Government of Canada Gouvernement du Canada

700TAG – Technical Advisory Note #8

Priority and Pre-emption mechanisms in LTE Broadband Communications Networks

700 MHz Mobile Broadband for Public Safety - Technology Advisory Group
Public Security Science and Technology **v.4** **September 25, 2012**

Partners: The Technology Advisory Group for 700 MHz Public Safety Spectrum (700TAG) is composed of a collaborative group of technical experts led by Centre for Security Science and includes scientific authorities from the Communications Research Center, and technical experts from Federal/Provincial/Territorial/-Municipal agencies.

Objectives

The objective of this Technical Advisory Note (TAN) is to inform the Canadian public safety community on how priority and pre-emption mechanisms of LTE technology can be used to manage the assignment of radio resources in situations where the demand for service exceeds the available capacity. The TAN discusses how, in such cases, the data communications service would be affected in a predictable and controllable manner. Examples are used to illustrate how priority and pre-emption mechanisms support the goals of the incident command when the public safety mobile broadband network is over-subscribed. This TAN is a supplement to TAN#5 titled, "Congestion Management Techniques for Public Safety Mobile Broadband Communications Networks" [1].

Chaotic degradation of service

Studies have shown that the demand for broadband data services by first responders during commonly recurring incidents will likely exceed the available capacity of LTE technology – even with 20 MHz of dedicated bandwidth[2]. In the absence of any mechanism to manage how the radio resources are allocated during periods of congestion there would be chaotic, ie unpredictable and uncontrollable degradation of service for all first responders.

In older circuit-switched cellular technology such as GSM (2G), network congestion would be manifested by the unavailability of radio channels. Users would hear a busy-tone when trying to make a voice call. The probability to seize a radio channel would depend inversely on the number of users competing to access the channel and the length of time a channel is held by users already on the network.

For newer packet-based systems, such as LTE, if congestion management mechanisms are not applied, then in times of network congestion, user traffic packets may be lost or re-transmitted, and the additional overhead may further exacerbate the congestion. Applications such as real time video would appear pixilated or as frozen images on screen. Non-real time traffic applications such as messaging would experience delays and access to web sites could potentially time-out.

The number of simultaneous users trying to access the network at any given time will contribute to network congestion. Furthermore, should a large number converge within a small geographical area, the capacity limit (assuming a single cell) may be quickly exceeded. Another factor contributing to network congestion is the nature of the applications itself. Applications, such as video streaming, which already require a high throughput, could require additional resources for timely (un-buffered) transmission as in the case of real time video. The resolution requirements also affect the amount of radio resources required to carry that information stream. As such, when bandwidth-intensive applications are accessed concurrently by a large number of users, the probability for congestion to occur increases.

Without any mechanism to manage the allocation of radio resources, the users will experience unpredictable degraded performance of the service they receive in times of network congestion. This degradation of service from the network becomes worse as the number of users trying to access the network resources grows and as they attempt to use bandwidth-intensive applications.

Priority and pre-emption mechanisms

LTE has a number of built-in mechanisms that a network operator can use to manage the allocation of network resources during periods of congestion. These mechanisms can be asserted by the network operator to act on two dimensions: a) priority assigned to user

---

[1] TAN#5 is available through Public Safety Canada and Centre for Security Science.

[2] "700MHz Spectrum Requirements for Canadian Public Safety Interoperable Mobile Broadband Data Communications", Centre for Security Science – Govt of Canada, Feb. 28, 2011.

equipment (UE), and b) applications' priority on the network. A user's ability to access different services over the radio network is established via service flows. Each UE can access multiple service flows (ie. applications) simultaneously. Priority can be assigned independently per application.

UEs' priority to access radio resources

The 3GPP defined a mechanism known as Allocation Retention Priority (ARP)[3]. ARP is used to indicate the relative priority of requests for radio resources and can assume any value from 1 to 15. Level 1 is the highest priority. Within a UE's subscriber profile, different ARP levels can be assigned to different service flows. Thus, a UE can have a high priority to access some services and a lower priority to access other services. If a UE is shared between users, the ARP priority remains unchanged. However, there may be security settings that are established at the application layer to assert additional controls for what information is accessible by whom based on their credentials.

