Random access procedures and radio access network (RAN) overload control in standard and advanced long-term evolution (LTE and LTE-A) networks **Author Details:** Nuno K. Pratas, Henning Thomsen and Petar Popovski, Aalborg University, Department of Electronic Systems, {nup,ht,petarp}@es.aau.dk

Abstract: In this chapter, we describe and discuss the current LTE random access procedure and the Radio Access Network Load Control solution within LTE/LTE-A. We provide an overview of the several considered load control solutions and give a detailed description of the standardized Extended Access Class Barring solution. We then provide a brief overview of the Load Control solutions provided by the Enhanced Packet Core (EPC) Network and how they intertwine with the Extended Access Barring at the Enhanced Universal Terrestrial Radio Access Network (E-UTRAN). We also provide an outlook on the current 3GPP efforts in regards to MTC related load control issues.

Keywords: M2M, MTC, Overload Control

Acknowledgment: The work presented in this paper was supported by the Danish Council for Independent Research (Det Frie Forskningsr˚ad) within the Sapere Aude Research Leader program, Grant No. 11-105159 "Dependable Wireless Bits for Machine-to-Machine (M2M) Communications" and by the framework of the FP7 project ICT-317669 METIS. The authors would like to acknowledge the contributions of their colleagues in METIS, although the views expressed are those of the authors and do not necessarily represent the project.

Table of Contents:

Contents

Introduction 3
E-UTRAN Access Reservation Protocol 7
1) Random Access Preamble 8
2) Random Access Response 9
3) RRC Connection Request 11
4) Contention Resolution 11
Extended Access Barring Protocol 12
Alternative E-UTRAN Load Control Principles 15
Overview of Core Network (CN) Challenges and Solutions for Load Control 16
Ongoing 3GPP work on Load Control 19
Resilience to Overload through Protocol Reengineering 21
Conclusion 25
References: 26

Introduction

With the advent of Machine-to-Machine (M2M) communications, denoted in 3GPP as Machine-Type Communications (MTC), and the densification of cellular networks (i.e. cell sizes comparable to local area networks) there will be a massive increase in the number of terminal devices connected to the network. One of the main outcomes of this growth will be the overload of the signalling plane of the cellular networks, if no proper load mechanisms and optimized signalling procedures are put in place. With that goal in mind, in 3GPP there have been in recent standard releases several study items [**3GPPTR-22.868**] where many of the MTC related challenges to the existing network architecture have been studied. These challenges include: the support of lower complexity LTE devices; efficient triggering support; efficient transmission of user packets with small amounts of data; simplification and reduction of signalling flows [**3GPPTS-24.368**]; and overload control mechanisms in both the Radio Access and Core Network.

Figure 1 depicts the Evolved Packet System (EPS) architecture model for MTC, showing the entities present in the control and user plane. The control plane is where the signalling flows take place, e.g. the activation and registration of user in a network, the establishment of voice call, the sending of a SMS, etc. In the user plane, the actual user data flows take place, e.g. data payload of a voice call, data files transfer, etc. The cellular system composed by the Long Term Evolution (LTE) and the System Architecture Evolution (SAE), jointly denoted as Evolved Packet System (EPS), supports only packet-switched services, providing IP connectivity between the User Equipment (UE) and the Packet Data Network (PDN). The radio network part is denoted as Evolved Universal Terrestrial Access Network (E-UTRAN), the Core Network (CN), which encompasses all non-radio aspects, is denoted as Evolved Packet Core (EPC). Finally, the External Services includes all entities not under the standardization umbrella of 3GPP, which include application servers, either local or through other external PDNs, such as the Internet. More recently, it has been proposed a cellular centric M2M architecture [**LYJ-13**].

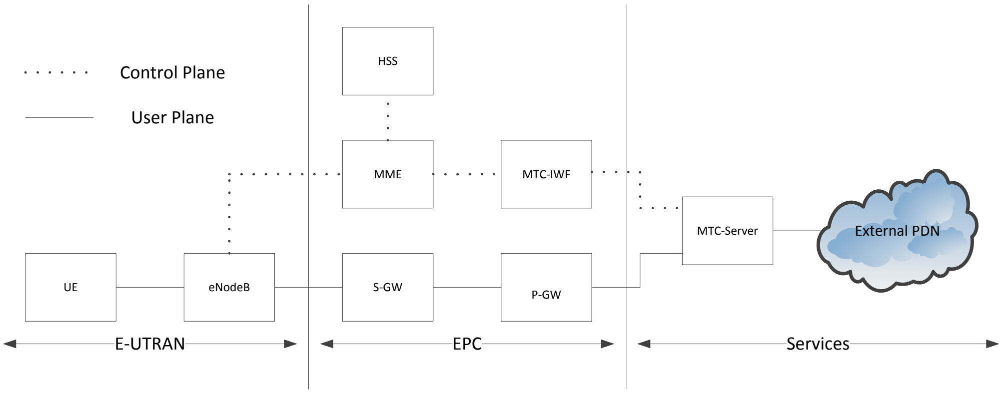


Figure 1 – EPS MTC architecture model [3GPPTR-23.888, ALUD-LTE].

The uplink overload in LTE/LTE-A, and in fact any cellular network, is caused by a very large number of UEs attempting access to the network. To understand the impact of a massive amount of users transmitting small packet to the network, consider Figure 2, which depicts the messages exchanged between the UE and the EPC network (eNB, MME and S-GW/P-GW). The first four message exchanges denote the Access Reservation Protocol (ARP) and occur within the E-UTRAN domain. The remaining ones occur within the EPC domain, although relayed through the E-UTRAN.

