have registered with the mobile ISP.

3.5 ACTIVATION OF APPLICATION

Under this process, a user could potentially first download the OTTphone application to the user device from an application store through the regular best-effort Internet connection using the default bearer. The user would then activate the application, connect with the edge provider, and register with the edge provider, if needed. Then the user device would signal to the OTTphone server (via the OTTphone application, connected using the default bearer) that it needs to make a phone call (Figure 5, step 1).[21]

3.6 AUTHENTICATION OF USER DEVICE

In this step, the OTTphone application would authenticate the user device to use the application, potentially having detected the wireless carrier of the user based on information from the device, the source address of the wireless carrier, or from information in the signup process. OTTphone would then identify the user and device information to the wireless carrier by sending a request to the PCRF with the Application ID, user device identity (e.g., IMSI, IP address), and requested QoS profile. (Figure 5, step 2).

The wireless carrier checks the subscription status of the user (Figure 5, step 3) and applies the rules established for the toll-quality voice profile to the user, user device, and application. The PCRF does this by querying the SPR, making a policy decision based on the information from OTTphone and the SPR, and then sending the rule to the PCEF within the PGW (Figure 5, steps 4 and 5). The PCRF authorizes the PCEF in the PGW to create a dedicated bearer with the toll-quality voice QoS profile.[22]

---

[21] If a user is going to a site to stream music or other content, this step may simply be a query from the media player on the user device to the streaming web site.

[22] The query to the SPR is not necessary if all wireless carrier users are entitled to prioritization or if there is no charging. In that case, the PCRF authorizes the dedicated bearer setup based on the information provided by the OTTphone application.

3.7 ACTIVATION OF QOS

The wireless carrier sets up QoS according to the QoS characteristics of the profile. The PGW binds the user device IP address with the new service data flow corresponding to OTTphone and sends a request to SGW to create a dedicated bearer along with the required QoS profile (Figure 5, step 6).

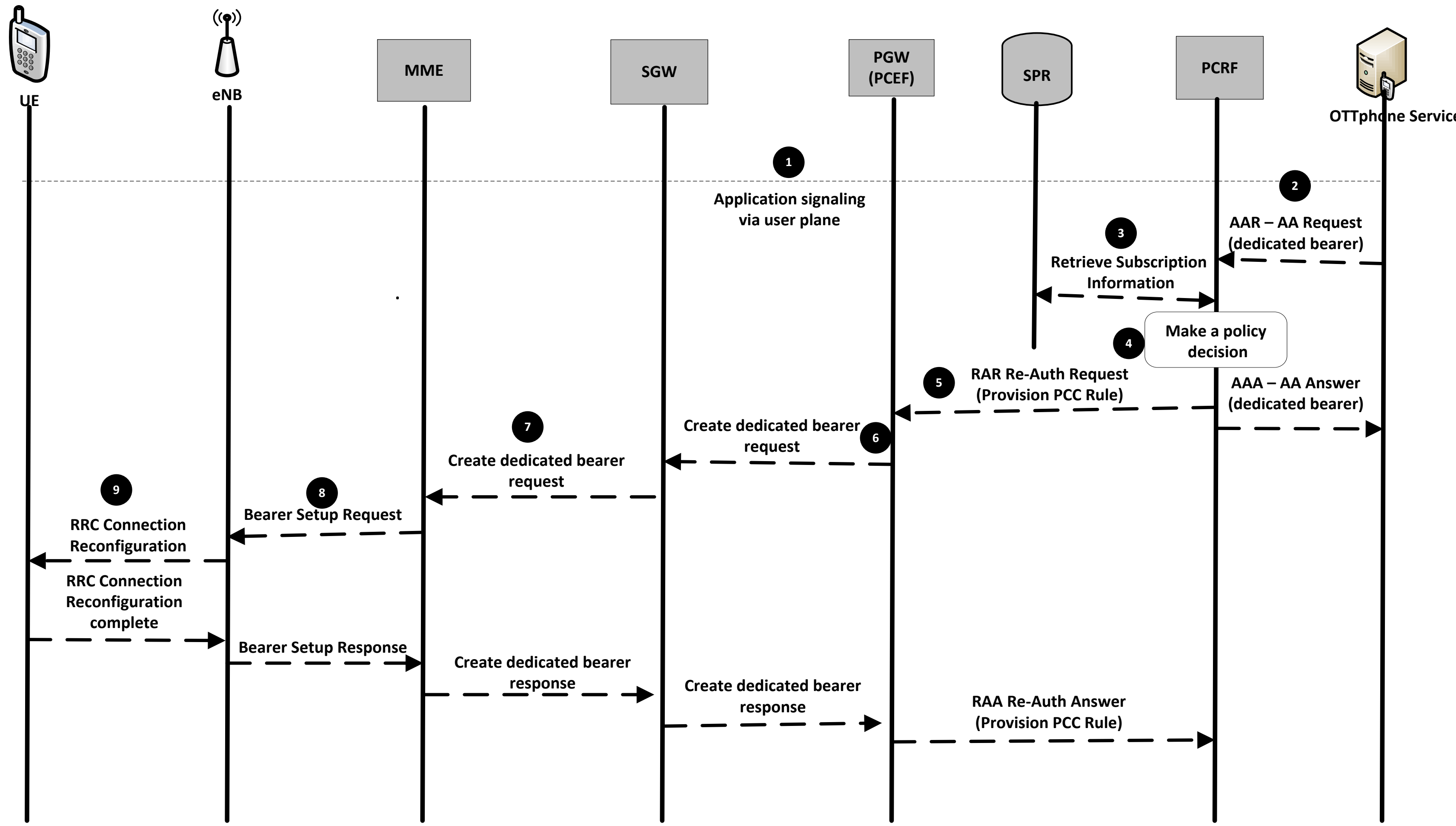
The SGW then sends a request to the MME (Figure 5, step 7) to coordinate setting up the requested bearer with the cell sites (eNodeBs). Then, the MME sends a request to the eNodeBs to establish a new dedicated radio bearer (Figure 5, step 8).

