4

User Plane Protocols

Patrick Fischer, SeungJune Yi, SungDuck Chun and YoungDae Lee

4.1 Introduction to the User Plane Protocol Stack

The LTE Layer 2 user-plane protocol stack is composed of three sublayers, as shown in Figure 4.1:

- **The Packet Data Convergence Protocol (PDCP) layer [1]:** This layer processes Radio Resource Control (RRC) messages in the control plane and Internet Protocol (IP) packets in the user plane. Depending on the radio bearer, the main functions of the PDCP layer are header compression, security (integrity protection and ciphering), and support for reordering and retransmission during handover. For radio bearers which are configured to use the PDCP layer, there is one PDCP entity per radio bearer.
- **The Radio Link Control (RLC) layer [2]:** The main functions of the RLC layer are segmentation and reassembly of upper layer packets in order to adapt them to the size which can actually be transmitted over the radio interface. For radio bearers which need error-free transmission, the RLC layer also performs retransmission to recover from packet losses. Additionally, the RLC layer performs reordering to compensate for out-of-order reception due to Hybrid Automatic Repeat reQuest (HARQ) operation in the layer below. There is one RLC entity per radio bearer.
- **The Medium Access Control (MAC) layer [3]:** This layer performs multiplexing of data from different radio bearers. Therefore there is only one MAC entity per UE. By deciding the amount of data that can be transmitted from each radio bearer and instructing the RLC layer as to the size of packets to provide, the MAC layer aims to achieve the negotiated Quality of Service (QoS) for each radio bearer. For the

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

uplink, this process includes reporting to the eNodeB the amount of buffered data for transmission.

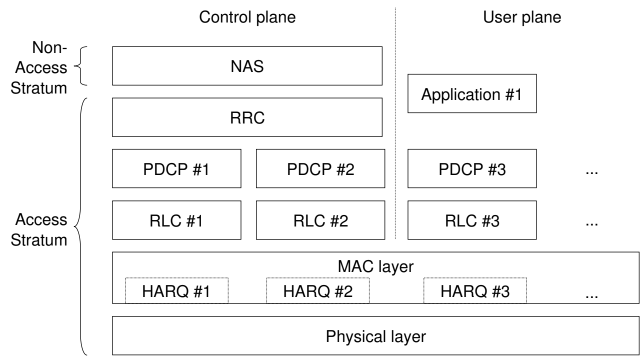

Figure 4.1: Overview of user-plane architecture.

At the transmitting side, each layer receives a Service Data Unit (SDU) from a higher layer, for which the layer provides a service, and outputs a Protocol Data Unit (PDU) to the layer below. The RLC layer receives packets from the PDCP layer. These packets are called PDCP PDUs from a PDCP point of view and represent RLC SDUs from an RLC point of view. The RLC layer creates packets which are provided to the layer below, i.e. the MAC layer. The packets provided by RLC to the MAC layer are RLC PDUs from an RLC point of view, and MAC SDUs from a MAC point of view. At the receiving side, the process is reversed, with each layer passing SDUs up to the layer above, where they are received as PDUs.

An important design feature of the LTE protocol stack is that all the PDUs and SDUs are *byte aligned*.[1] This is to facilitate handling by microprocessors, which are normally defined to handle packets in units of bytes. In order to further reduce the processing requirements of the user plane protocol stack in LTE, the headers created by each of the PDCP, RLC and MAC layers are also byte-aligned. This implies that sometimes unused padding bits are needed in the headers, and thus the cost of designing for efficient processing is that a small amount of potentially-available capacity is wasted.

[1]Byte alignment means that the lengths of the PDUs and SDUs are multiples of 8 bits.

4.2 Packet Data Convergence Protocol (PDCP)

4.2.1 Functions and Architecture

The PDCP layer performs the following functions:

- Header compression and decompression for user plane data;
- Security functions:
  - ciphering and deciphering for user plane and control plane data;
  - integrity protection and verification for control plane data;
- Handover support functions:
  - in-sequence delivery and reordering of PDUs for the layer above at handover;
  - lossless handover for user plane data mapped on RLC Acknowledged Mode (AM, see Section 4.3.1).
- Discard for user plane data due to timeout.

The PDCP layer manages data streams in the user plane, as well as in the control plane (i.e. the RRC protocol – see Section 3.2), only for the radio bearers using either a Dedicated Control CHannel (DCCH) or a Dedicated Transport CHannel (DTCH) — see Section 4.4.1.2. The architecture of the PDCP layer differs for user plane data and control plane data, as shown in Figures 4.2 and 4.3 respectively.

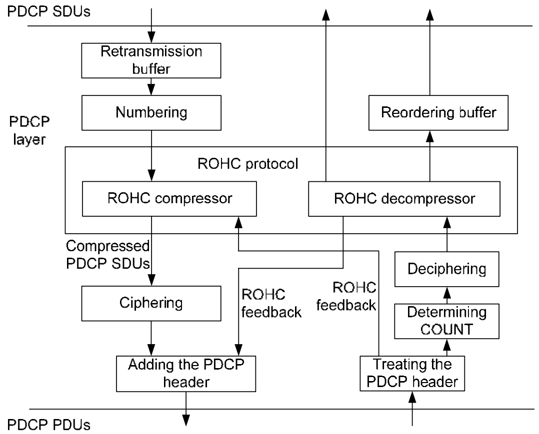

Figure 4.2: Overview of user-plane PDCP. Reproduced by permission of © 3GPP.

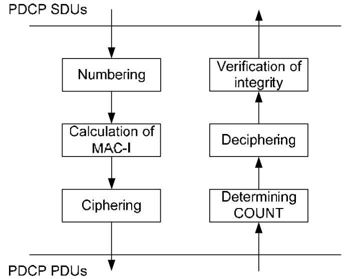

Figure 4.3: Overview of control-plane PDCP. Reproduced by permission of © 3GPP.

Each of the main functions is explained in the following subsections. Two different types of PDCP PDU are defined in LTE: PDCP Data PDUs and PDCP Control PDUs. PDCP Data PDUs are used for both control and user plane data. PDCP Control PDUs are only used to transport the feedback information for header compression, and for PDCP status reports which are used in case of handover (see Section 4.2.6) and hence are only used within the user plane.

4.2.2 Header Compression

One of the main functions of PDCP is header compression using the RObust Header Compression (ROHC) protocol defined by the IETF (Internet Engineering Task Force). In LTE, header compression is very important because there is no support for the transport of voice services via the Circuit-Switched (CS) domain.[2] Thus, in order to provide voice services on the Packet-Switched (PS) domain in a way that comes close to the efficiency normally associated with CS services it is necessary to compress the IP/UDP/RTP[3] header which is typically used for Voice over IP (VoIP) services.

The IETF specifies in 'RFC 4995'[4] a framework which supports a number of different header compression 'profiles' (i.e. sets of rules and parameters for performing the compression). The header compression profiles supported for LTE are shown in Table 4.1. This means that a UE may implement one or more of these ROHC profiles. It is important to notice that the profiles already defined in the IETF's earlier 'RFC 3095' have been redefined in RFC 4995 in order to increase robustness in some cases. The efficiency of RFC 3095 and RFC 4995 is similar, and UMTS[5] supports only RFC 3095.

[2]LTE does, however, support a CS FallBack (CSFB) procedure to allow an LTE UE to be handed over to a legacy RAT to originate a CS voice call, as well as a Single Radio Voice Call Continuity (SRVCC) procedure to hand over a Packet-Switched (PS) VoIP call to a CS voice call – see Sections 2.4.2.1 and 2.4.2.2.
[3]Internet Protocol / User Datagram Protocol / Real-Time Transport Protocol.
[4]Requests for Comments (RFCs) capture much of the output of the IETF.
[5]Universal Mobile Telecommunications System.