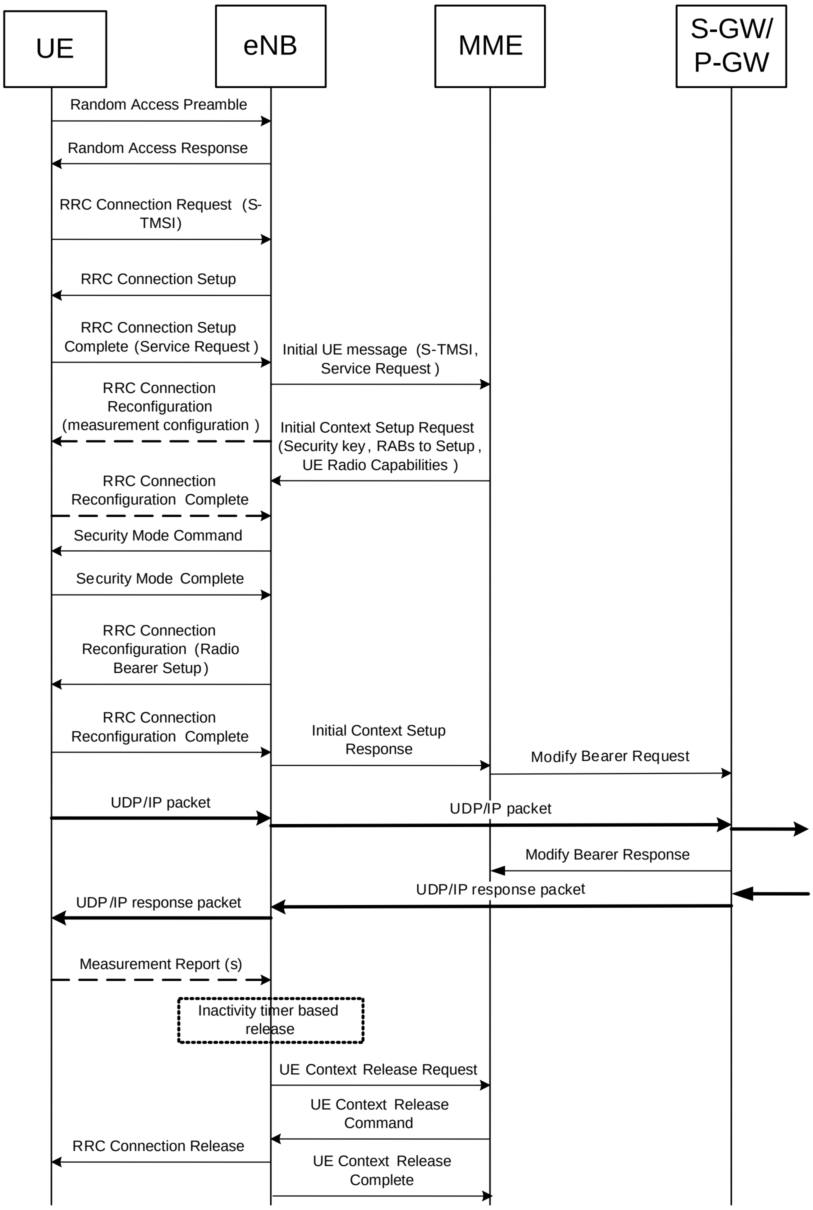


Figure 2 – 3GPP Release 11 message sequence for the transfer of a single IP packet [3GPPTR-23.887].

The overload can occur both at the E-UTRAN and EPC level, although due to different causes. At the E-UTRAN level, the overload occurs whenever the number of UEs attempting access to the network is much larger than the amount of radio resources available. At the EPC level, the overload occurs whenever the amount of UEs in a region (traffic incoming from multiple eNodeBs) attempting access is much larger than the processing capacity of the MME entity as well as the link(s) that connect it to the network. In the EPC the overload can also occur in any of the other entities, although in [**3GPPTR-23.888**] the MME was identified to be the more susceptible one.

Load control can be accomplished through proper downstream triggering of devices. Especially, if the MTC application is mostly based in collecting information from the MTC devices through triggering, then the application should distribute the triggering in such a way that the amount of users accessing the network does not exceed the amount of resources available. This can only occur if the application is aware of the network load at each of its points. Therefore, while referring to the deployment scenarios of a MTC application in 3GPP network, the indirect and hybrid model would allow such information to be available to the application. An alternative approach would be to let the network decide when each of these devices should be triggered. In this chapter we consider solely the challenges and mechanisms associated with overload control in the 3GPP context in the upstream direction, i.e. from the User Equipment (UE) to the network.

We note that although in this chapter we emphasize the load control mechanism in MTC context, the same mechanisms are applicable to other devices and traffic profiles.

In the remaining part of this chapter we describe in detail the LTE/LTE-A Access Reservation Protocol and the Extended Access Barring Scheme, followed by an overview of alternative load control principles not selected to be implemented in the E-UTRAN. Then, we give an overview of the challenges associated with overload in the EPC, the mechanism put in place to control it and how these trigger the Extended Access Barring in the E-UTRAN. Then we give a short synopsis of the past work done in 3GPP in the context of load control, as well as the future directions. Further, we introduce the concept of system reengineering, which allows cellular communication system such as LTE/LTE-A to be more resilient to overload conditions. Finally, we finalize the book chapter with a recap of the chapter contents and future outlook of the developments of MTC load control in cellular networks.

E-UTRAN Access Reservation Protocol

In this section we describe the baseline Access Reservation Protocol (ARP) employed in the E-UTRAN. We describe the random access procedure itself, and ancillary subjects that are relevant to the protocol, such as power ramping and preamble configurations.

The Access Reservation Protocol employed in the LTE/LTE-A standard can take on two forms, *contention-based* and *contention-free* [**3GPPTS-36.321**]. The contention-based method is used in the connection establishment, while the non-contention-based method is used when users have dedicated resources, in this case, orthogonal preambles. This latter case is used in e.g. handover purposes.

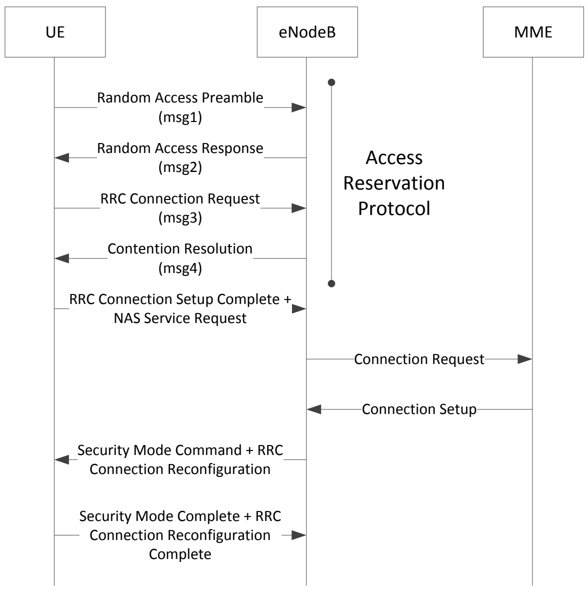


Figure 3 - Simplified Access Reservation procedure in E-UTRAN.

The contention-based Access Reservation is a four step procedure, which as shown in the message flow depicted Figure 3, where four different messages are exchanged between the UE and eNodeB [**3GPPTS-36.321**].

In [**LA-13**] the current proposals by the scientific community to allow the ARP to support efficiently Machine-Type Traffic are discussed.