Based on resource availability (here ARP may play a major role if the network is congested) the eNodeBs complete the bearer set-up procedure for toll-quality voice QoS, and an end-to-end dedicated bearer is established between the user device and the PGW (Figure 5, step 9).

Following Step 9 the user uses the OTTphone application with toll-quality voice QoS over the dedicated bearer for the phone call. Once the call is over, the dedicated bearer is removed. The above steps are repeated for every new phone call.

The ladder diagram below shows high-level protocol messages that are exchanged between various entities involved in the dedicated bearer setup.

Figure 6: Ladder Diagram of Dedicated Bearer Setup



This process creates QoS between the user device and the PGW (the interface between the service provider and the Internet, or other connection to the edge provider) for that application. It also creates a framework, if desired, for the wireless carrier to measure the utilization of the dedicated bearer or for the wireless carrier to bill the customer. A similar process can be used for other applications, such as streaming and video conferencing, each of which would have their own rules, their own QoS, and their own dedicated bearers.

3.8 QOS IN WI-FI AND OTHER NON-LTE ENVIRONMENTS

Wi-Fi provides high speed communications over short distances, uses unlicensed spectrum, and can be deployed at low cost by an individual, a business, a communications company, or a wireless carrier. It essentially acts as a virtual extension cord for a wireline network.

Almost all smartphones and tablets that work on a wireless carrier network can also use Wi-Fi networks. Users often configure their devices to operate in Wi-Fi mode when at home, at work, or in public spaces where Wi-Fi is available. Connecting to Wi-Fi not only provides faster service, but also lets individuals use high-bandwidth applications such as video without consuming data on their wireless plans (which typically have data caps, beyond which subscribers will be charged overage fees). Wi-Fi also often enables connectivity in indoor locations where carrier wireless signal quality is poor.