Table 4.1: Supported header compression protocols.

Reference	Usage
RFC 4995	No compression
RFC 3095, RFC 4815	RTP/UDP/IP
RFC 3095, RFC 4815	UDP/IP
RFC 3095, RFC 4815	ESP/IP
RFC 3843, RFC 4815	IP
RFC 4996	TCP/IP
RFC 5225	RTP/UDP/IP
RFC 5225	UDP/IP
RFC 5225	ESP/IP
RFC 5225	IP

The support of ROHC is not mandatory for the UE, except for those UEs which support VoIP. UEs which support VoIP have to support at least one profile for compression of RTP, UDP and IP.[6] The eNodeB controls by RRC signalling which of the ROHC profiles supported by the UE are allowed to be used. The ROHC compressors in the UE and the eNodeB then dynamically detect IP flows that use a certain IP header configuration and choose a suitable compression profile from the allowed and supported profiles.

ROHC header compression operates by allowing both the sender and the receiver to store the static parts of the header (e.g. the IP addresses of the sender/receiver), and to update these only when they change. Furthermore, dynamic parts (for example, the timestamp in the RTP header) are compressed by transmitting only the difference from a reference clock maintained in both the transmitter and the receiver.

As the non-changing parts of the headers are thus transmitted only once, successful decompression depends on their correct reception. Feedback is therefore used in order to confirm the correct reception of initialization information for the header decompression. Furthermore, the correct decompression of the received PDCP PDUs is confirmed periodically, depending on the experienced packet losses.

As noted above, the most important use case for ROHC is VoIP. Typically, for the transport of a VoIP packet which contains a payload of 32 bytes, the header added will be 60 bytes for the case of IPv6 and 40 bytes for the case of IPv4[7] – i.e. an overhead of 188% and 125% respectively. By means of ROHC, after the initialization of the header compression entities, this overhead can be compressed to four to six bytes, and thus to a relative overhead of 12.5–18.8%. This calculation is valid during the active periods, but during silence periods the payload size is smaller so the relative overhead is higher.

[6]ROHC is required for VoIP supported via the IP Multimedia Subsystem (IMS); in theory it could be possible to support raw IP VoIP without implementing ROHC.

[7]IPv6 is the successor to the original IPv4, for many years the dominant version of IP used on the Internet, and introduces a significantly expanded address space.

4.2.3 Security

The security architecture of LTE was introduced in Section 3.2.3.1. The implementation of security, by ciphering (of both control plane (RRC) data and user plane data) and integrity protection (for control plane (RRC) data only), is the responsibility of the PDCP layer.

A PDCP Data PDU counter (known as 'COUNT' in the LTE specifications) is used as an input to the security algorithms. The COUNT value is incremented for each PDCP Data PDU during an RRC connection; it has a length of 32 bits in order to allow an acceptable duration for an RRC connection.

During an RRC connection, the COUNT value is maintained by both the UE and the eNodeB by counting each transmitted/received PDCP Data PDU. In order to provide robustness against lost packets, each protected PDCP Data PDU includes a PDCP Sequence Number (SN) which corresponds to the least significant bits of the COUNT value.[8] Thus if one or more packets are lost, the correct COUNT value of a newly received packet can be determined using the PDCP SN. This means that the associated COUNT value is the next highest COUNT value for which the least significant bits correspond to the PDCP SN. A loss of synchronization of the COUNT value between the UE and eNodeB can then only occur if a number of packets corresponding to the maximum SN are lost consecutively. In principle, the probability of this kind of loss of synchronization occurring could be minimized by increasing the length of the SN, even to the extent of transmitting the whole COUNT value in every PDCP Data PDU. However, this would cause a high overhead, and therefore only the least significant bits are used as the SN; the actual SN length depends on the configuration and type of PDU, as explained in the description of the PDCP PDU formats in Section 4.2.6.

This use of a counter is designed to protect against a type of attack known as a *replay attack*, where the attacker tries to resend a packet that has been intercepted previously; the use of the COUNT value also provides protection against attacks which aim at deriving the used key or ciphering pattern by comparing successive patterns. Due to the use of the COUNT value, even if the same packet is transmitted twice, the ciphering pattern will be completely uncorrelated between the two transmissions, thus preventing possible security breaches.

Integrity protection is realized by adding a field known as 'Message Authentication Code for Integrity' (MAC-I)[9] to each RRC message. This code is calculated based on the Access Stratum (AS) derived keys (see Section 3.2.3.1), the message itself, the radio bearer ID, the direction (i.e. uplink or downlink) and the COUNT value.

If the integrity check fails, the message is discarded and the integrity check failure is indicated to the RRC layer so that the RRC connection re-establishment procedure can be executed (see Section 3.2.3.5).

Ciphering is realized by performing an XOR operation with the message and a ciphering stream that is generated by the ciphering algorithm based on the AS derived keys (see Section 3.2.3.1), the radio bearer ID, the direction (i.e. uplink or downlink), and the COUNT value.

Ciphering can only be applied to PDCP Data PDUs. PDCP Control PDUs (such as ROHC feedback or PDCP status reports) are neither ciphered nor integrity protected.

[8]In order to avoid excessive overhead, the most significant bits of the COUNT value, also referred to as the Hyper Frame Number (HFN), are not signalled but derived from counting overflows of the PDCP SN.

[9]Note that the MAC-I has no relation to the MAC layer.

Except for identical retransmissions, the same COUNT value is not allowed to be used more than once for a given security key. The eNodeB is responsible for avoiding reuse of the COUNT with the same combination of radio bearer ID, AS base-key and algorithm. In order to avoid such reuse, the eNodeB may for example use different radio bearer IDs for successive radio bearer establishments, trigger an intracell handover or trigger a UE state transition from connected to idle and back to connected again (see Section 3.2).

4.2.4 Handover

Handover is performed when the UE moves from the coverage of one cell to the coverage of another cell in RRC_CONNECTED state. Depending on the required QoS, either a seamless or a lossless handover is performed as appropriate for each user plane radio bearer, as explained in the following subsections.

4.2.4.1 Seamless Handover

Seamless handover is applied for user plane radio bearers mapped on RLC Unacknowledged Mode (UM, see Section 4.3.1). These types of data are typically reasonably tolerant of losses but less tolerant of delay (e.g. voice services). Seamless handover is therefore designed to minimize complexity and delay, but may result in loss of some SDUs.

At handover, for radio bearers to which seamless handover applies, the PDCP entities including the header compression contexts are reset, and the COUNT values are set to zero. As a new key is anyway generated at handover, there is no security reason to maintain the COUNT values. PDCP SDUs in the UE for which the transmission has not yet started will be transmitted after handover to the target cell. In the eNodeB, PDCP SDUs that have not yet been transmitted can be forwarded via the X2 interface[10] to the target eNodeB. PDCP SDUs for which the transmission has already started but that have not been successfully received will be lost. This minimizes the complexity because no context (i.e. configuration information) has to be transferred between the source and the target eNodeB at handover.

4.2.4.2 Lossless Handover

Based on the SN that is added to PDCP Data PDUs it is possible to ensure in-sequence delivery during handover, and even provide a fully lossless handover functionality, performing retransmission of PDCP SDUs for which reception has not yet been acknowledged prior to the handover. This lossless handover function is used mainly for delay-tolerant services such as file downloads where the loss of one PDCP SDU can result in a drastic reduction in the data rate due to the reaction of the Transmission Control Protocol (TCP).