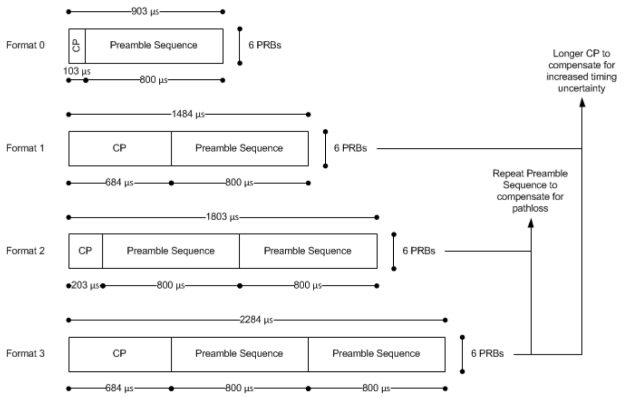
1) Random Access Preamble

The process begins with the UE selecting uniformly at random one preamble sequence from the set of preamble sequences used for contention-based access, and transmitting it at the next Physical Random Access Channel (PRACH) opportunity. There are 64 preamble sequences in total, for both contention-based and contention-free random access [**3GPPTS-36.211**]. Furthermore, the preambles used for contention-based random access are divided into two groups. In that way, by choosing a preamble from a particular group, the UE transmits a single bit of information in the RACH. The two groups are denoted as group A and B, respectively. From which group the preambles will be transmitted depends on the UEs message size.

The preambles used in the random access procedure are derived from Zadoff Chu (ZC) sequences [**3GPPTS-36.201**]. This is done by cyclically shifting a base sequence (also called a root sequence), to obtain the preambles. The autocorrelation function of a ZC sequence is periodic, which makes such sequences useful for preamble detection and timing. They also have very good cross-correlation properties.

The cyclic shifts of the preambles need to be sufficiently wide to take uplink timing uncertainty into account. In LTE there are four possible preamble format configurations possible. These are shown in Figure 4. The formats can have either short or long Cyclic Prefix (CP). The long CP, used in formats 1 and 3 enable an increased tolerance for timing uncertainty. By aggregating two preamble sequences, as done in formats 2 and 3, there can be better compensation for path loss, which also means that these two formats are better for cells of large radius.

The possible PRACH configurations are shown in Figure 5, for the case of preamble configuration 1 and frame structure type 1 [**3GPPTS-36.211**]. The PRACH configuration is selected by the eNodeB, and depends on the load, i.e. on the amount of users attempting to connect to the network [**STB-11**].



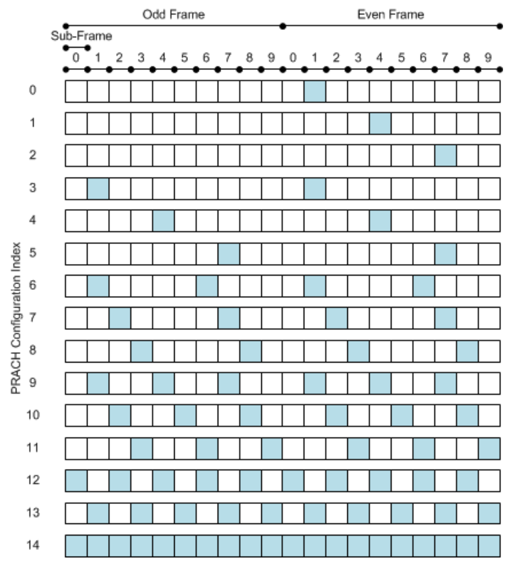


Figure 4 – PRACH preamble formats.

Figure 5 – PRACH preamble configurations.

2) Random Access Response

After the preamble transmission, the UE listens on the Physical Downlink Control Channel (PDCCH) for a Random Access Response message [**3GPPTS-36.211**]. The UE starts to listen to the PDCCH, three subframes after the subframe containing the end of the preamble, and listens for ra-ResponseWindowSize subframes, where ra-ResponseWindowSize is parameter set by configuration [**3GPPTS-36.331**]. If the eNodeB is able to detect that a preamble was transmitted within this period, it can reply with the time and frequency where the preamble was detected. This is called the Random Access Response (RAR) message. If several UEs transmitted the same preamble, they will all receive the RAR. In the case where the eNodeB is able to discern that multiple selected the same preamble then it can reply with a back-off message. This is only possible whenever the cell size is more than twice the distance corresponding to the maximum delay spread, the BS may, in some circumstances, be able to differentiate the transmission of the same preamble by two or more users, provided that the users are separable in terms of the Power Delay Profile [**Sesia-11**]. If the UE is not able to receive the RAR within the specified time period, it will increase the counter for the number of preamble transmissions by one. The UE may use power ramping, when the eNodeB does not receive the preamble transmission.

In step 1 of the random access procedure, contending UEs can use power ramping. This principle is shown in Figure 6.