Increasingly, communications companies are deploying Wi-Fi access points. These include wireless carriers, wireless companies that specialize in Wi-Fi (such as Boingo), and cable companies (which do not have licensed spectrum, yet can provide wireless services near their customers using Wi-Fi). Wireless carriers are using Wi-Fi in high-density areas such as crowded public spaces, airports, and transit stations to offload capacity from their licensed spectrum. Cable companies are currently focused on providing Wi-Fi services in public places to customers who already buy cable Internet services at home; these companies are beginning to sell “day passes” and other services to non-customers and may continue to expand their wireless offerings.

Unlike LTE, Wi-Fi was not conceived initially with end-to-end QoS over a large area, and Wi-Fi standards do not include a PCRF or other integrated large-scale policy control system. Like LTE, however, Wi-Fi has built-in protocols to provide prioritization—so Wi-Fi can prioritize devices, users and types of traffic on the network.

Wi-Fi QoS is under the control of the access point owner. For example, individual owners can use easy-to-use configuration screens to set up QoS to enhance the speed to particular devices in a home (phones, televisions), prioritize different types of latency-sensitive traffic, and configure security.[23] In addition, more sophisticated enterprise-grade and carrier-grade Wi-Fi access points used by wireless carriers and cable companies can manage QoS and security in an integrated way across a larger network. They can also extend the QoS policies of their wireline or wireless carrier networks or other policies to their Wi-Fi access points.[24]

For example, a cable operator might prioritize its own television content, a wireless carrier might prioritize its voice services, and an airline might block external video feeds and other high-bandwidth traffic that might saturate a satellite connection to the aircraft.

Because Wi-Fi is becoming an important part of the commercial wireless ecosystem, and because Wi-Fi may become the technology of choice for wireless carriers in some environments, it is important that Wi-Fi networks that are part of a commercial network, that act as extensions to the wireline or carrier wireless network, are subject to the Open Internet rules that correspond to the commercial network and not provide a loophole from Open Internet rules.

Carrying over wireless carrier rules to Wi-Fi is feasible from a technical perspective. As in the LTE case, whatever prioritization is done by the service provider for particular types of applications (voice, video) will need to be available to any like application. And, if the prioritization on a wireless carrier network is too complex to transfer the LTE rules exactly to the Wi-Fi network (for example if the Wi-Fi network cannot identify voice traffic from unaffiliated edge providers), the technical solution should be for the Wi-Fi network to not prioritize applications.

For example, if a cable company is limited to application-agnostic QoS, the connected Wi-Fi network should only prioritize on an application-agnostic basis, if at all. If a wireless carrier is limited to treating like applications in a like manner, the prioritization on the Wi-Fi network should carry over that limitation—and if it is not possible for the Wi-Fi network to match the like application prioritization on the LTE network, then the Wi-Fi network should be limited to application-agnostic prioritization, or no prioritization at all.

---

[23] *See, for example:* Linksys, "Overview of the Linksys Smart Wi-Fi Router's Media Prioritization settings," in the Linksys Knowledge Base. http://kb.linksys.com/Linksys/ukp.aspx?pid=80&vw=1&articleid=25665 (accessed Nov. 4, 2014).

[24] Cisco Systems, "Configuring QoS in a Wireless Environment," in *Cisco 3200 Series Wireless MIC Software Configuration Guide*. http://www.cisco.com/c/en/us/td/docs/routers/access/wireless/software/guide/QualityOfService.pdf (accessed Nov. 4, 2014).

4. STRATEGIES FOR TRANSPARENCY IN WIRELESS CARRIER PRACTICES

As discussed in Section 2.1, QoS in LTE is entered into and stored in the PCRF and other LTE components in the evolved packet core as a series of policy rules. The PCRF will push the QoS policies out to the various components and modify their use based on changing circumstances in the network.

In the proposed framework, the wireless carrier would establish and manage a set of rules corresponding to 1) the treatment of premium users under application-agnostic QoS, and 2) the QoS treatment of each of the QoS profiles. As discussed previously, the application-agnostic QoS rules would specify a level of congestion that triggers the prioritization and a set of agreed-upon rules, such as a different QCI or AMBR, for those users.