Pre-emption is used to determine if a new request for service is to be denied or granted, based on the ARP priority. If there are insufficient radio resources available to grant the request, then pre-emption can be used to allocate radio resources to higher priority service flows. This feature can be configured within ARP in terms of pre-emption *vulnerability* and pre-emption *capability*. The pre-emption flag can be set to 'yes' or 'no' for each attribute to indicate if a service flow can pre-empt a lower priority service flow or if it can be pre-empted by a higher priority service flow. When enabled, an existing service flow could lose the resources already assigned to it in favour of a request from a higher priority service flow. On the other hand, a request from a lower priority service flow could be denied should higher priority service flows already exist on the network.

The 3GPP recommends that the assignment of ARP levels be assigned to service flows in a consistent manner between cooperating operators in order for users to have similar experiences while in their home networks or while roaming on other networks.

Applications' priority on the network

In order for certain applications to function properly, the radio network must satisfy some minimum performance requirements for latency and packet error rate. Furthermore, some applications require a minimum guaranteed bit rate (GBR) in order to be usable, whereas others can operate using 'best effort' or non-guaranteed bit rate (non-GBR). The 3GPP has defined a quality of service (QoS) parameter that is applied at the packet level for LTE systems. This parameter is known as the Quality of Service Class Identifier (QCI). QCI can assume one of 9 different classes. Each class represents a different combination of packet error rate and latency. The 3GPP also assigned a priority value to each class as shown in Table 1.

Table 1: Standardized QCI values.[4]

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

The LTE network uses QCI values as a guideline for what resources to assign to service flows. From Table 1, services flows and corresponding applications that are categorized by QCI 1 to 4 typically require a guaranteed minimum bit rate in order to be usable. Real time voice and interactive video applications could require this level of service. If the minimum GBR is not available to support a requested service flow, and the requested service flow has a high priority ARP, then lower priority service flows will be allocated fewer resources until sufficient resources are made available in order to admit the request for the higher priority session. There is a point at which LTE mechanisms will pre-empt active service flows. The pre-emption action is a function of relative priorities of service flows, whether or not service flows are designated as GBR or non-GBR, and whether aggregated maximum bit rates have been reached.

Services flows and corresponding applications with QCI 5 to 9 are typically non-real time and can tolerate variable bit rates. When radio resources are allocated to service flows with higher QCI values, these service flows may still operate albeit at reduced throughput.

[3] 3GPP TS 23.203

[4] 3GPP TS 23.203

Use Cases

The following use-cases depict situations in which priority and pre-emption can be used. They are for illustrative purposes only and do not represent actual or intended uses of priority and pre-emption. In all cases, it is assumed that the public safety broadband network is congested and that priority and pre-emption mechanisms are invoked to manage the allocation of radio resources.

Use Case #1

A policeman arrives at the scene of an accident. She determines that emergency medical services are necessary and contacts the dispatcher to send an ambulance. While the ambulance is on route the policeman uses her lapel-mounted camera to transmit a live video feed of the scene and the injured person to the trauma centre, which is also relayed to the Emergency Medical Technician (EMT) while en route to the accident. When the EMT arrives, he attempts to establish a remote patient monitoring (RPM) session to the trauma centre to transmit the patient's vital signs. The accident occurred at the cell edge of the public safety broadband network and is thus serving that area at its lowest capacity. The consequence is that both sessions cannot be supported simultaneously. The RPM session has a higher priority level than the video session of the policeman. Therefore, the policeman's video is degraded in favour of the EMT's session. The policeman's session is degraded and not pre-empted because the streaming video service flow is a non-GBR-class service. Hence, the radio resource controller can assign fewer resource blocks to it.

Use Case #2

Police SWAT and emergency medical personnel converge on the scene of an active shooter inside a building. The tactical command centre has bridged into the indoor surveillance cameras and is relaying that information to the incident response team. Some officers are also capturing video from their positions and uploading the scenes to the tactical command centre. The network experiences congestion from the large number of feeds and the incident commander decides which feeds to allow through the network from the available pool of videos. Suddenly, the UE from an officer inside the building transmits an automatic 'man-down' signal. The UE of each officer is configured so that a 'man-down' condition, including location and vital signs, is signalled at the highest priority and can pre-empt any other traffic on the network. A sufficient number of radio resources are freed up to allow the 'man-down' session to be set up. The rules which govern which sessions are pre-empted follow a service priority policy that is pre-established by the network operator.