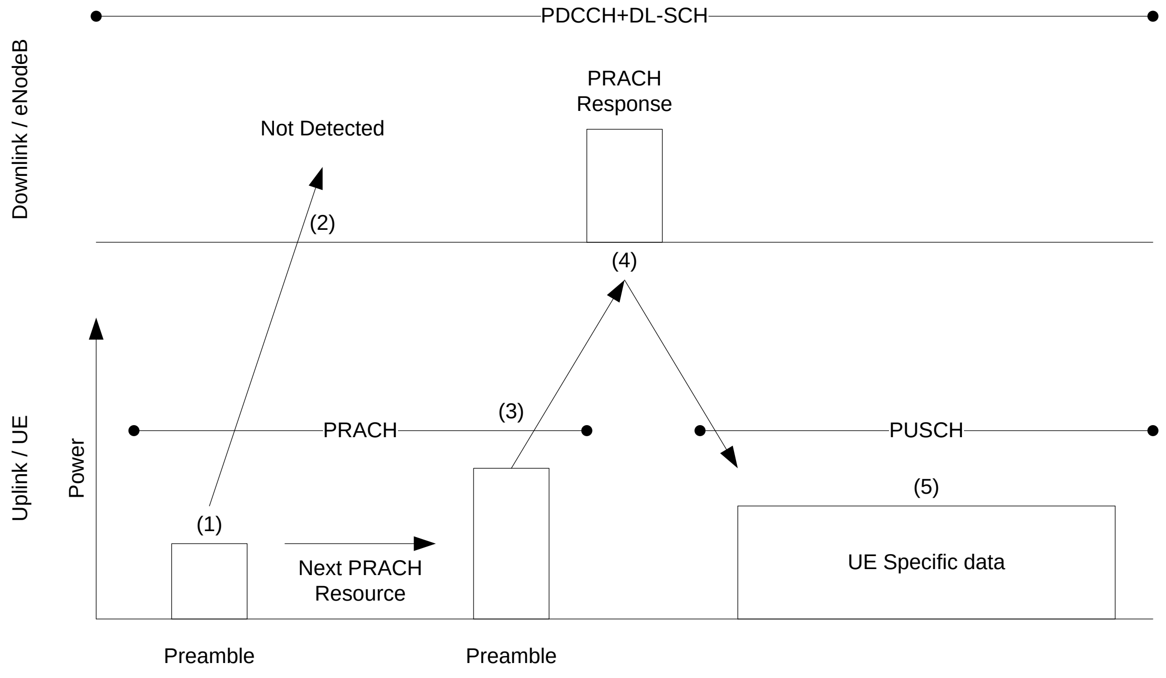


Figure 6 – Power ramping operation for random access.

The UE starts by transmitting a preamble (1), as explained in the previous section. In step (2) of the figure, the preamble is not detected at the eNodeB, due to e.g. collision, fading, or some other means. The UE then increases its preamble transmission counter by one. At the next PRACH resource, the UE transmits another preamble (3), with the transmission power increased, compared to the first transmission. The eNodeB then receives the preamble (4), and responds with a RAR. In step (5), the UE transmits its data (Layer 2 or 3 message) on the Physical Uplink Shared Channel (PUSCH).

3) RRC Connection Request

The UEs which got the RAR in step 3 each send a contention resolution message. This message includes Radio Resource Control (RRC) request and scheduling request. Hybrid Automatic Repeat request (HARQ) is used in the transmission. If several UEs sent the same preamble in step 1, they will transmit the same RRC and scheduling requests in this step and collide again.

4) Contention Resolution

Also denoted as Request Acknowledgement, here the eNodeB acknowledges any message received in the third step. This message, as the one in the previous step also uses HARQ.

More detailed information about each of the steps of the Access Reservation Procedure and the format and contents of each of the exchanged messages can be found in [**3GPPTS-36.321**] and [**3GPPTS-36.331**].

Extended Access Barring Protocol

In this section, we provide a description of the Extended Access Barring protocol, standardized in 3GPP to handle overload situations in the E-UTRAN.

Contrary to the GERAN overload control mechanism “Implicit Reject” [**3GPPTS-44.018**] that stops the UEs connection attempt after the UE has already made a request to the base station and therefore has spent radio resources, the Extended Access Barring (EAB) mechanism restricts completely the access of the UEs. Therefore, the Extended Access Class Barring (EAB) mechanism, works by stopping the radio access of UEs configured as EAB.

The Extended Access Class Barring applies only to UE (User Equipment) that are configured with low access priority and Extended Access Barring as defined in [3GPP**TS-22.011**]. This configuration is targeted primarily for usage by UEs that can tolerate being deferred when competing with other UEs for accessing network resources, as is the case during congestion situations.

The network distinguishes between UEs configured with low access priority and therefore eligible to Extended Access Barring by the indication of low priority when establishing a connection with the UTRAN. We note that this low priority flag is used by the load control mechanisms present at the CN entities, since the eNodeB when broadcasting the EAB flag does not require to know how many UEs are affected. The UEs themselves, when configured with low access priority, abstain from accessing the channel whenever the EAB flag is active.

An UE, configured for low access priority and Extended Access Barring, may be allowed to override the restrictions imposed as long as its configuration allows it. This exceptional behaviour is primarily for usage by applications that most of the time can tolerate being deferred due to low access priority when competing with other UEs for accessing network resources, but which occasionally requires access to the network when the low access priority configuration would prevent getting access. For activating this behaviour the UE requests the activation of Packet Data Network (PDN) connection without indicating that it is configured as low access priority.

The permission for overriding low access priority and Extended Access Barring restrictions by the application, still needs to be handled with care since as long as such a PDN connection without low access priority is active, the UE is not affected by any access restriction conditions that the network may set for access with low access priority. As the 3GPP system cannot determine whether any overriding of access restrictions by such UE is justified, the operator has to establish an overlaying mechanism that prevents abuse of such privileges, e.g. through specific tariffing to avoid excessive usage of overriding the low access priority, etc.

The EAB mechanism is shown in Figure 7, and it works as follows. The eNodeB starts by paging the devices. The indication of EAB is by a modification of the EAB System Information Block (SIB) [**3GPPTS-25.304**]. Only devices configured to EAB are allowed to read this block. The duration of the paging cycle is typically 2.56 seconds.

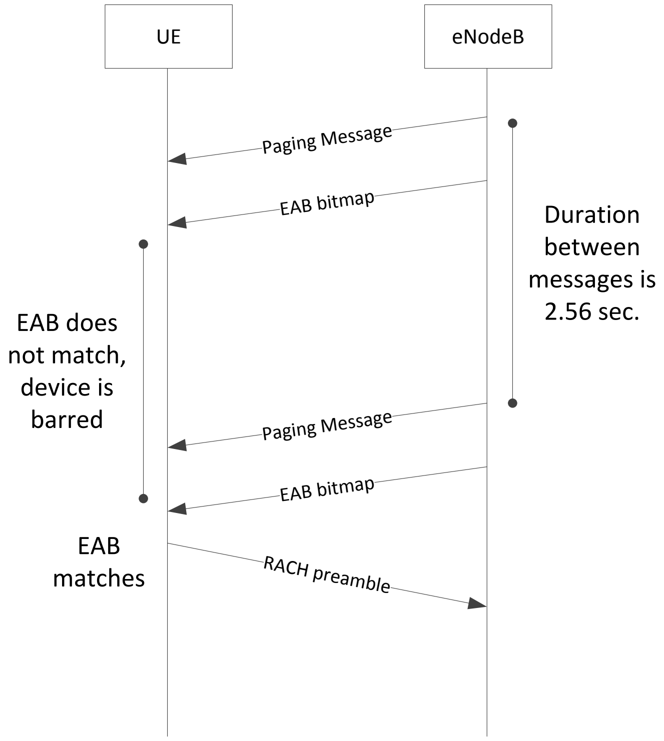


Figure 7 – Extended Access Barring mechanism used in 3GPP LTE/LTE-A networks.

After the broadcasting of the paging information, the eNodeB sends a barring bitmap, consisting of 10 bits numbered 0-9, representing different access classes. The eNodeB can also bar devices that are roaming, i.e. devices not in their home network [**JAIN-12**].