Returning to our example of "like" application QoS, the toll-quality voice profile will be some combination of QCI, ARP, GBR, and AMBR. There will likely be rules stating how to change these categories in the event of various circumstances, differing conditions, and network deployment within the service area of the EPC.

The wireless carrier may add, delete, or change rules to optimize the overall performance of the network or other metrics. The specifics of how the carrier does this are part of how the carrier differentiates from other carriers and, in this model, most changes can be done without necessarily having to consult with the regulator or the edge providers, although there should be a mechanism to report significant changes that may impact how an application works. Significant changes in the case of toll-quality voice may include reductions in the GBR per bearer below the rates required by the applications, or significant reductions in the number of users receiving the profile in a given service area.

Therefore, a transparency requirement in this model is different from the transparency models discussed for wireline and previous wireless models. Transparency is necessary in order to verify that the wireless carriers are fulfilling their commitments to 1) establish QoS service profiles for classes of applications that are bandwidth-intensive or latency sensitive, 2) enact those profiles in the same way for all like applications, whether from edge providers seeking prioritization or for like applications operated by the wireless carrier, and 3) put those profiles in place in this manner throughout the wireless carrier network. We therefore recommend that:

1) The Commission designate a neutral but informed party to serve as verifiers.

2) The verifiers be provided read-only access to all policy rules programmed into the PCRF for each EPC.

3) The verifiers perform regular audits on a spot basis, similar to RF and cable TV enforcement measures, to verify compliance.

Audits on an approximately monthly basis should be sufficient at the outset. The interval can be expanded or contracted based on the outcome of the audits, or in response to complaints.

There are several benefits to this approach:

1. It does not require effort by the wireless carriers to report. It uses standard tools already in place on the system as part of the LTE standard, with access simply extended to the verifiers.

2. It can be done flexibly at any time, and therefore is more likely to capture irregularities or problems than audits at a known fixed time. The verifiers could potentially create automated scripts to capture and analyze rule changes—enabling reporting to be done quickly and efficiently.

3. It does not require onsite field testing.

4. The metrics will not become obsolete—rules will be the mechanism for creating and enacting QoS as long as LTE exists.

5. By definition, transparency by review of rules demonstrates that the wireless carrier is or is not providing like QoS for like applications. If a wireless carrier intended to block an application or type of applications, it would need to create rules to do so, which would be captured in the audit. This would be a more effective means of capturing blocking than field tests—which may suspect blocking of an application or user when, in fact, the network may be fully saturated, with all like applications unable to operate.

There will be challenges in transparency. There will be many policy rules, and they may be complex, making it more likely that automated tools maybe necessary to read and analyze the rules. The verifiers must be trusted to be neutral and thorough, as well as protective of proprietary information.

5. CONCLUSION

This Report concludes that Long Term Evolution (LTE) technology is capable of managing congestion through prioritization protocols that are application-agnostic (e.g., user-directed prioritization) and which can, when faced with severe congestion, prioritize latency-sensitive traffic while avoiding discrimination among like applications, content, or services.

For conditions of moderate congestion, we propose an approach in which wireless carriers can choose to offer users one or more enhanced levels of prioritization and service with transparently reported performance metrics (a user-directed “premium” tier). This model will offer customers some improvements in performance in conditions of moderate congestion.

For conditions of severe congestion, when latency-sensitive applications are impaired or no longer work, we propose another neutral approach that also functions within the architecture of LTE and within the structure and philosophy of the Internet. In this neutral framework, an unaffiliated edge-provider application can be provided QoS; “like” applications are treated in a “like” fashion; and the LTE network provides QoS to latency-sensitive and/or high-bandwidth applications that benefit from it or require it to run consistently and effectively in congested environments.

While it is true that there are marked differences between the technical functionality and capability of fixed and wireless networks, LTE networks clearly have thethe technological capability to manage mobile networks that would allow equitable QoS treatment for like applications, even in the presence of severe congestion. LTE technologies and 3GPP LTE standards facilitate—and certainly do not preclude—non-discriminatory policy rules for mobile data traffic management.