Lossless handover is applied for user plane radio bearers that are mapped on RLC Acknowledged Mode (AM, see Section 4.3.1).

For lossless handover, the header compression protocol is reset in the UE because the header compression context is not forwarded from the source eNodeB to the target eNodeB. However, the PDCP SNs and the COUNT values associated with PDCP SDUs are maintained. For simplicity reasons, inter-eNodeB handover and intra-eNodeB handover are handled in the same way in LTE.

[10]For details of the X2 interface, see Section 2.6.

In normal transmission, while the UE is not handing over from one cell to another, the RLC layer in the UE and the eNodeB ensures in-sequence delivery. PDCP PDUs that are retransmitted by the RLC protocol, or that arrive out of sequence due to the variable delay in the HARQ transmission, are reordered based on the RLC SN. At handover, the RLC layer in the UE and in the eNodeB will deliver all PDCP PDUs that have already been received to the PDCP layer in order to have them decompressed before the header compression protocol is reset. Because some PDCP SDUs may not be available at this point, the PDCP SDUs that are not available in-sequence are not delivered immediately to higher layers in the UE or to the gateway in the network. In the PDCP layer, the PDCP SDUs received out of order are stored in the reordering buffer (see Figure 4.2). PDCP SDUs that have been transmitted but not yet been acknowledged by the RLC layer are stored in a retransmission buffer in the PDCP layer.

In order to ensure lossless handover in the *uplink*, the UE retransmits the PDCP SDUs stored in the PDCP retransmission buffer. This is illustrated in Figure 4.4. In this example the PDCP entity has initiated transmission for the PDCP SDUs with the sequence numbers 1 to 5; the packets with the sequence numbers 3 and 5 have not been received by the source eNodeB, for example due to the handover interrupting the HARQ retransmissions. After the handover, the UE restarts the transmission of the PDCP SDUs for which successful transmission has not yet been acknowledged to the target eNodeB. In the example in Figure 4.4 only the PDCP SDUs 1 and 2 have been acknowledged prior to the handover. Therefore, after the handover the UE will retransmit the packets 3, 4 and 5, although the network had already received packet 4.

In order to ensure in-sequence delivery in the uplink, the source eNodeB, after decompression, delivers the PDCP SDUs that are received in-sequence to the gateway, and forwards the PDCP SDUs that are received out-of-sequence to the target eNodeB. Thus, the target eNodeB can reorder the decompressed PDCP SDUs received from the source eNodeB and the retransmitted PDCP SDUs received from the UE based on the PDCP SNs which are maintained during the handover, and deliver them to the gateway in the correct sequence.

In order to ensure lossless handover in the *downlink*, the source eNodeB forwards the uncompressed PDCP SDUs for which reception has not yet been acknowledged by the UE to the target eNodeB for retransmission in the downlink. The source eNodeB receives an indication from the gateway that indicates the last packet sent to the source eNodeB. The source eNodeB also forwards this indication to the target eNodeB so that the target eNodeB knows when it can start transmission of packets received from the gateway. In the example in Figure 4.5, the source eNodeB has started the transmission of the PDCP SDUs 1 to 4; due to, for example, a handover occurring prior to the HARQ retransmissions of packet 3, packet 3 will not be received by the UE from the source eNodeB. Furthermore the UE has only sent an acknowledgment for packets 1 and 2, although packet 4 has been received by the UE. The target eNodeB then ensures that the PDCP SDUs that have not yet been acknowledged in the source eNodeB are sent to the UE. Thus, the UE can reorder the received PDCP SDUs and the PDCP SDUs that are stored in the reordering buffer, and deliver them to higher layers in sequential order.

The UE will expect the packets from the target eNodeB in ascending order of SNs. In the case of a packet not being forwarded from the source eNodeB to the target eNodeB, i.e. when one of the packets that the UE expects is missing during the handover operation, the UE can immediately conclude that the packet is lost and can forward the packets which have already been received in sequence to higher layers. This avoids the UE having to retain

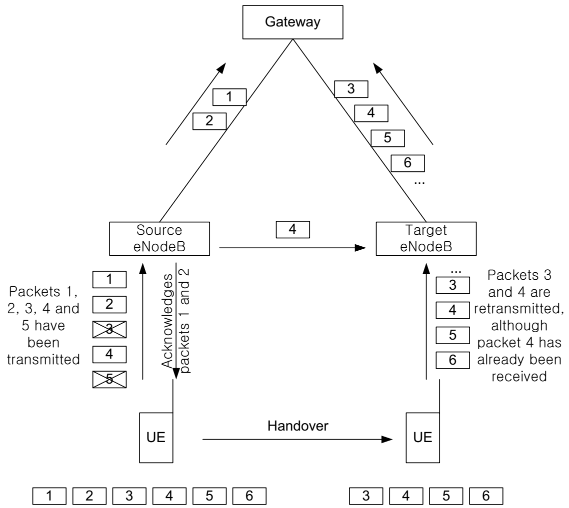

Figure 4.4: Lossless handover in the uplink.

already-received packets in order to wait for a potential retransmission. Thus the forwarding of the packets in the network can be decided without informing the UE.

In some cases it may happen that a PDCP SDU has been successfully received, but a corresponding RLC acknowledgement has not. In this case, after the handover, there may be unnecessary retransmissions initiated by the UE or the target eNodeB based on the incorrect status received by the RLC layer. In order to avoid these unnecessary retransmissions a PDCP status report can be sent from the eNodeB to the UE and from the UE to the eNodeB as described in Section 4.2.6. Additionally, a PDCP 'Status Report' can request retransmission of PDCP SDUs which were correctly received but failed in header decompression. Whether to send a PDCP status report after handover is configured independently for each radio bearer.

4.2.5 Discard of Data Packets

Typically, the data rate that is available on the radio interface is smaller than the data rate available on the network interfaces (e.g. S1[11]). Thus, when the data rate of a given service is higher than the data rate provided by the LTE radio interface, this leads to buffering in the UE and in the eNodeB. This buffering allows the scheduler in the MAC layer some freedom to vary the instantaneous data rate at the physical layer in order to adapt to the current radio

[11]For details of the S1 interface, see Section 2.5.

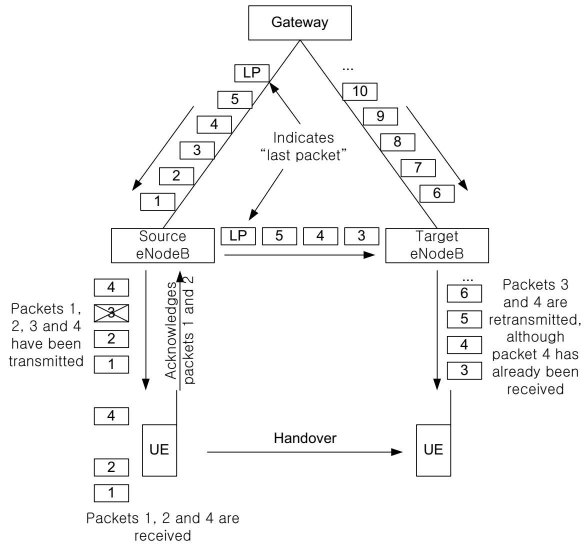

Figure 4.5: Lossless handover in the downlink.

channel conditions. Thanks to the buffering, the variations in the instantaneous data rate are then seen by the application only as some jitter in the transfer delay.

However, when the data rate provided by the application exceeds the data rate provided by the radio interface for a long period, large amounts of buffered data can result. This may lead to a large loss of data at handover if lossless handover is not applied to the bearer, or to an excessive delay for real time applications.