Use Case #3

A regional service delivery entity has decided to lease unused public safety spectrum to a commercial operator. The lease agreement states that first responders will have priority access over consumers during an emergency, with special handling for 911 messages initiated by consumers.

A serious accident has occurred between 2 motorists on a highway and both vehicles are disabled with passengers trapped in the vehicles. In the vicinity of the accident, residential consumers are streaming music and video, accessing emails and web sites, a teleworker is in a video-conference session. The vehicles are equipped with on-board telemetry systems which is a service offered by the commercial carrier to its consumer clients. The on-board systems of the vehicles communicate the emergency status and location of the vehicles to the monitoring centre. This session has high priority over the active sessions in the nearby residences and if necessary would pre-empt those active sessions. Soon, by-standers arrive and begin uploading videos of the accident via NG911 calls. Those consumers that are watching movies and listening to audio tracks begin to experience degraded performance because NG911 video has a higher priority (ARP and QCI) than streaming video that is not part of the NG911 call. The emergency call taker, realising that many feeds are duplicated because the by-standers are next to each other, decides to terminate some NG911 sessions[5]. Radio resources are freed and the streaming sessions resume normally.

Emergency responders arrive on the scene and initiate sessions on the network. The movie and audio streams are once again degraded. The teleworker's session, having higher priority in QCI than the streaming movie and audio sessions, will remain active only for as long as there are sufficient resources to support it.

While the incident is in progress, a resident's house alarm detects an intruder and the alarm system attempts to initiate a session with the alarm monitoring centre, which includes uploading the real-time streaming video images captured by the security cameras . At the same time the resident, aware of the intruder, attempts to send a text message to 911.

In this example there are multiple competing requests for radio resources. Through ARP and QCI mechanisms, the LTE network will enable those sessions that its radio resources can support according to the operator-defined rules for what service flows have priority over others. It can pre-empt some service flows and degrade others in order to accommodate higher priority service flows as they are requested. The rules can be dynamically modified by the Incident

[5] It is presumed that the emergency call-taker would not hold identical sessions active on his/her console.

Command team to better suit the specific needs of the incident response.

Conclusion

The 3GPP has specified a sophisticated set of controls for a network operator to manage the allocation of resources, ie bandwidth, on an LTE network. These controls, ARP and QCI, are vital in order for the network to respond in a predictable manner to the inevitable situations when the demand for services exceeds the available capacity. ARP and QCI values are used to prioritize service flows so that when the network capacity is insufficient to meet the requests for sessions the network may apply pre-established rules to preserve the service flows that have a higher priority. The mechanisms prioritize certain packets over others and may pre-empt lower priority service flows.

References

Mode detailed information can be found in the documents referenced below.

1. Wim Brouwer, "QoS in LTE – PSCR Demo Days", Alcatel-Lucent, Dec.2010

   http://www.pscr.gov/projects/broadband/700mhz_demo_net/stakeholder_mtg_122010/day_1/5.2_qos_priority_preemption-alu.pdf

2. 3GPP TS 22.011, "Technical Specifications Group Services and Systems Aspects – Service Accessibility (Release 8)", Chapter 4 – Access Control, V8.9.0, Sept.2009

   http://www.3gpp.org/ftp/Specs/html-info/22011.htm

3. 3GPP TS 23.203, "Policy and charging control architecture (Release 9)", July 2012

   http://www.etsi.org/deliver/etsi_ts/123200_123299/123203/09.12.00_60/ts_123203v091200p.pdf

Author: Claudio Lucente, P.Eng

Scientific Authorities: Dr. Daniel Charlebois,
Portfolio Manager – Interoperable Communications

Reviewers:

Joe Fournier, Communications Research Centre
Eric Lafond, Communications Research Centre
Jacob Gurnick, Communications Research Centre

Approval for Release: Dr. Andrew Vallerand