The devices do the barring check by comparing the detected bitmap to the Access Class (AC) value of the device. If the AC value and the broadcasted bitmap match, the device will not initiate any communication until the EAB SIB is changed. The network rotates access classes, by broadcasting a different EAB bitmap each time. The duration of barring varies. If the bitmap matches, the device proceeds with the communication.

Alternative E-UTRAN Load Control Principles

This section reviews the various alternative load control principles considered by 3GPP prior to the adoption of the EAB protocol [**3GPPTR-37.868, LLJK-11**].

These principles are the following:

- **Access Class Barring** - This method consists of separating users into groups, also called access classes. This method is used by the eNodeB to control the load. It does so by blocking one or several user classes. The number of classes is optional, and depends on the required granularity. The EAB protocol follows this principle.
- **Separate RACH resources for MTC** - The PRACH resources can be separated into two groups, one for H2H traffic and the other for M2M traffic.
- **Dynamic Allocation of RACH resources** - In this solution, the network can adjust the amount of PRACH resources for M2M traffic based on the load, in a dynamical manner.
- **Backoff Schemes** - In overload conditions, the network can request the M2M devices to back off. These devices will then attempt transmission at a later stage. This scheme is less suitable for massive batch arrivals than other schemes [**ShJiLi-12**].
- **Slotted Access** - Here, resources are assigned by the network for exclusive use by either individual M2M devices or groups of such devices. An M2M can be assigned an access slot based on its ID [**3GPPTR-37.868**].
- **Pull-based Schemes** - This scheme is a paging scheme, where the eNodeB triggers the M2M devices to transmit. Such triggering request is enabled by the core network. This method is only used under favourable traffic conditions.

Overview of Core Network (CN) Challenges and Solutions for Load Control

In this section we describe solely the CN mechanism in place that trigger the load control in the UTRAN and although the GERAN has also an overload control mechanism denoted as “Implicit Reject” [**3GPPTS-44.018**], this is not described here. The overload control general principle in 3GPP networks is that whenever the network starts to become overloaded, the network starts to discard/reject the access requests from low priority devices and only if still overloaded starts discarding the higher priority devices.

The introduction of devices classified as low priority was triggered by recognizing that many of the M2M application use cases considered by 3GPP are assumed to be delay tolerant [**3GPPTS-22.368**]. This led to the introduction of Low Access Priority Indication (LAPI) in the control plane signalling and at the Radio Resource Control (RRC) level.

The classification of low priority device is in general applicable to all the applications running in that same device [**JAIN-12**]. Although, there were introduced exceptions after 3GPP Release 11, where the device is able to override the low access priority configuration is case it requires emergency access. This facility is a consequence of the regulatory requirements that dictate such behaviour in normal cellular systems. This overriding facility will be available for a certain number of restricted applications in these low access priority devices.

We now describe the different overload protection mechanisms introduced to protect the 3GPP networks, although with a focus on the EPS. In Figure 8 is depicted a simplified view of the Evolved Packet Core (EPC) and the entities that are able to perform overload protection mechanisms.

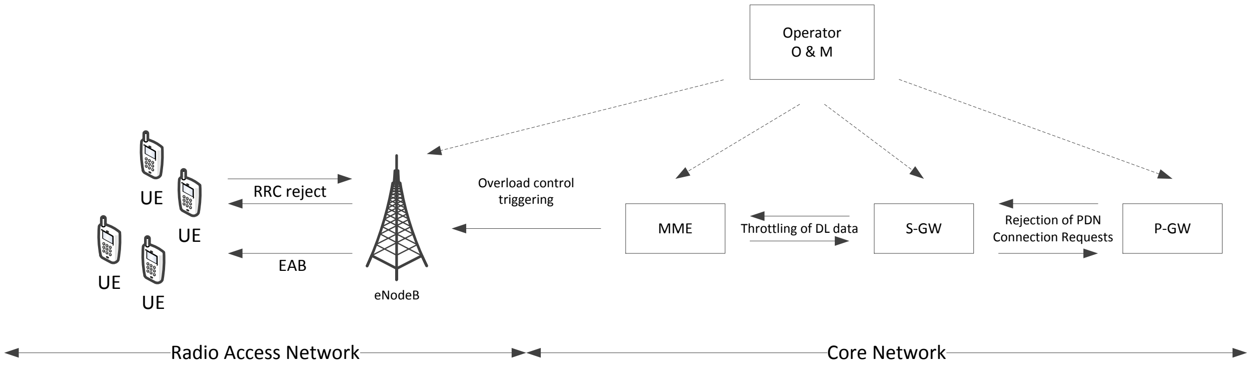


Figure 8 - Overview of overload protection mechanism in the evolved packet core [JAIN-12].

To enact the overload protection mechanism, it is first necessary to detect if there is an overload occurring at any node or nodes of the network. There are two overload detection and minimization strategies that can be put in place in a 3GPP network [**JAIN-12**].

The first is an external one, where the operator can deploy traffic analysis tools to measure the load conditions at each point and node of the network, which makes the strategy rather comprehensive. Upon detection or in the case of prior knowledge of outside world events (such as large population gatherings, etc) the operator may act directly by sending commands to the RAN or CN entities so to offload some of the traffic being generated. These operator originating commands will then externally trigger the overload control functionalities standardized by 3GPP at each of the network nodes. This strategy is traditional denoted as Operation and Management (O&M).

The second strategy is the local detection and triggering at the 3GPP network nodes, which is in line with the Self-Organizing Network (SON) paradigm [**3GPPTS-32.500**]. Upon overload detection, the PDN gateway (P-GW) may start rejecting new connections and indicate a back-off time for a certain Access Point Name (APN) to the Mobility Management Entity (MME). The MME then rejects all incoming PDN connection requests, for the specified APN during the “PDN GW back-off time.” Upon receiving such rejection including a back-off time, the device starts a back-off timer and will not try to access that APN until the back-off timer has expired.

When the MME detects an overload it can act on it by rejecting Non Access Stratum (NAS) requests and include back-off time values to offload the CN. An MME may also reject Downlink Data Notification requests for low-priority traffic for devices in idle mode. Also, to further offload the MME, the MME can request the S-GWs to selectively reduce the number of downlink data notification requests sent for downlink low-priority traffic for devices in idle mode. If the MME experiences excessive load, it can issue an overload message with a load level indication towards the RAN and whether low access priority devices are to be rejected. The RAN will then start blocking some of the traffic incoming towards the CN node.