In the fixed internet, one of the roles typically performed by the routers is to drop packets when the data rate of an application exceeds the available data rate in a part of the internet. An application may then detect this loss of packets and adapt its data rate to the available rate. A typical example is the TCP transmit window handling, where the transmit window of TCP is reduced when a lost packet is detected, thus adapting to the available rate. Other applications such as video or voice calls via IP can also detect lost packets, for example via RTCP (Real-time Transport Control Protocol) feedback, and can adapt the data rate accordingly.

In order to allow these mechanisms to work, and to prevent excessive delay, a discard function is included in the PDCP layer for LTE. This discard function is based on a timer, where for each PDCP SDU received from the higher layers in the transmitter a timer is started, and when the transmission of the PDCP SDU has not yet been initiated in the UE at the expiry of this timer the PDCP SDU is discarded. If the timer is set to an appropriate value for the required QoS of the radio bearer, this discard mechanism can prevent excessive delay and queuing in the transmitter.

4.2.6 PDCP PDU Formats

PDCP PDUs for user plane data comprise a 'D/C' field in order to distinguish Data and Control PDUs, the formats of which are shown in Figures 4.6 and 4.7 respectively. PDCP Data PDUs comprise a 7- or 12-bit SN as shown in Table 4.2. PDCP Data PDUs for user plane data contain either an uncompressed (if header compression is not used) or a compressed IP packet.

Figure 4.6: Key features of PDCP Data PDU format. See Table 4.2 for presence of D/C and MAC-I fields.

Table 4.2: PDCP Data PDU formats.

PDU type	D/C field	Sequence number length	MAC-I	Applicable RLC Modes (see Section 4.3.1)
User plane long SN	Present	12 bits	Absent	AM / UM
User plane short SN	Present	7 bits	Absent	UM
Control plane	Absent	5 bits	32 bits	AM

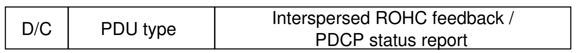

Figure 4.7: Key features of PDCP Control PDU format.

PDCP Data PDUs for control plane data (i.e. RRC signalling) comprise a MAC-I field of 32-bit length for integrity protection. PDCP Data PDUs for control plane data contain one complete RRC message.

As can be seen in Table 4.2 there are three types of PDCP Data PDU, distinguished mainly by the length of the PDCP SN and the presence of the MAC-I field. As mentioned in Section 4.2.3, the length of the PDCP SN in relation to the data rate, the packet size and the packet inter-arrival rate determines the maximum possible interruption time without desynchronizing the COUNT value which is used for ciphering and integrity protection.

The PDCP Data PDU for user plane data using the long SN allows longer interruption times, and makes it possible, when it is mapped on RLC Acknowledged Mode (AM) (see Section 4.3.1.3), to perform lossless handover as described in Section 4.2.4, but implies a higher overhead. Therefore it is mainly used for data applications with a large IP packet size where the overhead compared to the packet size is not too significant, for example for file transfer, web browsing, or e-mail traffic.

The PDCP Data PDU for user plane data using the short SN is mapped on RLC Unacknowledged Mode (UM) (see Section 4.3.1.2) and is typically used for VoIP services, where only seamless handover is used and retransmission is not necessary.

PDCP Control PDUs are used by PDCP entities handling user plane data (see Figure 4.1). There are two types of PDCP Control PDU, distinguished by the PDU Type field in the PDCP header. PDCP Control PDUs carry either PDCP 'Status Reports' for the case of lossless handover, or ROHC feedback created by the ROHC header compression protocol. PDCP Control PDUs carrying ROHC feedback are used for user plane radio bearers mapped on either RLC UM or RLC AM, while PDCP Control PDUs carrying PDCP Status Reports are used only for user plane radio bearers mapped on RLC AM.

In order to reduce complexity, a PDCP Control PDU carrying ROHC feedback carries exactly one ROHC feedback packet – there is no possibility to transmit several ROHC feedback packets in one PDCP PDU.

A PDCP Control PDU carrying a PDCP Status Report for the case of lossless handover is used to prevent the retransmission of already-correctly-received PDCP SDUs, and also to request retransmission of PDCP SDUs which were correctly received but for which header decompression failed. This PDCP Control PDU contains a bitmap indicating which PDCP SDUs need to be retransmitted and a reference SN, the First Missing SDU (FMS). In the case that all PDCP SDUs have been received in sequence this field indicates the next expected SN, and no bitmap is included.

4.3 Radio Link Control (RLC)

The RLC layer is located between the PDCP layer (the 'upper' layer) and the MAC layer (the 'lower' layer). It communicates with the PDCP layer through a Service Access Point (SAP), and with the MAC layer via logical channels. The RLC layer reformats PDCP PDUs in order to fit them into the size indicated by the MAC layer; that is, the RLC transmitter segments and/or concatenates the PDCP PDUs, and the RLC receiver reassembles the RLC PDUs to reconstruct the PDCP PDUs.

In addition, the RLC reorders the RLC PDUs if they are received out of sequence due to the HARQ operation performed in the MAC layer. This is the key difference from UMTS, where the HARQ reordering is performed in the MAC layer. The advantage of HARQ reordering in RLC is that no additional SN and reception buffer are needed for HARQ reordering. In LTE, the RLC SN and RLC reception buffer are used for both HARQ reordering and RLC-level ARQ related operations.

The functions of the RLC layer are performed by 'RLC entities'. An RLC entity is configured in one of three data transmission modes: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM). In AM, special functions are defined to support retransmission. When UM or AM is used, the choice between the two modes is made by the eNodeB during the RRC radio bearer setup procedure (see Section 3.2.3.3), based on the QoS requirements of the EPS bearer.[12] The three RLC modes are described in detail in the following sections.

4.3.1 RLC Entities

4.3.1.1 Transparent Mode (TM) RLC Entity

As the name indicates, the TM RLC entity is transparent to the PDUs that pass through it – no functions are performed and no RLC overhead is added. Since no overhead is added, an RLC SDU is directly mapped to an RLC PDU and vice versa. Therefore, the use of TM RLC is very restricted. Only RRC messages which do not need RLC configuration can utilize the TM RLC, such as broadcast System Information (SI) messages, paging messages, and RRC messages which are sent when no Signalling Radio Bearers (SRBs) other than SRB0 (see Section 3.2.1) are available. TM RLC is not used for user plane data transmission in LTE.

TM RLC provides a unidirectional data transfer service – in other words, a single TM RLC entity is configured either as a transmitting TM RLC entity or as a receiving TM RLC entity.

4.3.1.2 Unacknowledged Mode (UM) RLC Entity

UM RLC provides a unidirectional data transfer service like TM RLC. UM RLC is mainly utilized by delay-sensitive and error-tolerant real-time applications, especially VoIP, and other delay-sensitive streaming services. Point-to-multipoint services such as MBMS (Multimedia Broadcast/Multicast Service) also use UM RLC – since no feedback path is available in the case of point-to-multipoint services, AM RLC cannot be utilized by these services.

A block diagram of the UM RLC entity is shown in Figure 4.8.

The main functions of UM RLC can be summarized as follows:

- Segmentation and concatenation of RLC SDUs;
- Reordering of RLC PDUs;
- Duplicate detection of RLC PDUs;
- Reassembly of RLC SDUs.

Segmentation and concatenation. The transmitting UM RLC entity performs segmentation and/or concatenation on RLC SDUs received from upper layers, to form RLC PDUs. The size of the RLC PDU at each transmission opportunity is decided and notified by the MAC layer depending on the radio channel conditions and the available transmission resources; therefore, the size of each transmitted RLC PDU can be different.

The transmitting UM RLC entity includes RLC SDUs into an RLC PDU in the order in which they arrive at the UM RLC entity. Therefore, a single RLC PDU can contain RLC SDUs or segments of RLC SDUs according to the following pattern:

(zero or one) SDU segment + (zero or more) SDUs + (zero or one) SDU segment.

[12]Evolved Packet System – see Section 2.

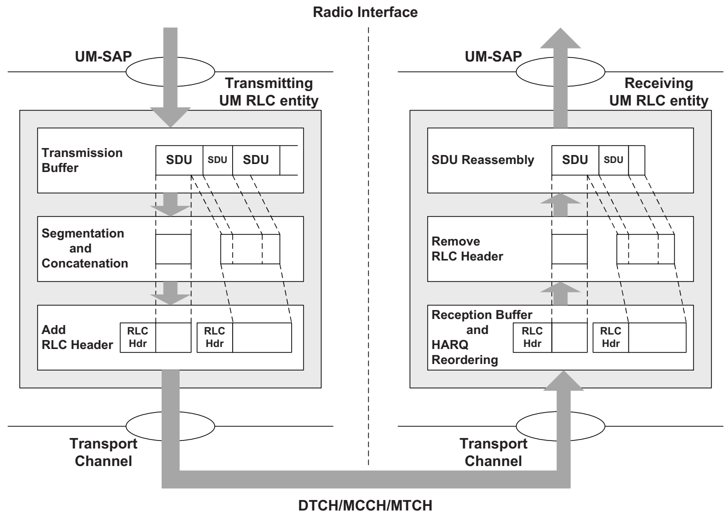

Figure 4.8: Model of UM RLC entities. Reproduced by permission of © 3GPP.

The constructed RLC PDU is always byte-aligned and has no padding.

After segmentation and/or concatenation of RLC SDUs, the transmitting UM RLC entity includes relevant UM RLC headers in the RLC PDU to indicate the sequence number[13] of the RLC PDU, and additionally the size and boundary of each included RLC SDU or RLC SDU segment.

Reordering, duplicate detection, and reassembly. When the receiving UM RLC entity receives RLC PDUs, it first reorders them if they are received out of sequence. Out-of-sequence reception is unavoidable due to the fact that the HARQ operation in the MAC layer uses multiple HARQ processes (see Section 4.4). Any RLC PDUs received out of sequence are stored in the reception buffer until all the previous RLC PDUs are received and delivered to the upper layer.

During the reordering process, any duplicate RLC PDUs received are detected by checking the SNs and discarded. This ensures that the upper layer receives upper layer PDUs only once. The most common cause of receiving duplicates is HARQ ACKs for MAC PDUs being misinterpreted as NACKs, resulting in unnecessary retransmissions of the MAC PDUs, which causes duplication in the RLC layer.

To detect reception failures and avoid excessive reordering delays, a reordering timer is used in the receiving UM RLC entity to set the maximum time to wait for the reception of RLC PDUs that have not been received in sequence. The receiving UM RLC entity starts

[13]Note that the RLC sequence number is independent from the sequence number added by PDCP.

the reordering timer when a missing RLC PDU is detected, and it waits for the missing RLC PDUs until the timer expires. When the timer expires, the receiving UM RLC entity declares the missing RLC PDUs as lost and starts to reassemble the next available RLC SDUs from the RLC PDUs stored in the reception buffer.

The reassembly function is performed on an RLC SDU basis; only RLC SDUs for which all segments are available are reassembled from the stored RLC PDUs and delivered to the upper layers. RLC SDUs that have at least one missing segment are simply discarded and not reassembled. If RLC SDUs were concatenated in an RLC PDU, the reassembly function in the RLC receiver separates them into their original RLC SDUs. The RLC receiver delivers reassembled RLC SDUs to the upper layers in increasing order of SNs.

An example scenario of a lost RLC PDU with HARQ reordering is shown in Figure 4.9. A reordering timer is started when the RLC receiver receives PDU#8. If PDU#7 has not been received before the timer expires, the RLC receiver decides that the PDU#7 is lost, and starts to reassemble RLC SDUs from the next received RLC PDU. In this example, SDU#22 and SDU#23 are discarded because they are not completely received, and SDU#24 is kept in the reception buffer until all segments are received. Only SDU#21 is completely received, so it is delivered up to the PDCP layer.

The reordering and duplicate detection functions are not applicable to RLC entities using the Multicast Control CHannel (MCCH) or Multicast Traffic CHannel (MTCH) (see Section 4.4.1.2). This is because the HARQ operation in the MAC layer is not used for these channels.

4.3.1.3 Acknowledged Mode (AM) RLC Entity

In contrast to the other RLC transmission modes, AM RLC provides a bidirectional data transfer service. Therefore, a single AM RLC entity is configured with the ability both to transmit and to receive – we refer to the corresponding parts of the AM RLC entity as the *transmitting side* and the *receiving side* respectively.

The most important feature of AM RLC is ‘retransmission’. An ARQ operation is performed to support error-free transmission. Since transmission errors are corrected by retransmissions, AM RLC is mainly utilized by error-sensitive and delay-tolerant non-real-time applications. Examples of such applications include most of the interactive/background type services, such as web browsing and file downloading. Streaming-type services also frequently use AM RLC if the delay requirement is not too stringent. In the control plane, RRC messages typically utilize the AM RLC in order to take advantage of RLC acknowledgements and retransmissions to ensure reliability.

A block diagram of the AM RLC entity is shown in Figure 4.10.

Although the AM RLC block diagram looks complicated at first glance, the transmitting and receiving sides are similar to the UM RLC transmitting and receiving entities respectively, except for the retransmission-related blocks. Therefore, most of the UM RLC behaviour described in the previous section applies to AM RLC in the same manner. The transmitting side of the AM RLC entity performs segmentation and/or concatenation of RLC SDUs received from upper layers to form RLC PDUs together with relevant AM RLC headers, and the receiving side of the AM RLC entity reassembles RLC SDUs from the received RLC PDUs after HARQ reordering.

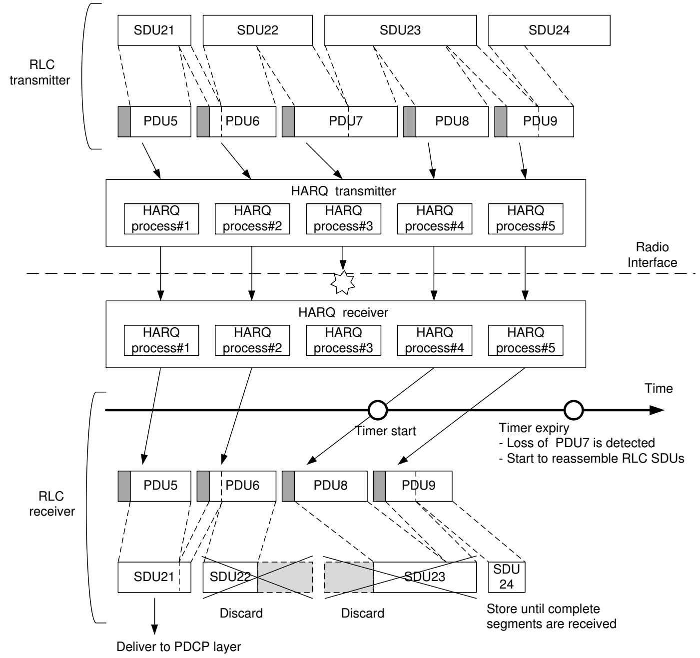

Figure 4.9: Example of PDU loss detection with HARQ reordering.