When offloading due to an overloaded CN, the RAN can reject the UEs' RRC access attempts while informing them of a waiting timer, during which the UE will not access the network until the timer expires, unless the access attempt is for an emergency or high-priority access request. Unfortunately, even when the UE moves out from the cell where its access request was rejected, it will still wait for the timer to expire before re-attempting the access request, preventing therefore the UE to reconnect to the network. Even though this waiting time is available in devices configured for low access priority, the timer value cannot exceed 30 minutes [**3GPPTS-23.401**].

Note that the RRC rejection approach is not efficient from a radio resource utilization perspective, since at that point the UE has already wasted the radio resources associated with the access reservation request (or in the case the Packet Switch is already established the resources associated with the request). This is why the Extended Access Barring (EAB) mechanism is so important, since it stops the access reservation from occurring, i.e. the UE will not even send the Random Access Preamble message as depicted in Figure 2.

Full description of the overload control mechanism at each of the 3GPP network nodes can be found within [**3GPPTS-23.401**] and [**3GPPTS-23.060**].

Ongoing 3GPP work on Load Control

The work in 3GPP in the context of network improvements for the handling Machine-Type Communications started with the study reported in [**3GPPTR-22.868**]. From that report were highlighted a possible set of requirements to be agreed on, although no specification phase was started. In Release 10 [**3GPP-R10Description**], a specification phase was started where the goal was to provide a high level functional description of the network improvements; the introduction of device with low priority indication; and finally of Overload and Congestion control at different levels of the CN and RAN. The results of the findings where reported on [3GPP**TR-23.888**] and network improvements at the CN captured in the specifications [3GPP**TS-23.401**], [3GPP**TS-23.060**] and [3GPP**TS-23.236**]. Still, at the Release 10 no RAN mechanism was specified.

In Release 11 [**3GPP-R11Description**], the findings of the study item on improvements on the RAN for Machine-Type Communications with emphasis on load control were documented on [**3GPPTR-37.868**]. From the several identified load control mechanism candidates, the Extended Access Barring (EAB) was select for specification and the necessary RAN improvements captured on [**3GPPTS-22.368**], [**3GPPTS-36.331**] and [**3GPPTS-25.331**].

In Release 12 [**3GPP-R12Description**], the study on overload control mechanism was resumed since the overload control solutions found in the previous releases act only on UEs that are on RRC_IDLE state (i.e., in the case of overload prevents new users to become activated in the network). The observed trend is to have a high number of smart-phones “always on” in the network, which results in having a majority of these devices in RRC_CONNECTED mode, making the access control mechanism in plane ineffective, since they only apply to devices trying to switch from RRC_IDLE to RRC_CONNECTED mode. This leaves an overloaded network susceptible to become even more overloaded, since, especially in situations of e.g. disaster, people will try to send requests to the network again and again, and since their UEs are in RRC_CONNECTED mode the network is forced to process those same requests. The end result is that the network starts discarding even emergency calls and high priority calls, which has to be avoided. The current solutions being studied are going in the direction of defining application-specific access barring mechanism, where based on specific service requirements for system that shall be able to allow or prohibit the communication initiation of particular applications, as defined by the operator and subject to regional regulations that can change according with context, i.e. in case of disaster. Further, in [**3GPPTR-23.843**] and [**3GPPTR-29.809**] are studied overload control solutions at the CN focusing on tighten the control in of generated signalling streams originated from multiple Radio Access Technologies (RATs).

Resilience to Overload through Protocol Reengineering

As mentioned previously, a majority of MTC application are seen to be delay tolerant, require very low data rates and activate the radio connection very sporadically. Therefore, it is of no surprise to verify that the sum of the total data rate required by all MTC connections in a cell should be, typically, well supported by the maximal data rate available in the cell.

Consider the following illustrative example: assuming that if there is a single UE in a cell, the base station can support a connection of at most 10 Mbps in the downstream and upstream directions. While, if there are 10^4 UEs requiring 1000bps each in the uplink and downlink, the cellular system will not be able to handle the request traffic, mainly due to the lack of resource granularity due to the signalling overhead associated with the protocol of establishing and maintaining the connection of each of the devices to the network.

Although the physical layer in place would theoretical be enough to serve the incoming traffic, the overhead of the upper layers protocol put on top of it introduces a limitation to the system capacity. This leads us to the following question: is it possible to reengineer the protocol of an existing communication system while keeping its physical layer specifications essentially intact, so that the system can support a massive number of devices with low rate connections? The rationale is that by reusing the physical layer, we are reusing a major part of the infrastructure and device structure.

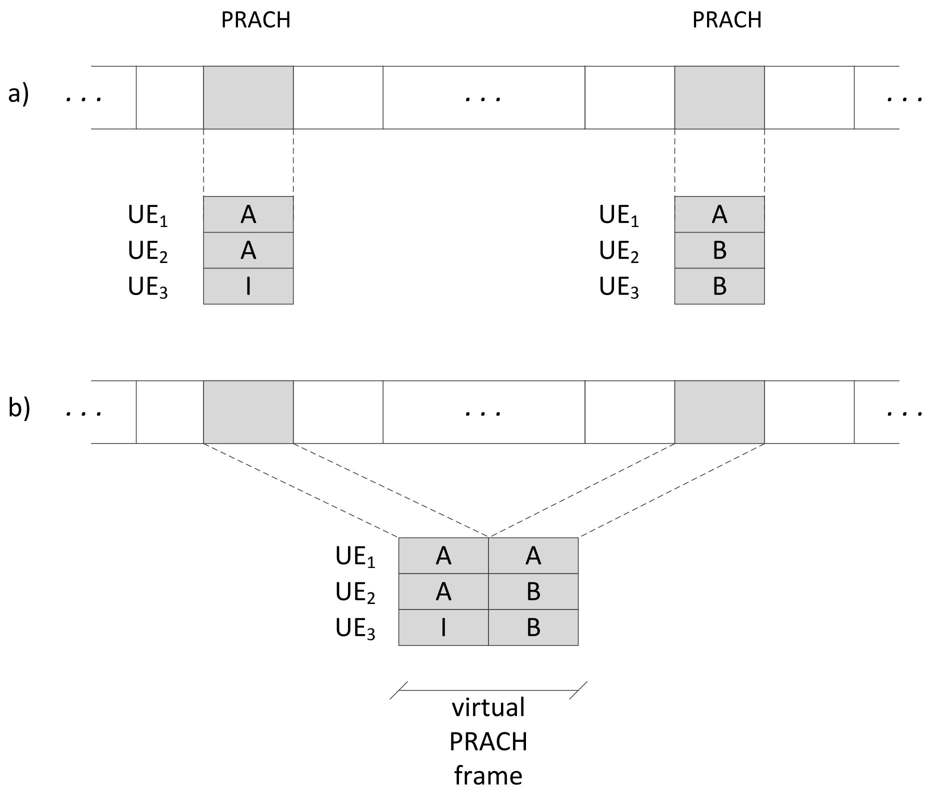


Figure 9 – Example of reenginering of the "Random Access Preamble" phase using the coded reservation principles. (a) standard LTE ARP; (b) Coded Reservation LTE ARP.