In addition to performing the functions of UM RLC, the main functions of AM RLC can be summarized as follows:

- Retransmission of RLC Data PDUs;
- Re-segmentation of retransmitted RLC Data PDUs;
- Polling;
- Status reporting;
- Status prohibit.

Retransmission and resegmentation. As mentioned before, the most important function of AM RLC is *retransmission*. In order that the transmitting side retransmits only the missing RLC PDUs, the receiving side provides a ‘Status Report’ to the transmitting side indicating ACK and/or NACK information for the RLC PDUs. Status reports are sent by

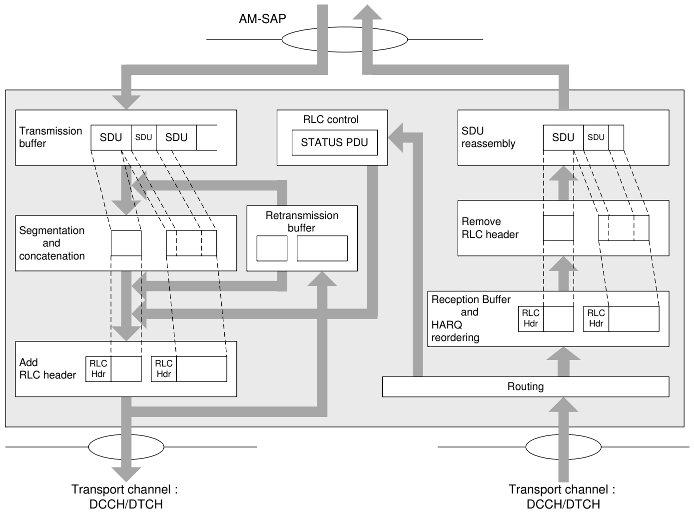

Figure 4.10: Model of AM RLC entity. Reproduced by permission of © 3GPP.

the transmitting side of the AM RLC entity whose receiving side received the corresponding RLC PDUs. Hence, the AM RLC transmitting side is able to transmit two types of RLC PDU, namely RLC Data PDUs containing data received from upper layers and RLC Control PDUs generated in the AM RLC entity itself. To differentiate between Data and Control PDUs, a 1-bit flag is included in the AM RLC header (see Section 4.3.2.3).

When the transmitting side transmits RLC Data PDUs, it stores the PDUs in the retransmission buffer for possible retransmission if requested by the receiver through a status report. In case of retransmission, the transmitter can resegment the original RLC Data PDUs into smaller PDU segments if the MAC layer indicates a size that is smaller than the original RLC Data PDU size.

An example of RLC resegmentation is shown in Figure 4.11. In this example, an original PDU of 600 bytes is resegmented into two PDU segments of 200 and 400 bytes at retransmission.

The original RLC PDU is distinguished from the retransmitted segments by another 1-bit flag in the AM RLC header: in the case of a retransmitted segment, some more fields are included in the AM RLC header to indicate resegmentation related information. The receiver can use status reports to indicate the status of individual retransmitted segments, not just full PDUs.

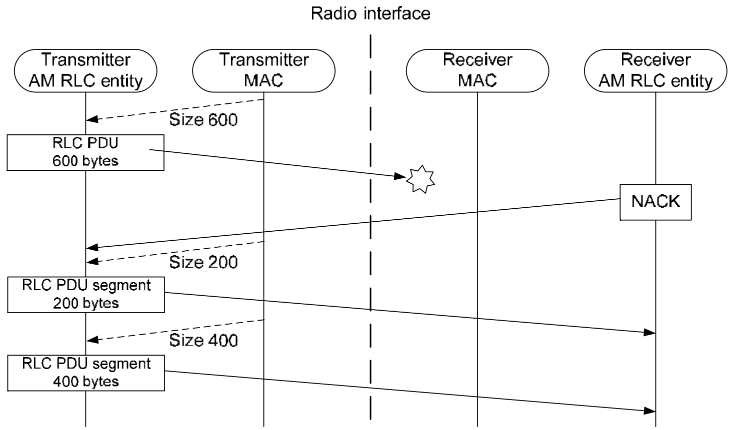

Figure 4.11: Example of RLC resegmentation.

Polling, status report and status prohibit. The transmitting side of the AM RLC entity can proactively request a status report from the peer receiving side by *polling* using a 1-bit indicator in the AM RLC header. The transmitting side can then use the status reports to select the RLC Data PDUs to be retransmitted, and manage transmission and retransmission buffers efficiently. Typical circumstances in which the transmitting side may initiate a poll include, for example, the last PDU in the transmitting side having been transmitted, or a predefined number of PDUs or data bytes having been transmitted.

When the receiving side of the AM RLC entity receives a poll from the peer transmitting side, it checks the reception buffer status and transmits a status report at the earliest transmission opportunity.

The receiving side can also generate a status report of its own accord if it detects a reception failure of an AM RLC PDU. For the detection of a reception failure, a similar mechanism is used as in the case of UM RLC in relation to the HARQ reordering delay. In AM RLC, however, the detection of a reception failure triggers a status report instead of considering the relevant RLC PDUs as permanently lost.

Note that the transmission of status reports needs to be carefully controlled according to the trade-off between transmission delay and radio efficiency. To reduce the transmission delay, status reports need to be transmitted frequently, but on the other hand frequent transmission of status reports wastes radio resources. Moreover, if further status reports are sent whilst the retransmissions triggered by a previous status report have not yet been received, unnecessary retransmissions may result, thus consuming further radio resources; in AM RLC this is in fact a second cause of duplicate PDUs occurring which have to be discarded by the duplicate-detection functionality. Therefore, to control the frequency of status reporting in an effective way, a 'status prohibit' function is available in AM RLC, whereby the transmission of new status reports is prohibited while a timer is running.

4.3.2 RLC PDU Formats

As mentioned above, the RLC layer provides two types of PDU, namely the RLC Data PDU and the RLC Control PDU. The RLC Data PDU is used to transmit PDCP PDUs and is defined in all RLC transmission modes. The RLC Control PDU delivers control information between peer RLC entities and is defined only in AM RLC. The RLC PDUs used in each RLC transmission mode are summarized in Table 4.3.

Table 4.3: PDU types used in RLC.

RLC Mode	Data PDU	Control PDU
TM	TMD (TM Data)	N/A
UM	UMD (UM Data)	N/A
AM	AMD (AM Data)/AMD segment	STATUS

In the following subsections, each of the RLC PDU formats is explained in turn.

4.3.2.1 Transparent Mode Data PDU Format

The Transparent Mode Data (TMD) PDU consists only of a data field and does not have any RLC headers. Since no segmentation or concatenation is performed, an RLC SDU is directly mapped to a TMD PDU.

4.3.2.2 Unacknowledged Mode Data PDU Format

The Unacknowledged Mode Data (UMD) PDU (Figure 4.12) consists of a data field and UMD PDU header. PDCP PDUs (i.e. RLC SDUs) can be segmented and/or concatenated into the data field. The UMD PDU header is further categorized into a fixed part (included in each UMD PDU) and an extension part (included only when the data field contains more than one SDU or SDU segment – i.e. only when the data field contains any SDU borders).

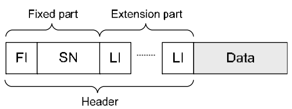

Figure 4.12: Key features of UMD PDU format.

- **Framing Info (FI).** This 2-bit field indicates whether the first and the last data field elements are complete SDUs or partial SDUs (i.e. whether the receiving RLC entity needs to receive multiple RLC PDUs in order to reassemble the corresponding SDU).
- **Sequence Number (SN).** For UMD PDUs, either a short (5 bits) or a long (10 bits) SN field can be used. This field allows the receiving RLC entity unambiguously to identify a UMD PDU, which allows reordering and duplicate-detection to take place.
- **Length Indicator (LI).** This 11-bit field indicates the length of the corresponding data field element present in the UMD PDU. There is a one-to-one correspondence between each LI and a data field element, except for the last data field element, for which the LI field is omitted because its length can be deduced from the UMD PDU size.

4.3.2.3 Acknowledged Mode Data PDU Format

In addition to the UMD PDU header fields, the Acknowledged Mode Data (AMD) PDU header (Figure 4.13) contains fields to support the RLC ARQ mechanism. The only difference in the PDU fields is that only the long SN field (10 bits) is used for AMD PDUs. The additional fields are as follows:

- **Data/Control (D/C).** This 1-bit field indicates whether the RLC PDU is an RLC Data PDU or an RLC Control PDU. It is present in all types of PDU used in AM RLC.
- **Resegmentation Flag (RF).** This 1-bit field indicates whether the RLC PDU is an AMD PDU or an AMD PDU segment.
- **Polling (P).** This 1-bit field is used to request a status report from the peer receiving side.

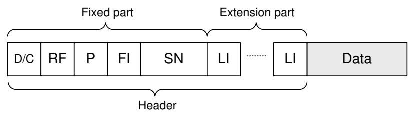

Figure 4.13: Key features of AMD PDU format.

4.3.2.4 AMD PDU Segment Format

The AMD PDU segment format (Figure 4.14) is used in case of resegmented retransmissions (when the available resource for retransmission is smaller than the original PDU size), as described in Section 4.3.1.3.

If the RF field indicates that the RLC PDU is an AMD PDU segment, the following additional resegmentation related fields are included in the fixed part of the AMD PDU header to enable correct reassembly:

- **Last Segment Flag (LSF).** This 1-bit field indicates whether or not this AMD PDU segment is the last segment of an AMD PDU.
- **Segmentation Offset (SO).** This 15-bit field indicates the starting position of the AMD PDU segment within the original AMD PDU.

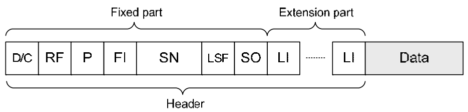

Figure 4.14: Key features of AMD PDU segment format.

4.3.2.5 STATUS PDU Format

The STATUS PDU (Figure 4.15) is designed to be very simple, as the RLC PDU error rate should normally be low in LTE due to the use of HARQ in the MAC layers. Therefore, the STATUS PDU simply lists all the missing portions of AMD PDUs by means of the following fields:

- **Control PDU Type (CPT).** This 3-bit field indicates the type of the RLC Control PDU, allowing more RLC Control PDU types to be defined in a later release of the LTE specifications. (The STATUS PDU is the only type of RLC Control PDU defined in the first version of LTE.)
- **ACK_SN.** This 10-bit field indicates the SN of the first AMD PDU which is neither received nor listed in this STATUS PDU. All AMD PDUs up to but not including this AMD PDU are correctly received by the receiver except the AMD PDUs or portions of AMD PDUs listed in the NACK_SN List.
- **NACK_SN List.** This field contains a list of SNs of the AMD PDUs that have not been completely received, optionally including indicators of which bytes of the AMD PDU are missing in the case of resegmentation.

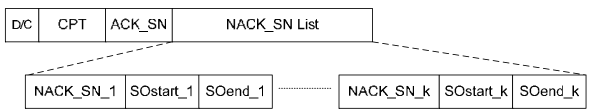

Figure 4.15: STATUS PDU format.

4.4 Medium Access Control (MAC)

The MAC layer is the lowest sublayer in the Layer 2 architecture of the LTE radio protocol stack. The connection to the physical layer below is through transport channels, and the connection to the RLC layer above is through logical channels. The MAC layer therefore performs multiplexing and demultiplexing between logical channels and transport channels: the MAC layer in the transmitting side constructs MAC PDUs, known as Transport Blocks (TBs), from MAC SDUs received through logical channels, and the MAC layer in the receiving side recovers MAC SDUs from MAC PDUs received through transport channels.

4.4.1 MAC Architecture

4.4.1.1 Overall Architecture

Figure 4.16 shows a conceptual overview of the architecture of the MAC layer.

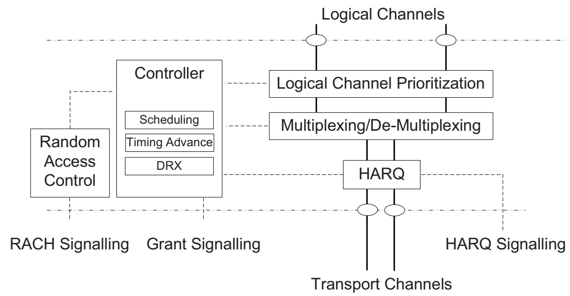

Figure 4.16: Conceptual overview of the UE-side MAC architecture.

The MAC layer consists of a HARQ entity, a multiplexing/demultiplexing entity, a logical channel prioritization entity, a random access control entity, and a controller which performs various control functions.

The HARQ entity is responsible for the transmit and receive HARQ operations. The transmit HARQ operation includes transmission and retransmission of TBs, and reception and processing of ACK/NACK signalling. The receive HARQ operation includes reception of TBs, combining of the received data and generation of ACK/NACK signalling. In order to enable continuous transmission while previous TBs are being decoded, up to eight HARQ processes in parallel are used to support multiprocess 'Stop-And-Wait' (SAW) HARQ operation.

SAW operation means that upon transmission of a TB, a transmitter stops further transmissions and awaits feedback from the receiver. When a NACK is received, or when a certain time elapses without receiving any feedback, the transmitter retransmits the TB. Such a simple SAW HARQ operation cannot on its own utilize the transmission resources during the period between the first transmission and the retransmission. Therefore multiprocess HARQ interlaces several independent SAW processes in time so that all the transmission resources can be used. Each HARQ process is responsible for a separate SAW operation and manages a separate buffer.

In general, HARQ schemes can be categorized as either *synchronous* or *asynchronous*, with the retransmissions in each case being either *adaptive* or *non-adaptive*.

In a synchronous HARQ scheme, the retransmission(s) for each process occur at predefined times relative to the initial transmission. In this way, there is no need to signal information such as HARQ process number, as this can be inferred from the transmission timing. By contrast, in an asynchronous HARQ scheme, the retransmissions can occur at any time relative to the initial transmission, so additional explicit signalling is required to indicate the HARQ process number to the receiver, so that the receiver can correctly associate each retransmission with the corresponding initial transmission. In summary, synchronous HARQ schemes reduce the signalling overhead while asynchronous HARQ schemes allow more flexibility in scheduling.

In an adaptive HARQ scheme, transmission attributes such as the Modulation and Coding Scheme (MCS) and transmission resource allocation in the frequency domain can be changed at each retransmission in response to variations in the radio channel conditions. In a non-adaptive HARQ scheme, the retransmissions are performed without explicit signalling of new transmission attributes – either by using the same transmission attributes as those of the previous transmission, or by changing the attributes according to a predefined rule. Accordingly, adaptive schemes bring more scheduling gain at the expense of increased signalling overhead.

In LTE, asynchronous adaptive HARQ is used for the downlink, and synchronous HARQ for the uplink. In the uplink, the retransmissions may be either adaptive or non-adaptive, depending on whether new signalling of the transmission attributes is provided.