Similar efforts have been observed in the standardization work towards adapting the LTE system to support low cost MTC devices [**JAIN-12**] and other EU research projects (e.g. EXALTED); however, the changes introduced to the LTE system are limited and may only satisfy the MTC requirements in the short term.

A concrete example of the LTE protocol reengineering has been put forward in [**PTSP-12,** T**PSP-13**], denoted as Coded Reservation. Herein, the LTE's Access Reservation Protocol was reengineered so to increase the capacity of the phase associated with the transmission of the Random Access Preamble. The core difference between this scheme and the standard one is on how the UEs start the Access Reservation Procedure, while in the standard LTE the UEs do so through the transmission of a Random Access Preamble in a PRACH slot, in the Coded Reservation the UEs now do so by transmitting over several PRACH slots up to one Random Access Preamble.

To exemplify the Coded Reservation scheme, consider the illustration in Figure 9, which depicts the Random Access Preamble phase in the LTE standard Access Reservation Protocol and in the Coded Reservation. In Figure 9(a), corresponding to the LTE standard scheme, the UE1 and UE2 select the same preamble, here denoted as “A”, in the first PRACH slot, while UE3 selects to not transmit any preamble (not initiating the ARP procedure); this is identified by the idle symbol “I”. The eNodeB would then detect that the preamble “A” was activated and then proceed with the remaining phases of the ARP, resulting on UE1 and UE2 colliding in the third phase of the ARP procedure. In the subsequent PRACH slot, UE1 selects the preamble “A”, while UE2 and UE3 select the preamble “B”, leading UE2 and UE3 to collide in the ARP’s third phase. In Figure 9 (b), is shown the same contention example, but now reengineered using the Coded Reservation principle. The two PRACH slots are now combined in one virtual PRACH frame and the UEs are allowed to transmit up to one preamble in each slot of the frame. The users no longer contend in a single slot but at the frame level, with a contention codeword. The eNodeB observes the whole contention frame (i.e. the preambles transmitted at each RACH slot) and from there generates the list of possible contention codewords: AI, AA, AB, IA and IB, where the idle symbol “I” accounts for the case where the UEs remained silent in the respective slot. In this example, all three UEs no longer collide in the third phase of the ARP procedure, as the contention codewords AA, AB and IB correspond respectively to UE1, UE2 and UE3.

Through the Coded Reservation scheme, the first phase of the LTE ARP procedure is reengineered resulting in the expansion of the available contention resources following the law $(M+1)^L-1$, where M is the number of preambles and L the length of the virtual frame. The term “M+1” accounts for the idle symbol and the term “-1” accounts for exclusion of the contention codeword composed solely by idle symbols. The available contention resource scale then exponentially with L, which is a substantial improvement over the linear increase in the standard LTE with L and M achieved by selecting different PRACH configurations, as depicted in Figure 5.

In Figure 10 are shown three throughput curves in respect to the LTE Baseline and Coded Reservation, when L=4 contention slots are used. It can be seen that even when using a rather small number of preambles in the latter case, it is possible to achieve a larger contention spaces and, consequentially, increase the throughput. Therefore, this approach lends itself readily to the creation of several QoS classes by proper partitioning of the contention space in the code domain, as discussed in [**PTSP-12,** T**PSP-13**].

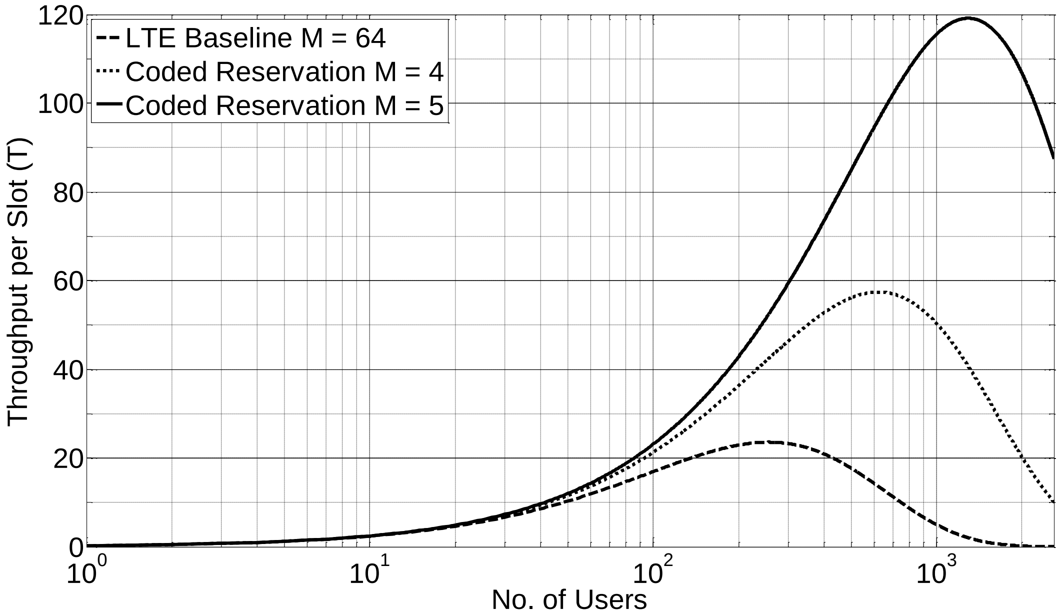


Figure 10- LTE Baseline and Coded Reservation expected asymptotic throughputs for L = 4.

The reengineering approach has also been applied in a GSM ARP context [**MSP-13**], although in the third phase of the ARP.

Existing cellular system can therefore be made resilient to overload conditions, by proper reengineering of the existing system. This is of practical interest since it allows reusing the existing cellular technologies and their deployed infrastructure.

Conclusion

In this chapter we provided a detailed description of the LTE/LTE-A Access Reservation procedure and of the load control mechanism in place in the E-UTRAN and in the EPC. Further, we have introduced how a cellular system such as the LTE/LTE-A can be reengineered into a system that is resilient to overload situations.

It is interesting to note that the future efforts, in the context of load control, by 3GPP is not in enhancing the existing load control mechanism for controlling the incoming traffic of non-active UEs (in RRC_IDLE), but instead on providing load control for users already active in the network (in RRC_CONNECTED). This comes in sequence of the LTE/LTE-A being completely packet-switched, while the previous cellular generations were predominantly circuit-switched. Therefore, with the possibility of a single UE generating several application streams, due to multiple apps running on it, it is now necessary to extend the load control to be application aware, which was not a priority in previous 3GPP releases.

This therefore enables to extend the definition of MTC beyond the one of sensor type devices attempting to connect to the network, to normal UE devices that are running in the background several APPs that requires short communication burst with the PDN, emulating the behaviour of a sensor network.

References:

[**3GPPTR-22.868**] 3GPP TR 22.868, "Study on facilitating machine to machine communication in 3GPP systems."

[**3GPPTS-24.368**] 3GPP TS 24.368,"Non-Access Stratum (NAS) configuration Management Object (MO)."

[**3GPPTR-23.888**] 3GPP TR 23.888, "System improvements for Machine-Type Communications"

[**ALUD-LTE**] "The LTE Network Architecture", Strategic White Paper

[**3GPPTR-23.887**] 3GPP TR 23.887, "Architectural Enhancements for Machine Type and other mobile data applications Communications."

[**3GPPTS-36.321**] 3GPP TS 36.321, "Medium Access Control (MAC) protocol specification"

[**3GPPTS-36.211**] 3GPP TS 36.211,"Physical Channels and Modulation"

[**3GPPTS-36.201**] 3GPP TS 36.201, "LTE physical layer; general description."

[**STB-11**] Stefania, S., Issam, T., & Matthew, B. (2009). LTE-The UMTS long term evolution from theory to practice. A John Wiley and Sons, Ltd.

[**3GPPTS-36.331**] 3GPP TS 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification"

[**Sesia-11**] S. Sesia, I. Toufik, and M. Baker, LTE—The UMTS Long Term Evolution: From Theory to Practice. Wiley, 2011.

[**3GPPTR-37.868**] 3GPP TR 37.868, “Study on RAN Improvements for Machine-type Communications”

[**ShJiLi-12**] Shafiq, M. Z., Ji, L., Liu, A. X., Pang, J., & Wang, J. (2012, June). “A first look at cellular machine-to-machine traffic: large scale measurement and characterization”. In ACM SIGMETRICS Performance Evaluation Review (Vol. 40, No. 1, pp. 65-76). doi: 10.1145/2318857.2254767

[**3GPPTS-44.018**] 3GPP TS 44.018, “Mobile radio interface layer 3 specification; Radio Resource Control (RRC) protocol”

[**3GPPTS-22.368**] 3GPP TS 22.368, “Service requirements for Machine-Type Communications (MTC)”.

[**JAIN-12**] Jain, P.; Hedman, P.; Zisimopoulos, H., "Machine type communications in 3GPP systems," Communications Magazine, IEEE , vol.50, no.11, pp.28,35, November 2012 doi: 10.1109/MCOM.2012.6353679

[**3GPPTS-32.500**] 3GPP TS 32.500, “Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements”

[**3GPPTS-23.401**] 3GPP TS 23.401, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access.”

[**3GPPTS-23.060**] 3GPP TS 23.060, “General Packet Radio Service (GPRS); Service description; Stage 2”

[**3GPP-R10Description**] 3GPP Overview of 3GPP Release 10

[**3GPP-R11Description**] 3GPP Overview of 3GPP Release 11

[**3GPP-R12Description**] 3GPP Overview of 3GPP Release 12

[**3GPPTS-23.236**] 3GPP TS 23.236 “Intra-domain connection of Radio Access Network (RAN) nodes to multiple Core Network (CN) nodes”

[**3GPPTS-25.331**] 3GPP TS 25.331 “Radio Resource Control (RRC); Protocol specification”

[**3GPPTR-23.843**] 3GPP TR 23.843 “Study on Core Network (CN) overload solutions”

[**3GPPTR-29.809**] 3GPP TR 29.809 “Study on Diameter overload control mechanisms”

[**3GPPTS-25.304**] 3GPP TS 25.304, “User Equipment (UE) procedures in idle mode and procedures for cell reselection in connected mode.”

[**PTSP-12**] N. K. Pratas, H. Thomsen, C. Stefanovic, and P. Popovski, “Code-expanded random access for machine-type communications,” in Globecom Workshops (GC Workkshops), 2012 IEEE. IEEE, 2012, pp. 1681–1686. DOI: 10.1109/GLOCOMW.2012.6477838

[T**PSP-13**] H. Thomsen, N. K. Pratas, C. Stefanovic, and P. Popovski, “Code-expanded radio access protocol for machine-to-machine communications,” in Transactions on Emerging Telecommunications Technologies Special Issue: Machine-to-Machine: An Emerging Communication Paradigm Volume 24, Issue 4, pages 355–365, June 2013. DOI: 10.1002/ett.2656

[**MSP-13**] G. C. Madueno, C. Stefanovic, and P. Popovski, “How Many Smart Meters can be Deployed in a GSM Cell?” in IEEE ICC 2013 - Second IEEE Workshop on Telecommunication Standards: From Research to Standards, Budapest, Hungary, Jun. 2013. Doi: 10.1109/ICCW.2013.6649431

[**LA-13**] Laya, A.; Alonso, L.; Alonso-Zarate, J., "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives," Communications Surveys & Tutorials, IEEE , vol.PP, no.99, pp.1,13 doi: 10.1109/SURV.2013.111313.00244

[**LLJK-11**] Lo, Anthony; Law, Yee Wei; Jacobsson, Martin; Kucharzak, Michal, ”Enhanced LTE-Advanced Random-Access Mechanism for Massive Machine-to-Machine (M2M) Communications” In the 27th World Wireless Research Forum (WWRF) Meeting, Düsseldorf, Germany, October 18-20, 2011.

[**LYJ-13**] Lo, A.; Yee Law; Jacobsson, M., "A cellular-centric service architecture for machine-to-machine (M2M) communications," Wireless Communications, IEEE , vol.20, no.5, pp.143,151, October 2013
doi: 10.1109/MWC.2013.6664485

[3GPP**TS-22.011**] 3GPP TS 22.011, “Service Accessibility.”