The details of the HARQ incremental redundancy schemes and timing for retransmissions are explained in Section 10.3.2.5.

In the multiplexing and demultiplexing entity, data from several logical channels can be (de)multiplexed into/from one transport channel. The multiplexing entity generates MAC PDUs from MAC SDUs when radio resources are available for a new transmission, based on the decisions of the logical channel prioritization entity. The demultiplexing entity reassembles the MAC SDUs from MAC PDUs and distributes them to the appropriate RLC entities. In addition, for peer-to-peer communication between the MAC layers, control messages called 'MAC Control Elements' can be included in the MAC PDU as explained in Section 4.4.2.7 below.

The logical channel prioritization entity prioritizes the data from the logical channels to decide how much data and from which logical channel(s) should be included in each MAC PDU, as explained in Section 4.4.2.6. The decisions are delivered to the multiplexing and demultiplexing entity.

The random access control entity is responsible for controlling the Random Access CHannel (RACH) procedure (see Section 4.4.2.3). The controller entity is responsible for a number of functions including Discontinuous Reception (DRX) , the Data Scheduling procedure, and for maintaining the uplink timing alignment. These functions are explained in the following sections.

4.4.1.2 Logical Channels

The MAC layer provides a data transfer service for the RLC layer through logical channels, which are either Control Logical Channels (for the transport of control data such as RRC signalling), or Traffic Logical Channels (for user plane data).

Control logical channels.

- **Broadcast Control CHannel (BCCH).** This is a downlink channel which is used to broadcast System Information (SI) and any Public Warning System (PWS) messages (see Section 13.7). In the RLC layer, it is associated with a TM RLC entity (see Section 4.3.1).
- **Paging Control CHannel (PCCH).** This is a downlink channel which is used to notify UEs of an incoming call or a change of SI. In the RLC layer, it is associated with a TM RLC entity (see Section 4.3.1).
- **Common Control CHannel (CCCH).** This channel is used to deliver control information in both uplink and downlink directions when there is no confirmed association between a UE and the eNodeB – i.e. during connection establishment. In the RLC layer, it is associated with a TM RLC entity (see Section 4.3.1).
- **Multicast Control CHannel (MCCH).** This is a downlink channel which is used to transmit control information related to the reception of MBMS services (see Chapter 13). In the RLC layer, it is associated with a UM RLC entity (see Section 4.3.1).
- **Dedicated Control CHannel (DCCH).** This channel is used to transmit dedicated control information relating to a specific UE, in both uplink and downlink directions. It is used when a UE has an RRC connection with eNodeB. In the RLC layer, it is associated with an AM RLC entity (see Section 4.3.1).

Traffic logical channels.

- **Dedicated Traffic CHannel (DTCH).** This channel is used to transmit dedicated user data in both uplink and downlink directions. In the RLC layer, it can be associated with either a UM RLC entity or an AM RLC entity (see Section 4.3.1).
- **Multicast Traffic CHannel (MTCH).** This channel is used to transmit user data for MBMS services in the downlink (see Chapter 13). In the RLC layer, it is associated with a UM RLC entity (see Section 4.3.1).

4.4.1.3 Transport Channels

Data from the MAC layer is exchanged with the physical layer through transport channels. Data is multiplexed into transport channels depending on how it is transmitted over the air. Transport channels are classified as downlink or uplink as follows:

Downlink transport channels.

- **Broadcast CHannel (BCH).** This channel is used to transport the parts of the SI which are essential for access the Downlink Shared CHannel (DL-SCH). The transport format is fixed and the capacity is limited.
- **Downlink Shared CHannel (DL-SCH).** This channel is used to transport downlink user data or control messages. In addition, the remaining parts of the SI that are not transported via the BCH are transported on the DL-SCH.
- **Paging CHannel (PCH).** This channel is used to transport paging information to UEs, and to inform UEs about updates of the SI (see Section 3.2.2) and PWS messages (see Section 13.7).
- **Multicast CHannel (MCH).** This channel is used to transport MBMS user data or control messages that require MBSFN combining (see Chapter 13).

The mapping of the downlink transport channels onto physical channels is explained in Section 6.4.

Uplink transport channels.

- **Uplink Shared CHannel (UL-SCH).** This channel is used to transport uplink user data or control messages.
- **Random Access CHannel (RACH).** This channel is used for access to the network when the UE does not have accurate uplink timing synchronization, or when the UE does not have any allocated uplink transmission resource (see Chapter 17).

The mapping of the uplink transport channels onto physical channels is explained in Chapter 16.

4.4.1.4 Multiplexing and Mapping between Logical Channels and Transport Channels

Figures 4.17 and 4.18 show the possible multiplexing between logical channels and transport channels in the downlink and uplink respectively.

Note that in the downlink, the DL-SCH carries information from all the logical channels except the PCCH, MCCH and MTCH.

In the uplink, the UL-SCH carries information from all the logical channels.

4.4.2 MAC Functions

4.4.2.1 Scheduling

The scheduler in the eNodeB distributes the available radio resources in one cell among the UEs, and among the radio bearers of each UE. The details of the scheduling algorithm are left to the eNodeB implementation, but the signalling to support the scheduling is standardized. Some possible scheduling algorithms are discussed in Chapter 12.

In principle, the eNodeB allocates downlink or uplink radio resources to each UE based respectively on the downlink data buffered in the eNodeB and on Buffer Status Reports

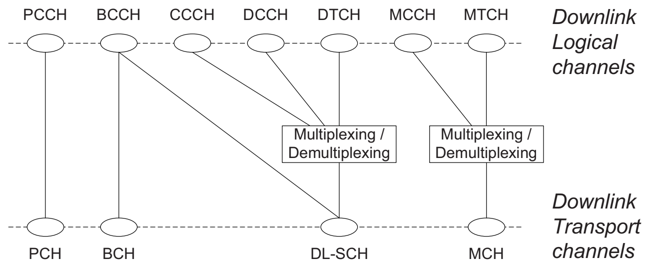

Figure 4.17: Downlink logical channel multiplexing. Reproduced by permission of © 3GPP.

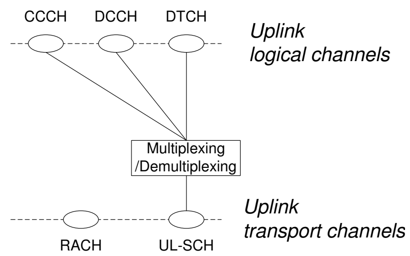

Figure 4.18: Uplink logical channel multiplexing. Reproduced by permission of © 3GPP.

(BSRs) received from the UE. In this process, the eNodeB considers the QoS requirements of each configured radio bearer, and selects the size of the MAC PDU.

The usual mode of scheduling is *dynamic scheduling*, by means of downlink assignment messages for the allocation of downlink transmission resources and uplink grant messages for the allocation of uplink transmission resources; these are valid for specific single subframes.[14] These messages also indicate whether the scheduled data is to be the first transmission of a new TB or a retransmission, by means of a 1-bit New Data Indicator (NDI); if the value of the NDI is changed relative to its previous value for the same HARQ process, the transmission is the start of a new TB. These messages are transmitted on the Physical Downlink Control CHannel (PDCCH) using a Cell Radio Network Temporary Identifier (C-RNTI) to identify the intended UE, as described in Section 9.3. This kind of scheduling is efficient for service types such as TCP or the SRBs, in which the traffic is bursty and dynamic in rate.

[14]The dynamic uplink transmission resource grants are valid for specific single subframes for initial transmissions, although they may also imply a resource allocation in later subframes for HARQ retransmissions.

In addition to the dynamic scheduling, *Semi-Persistent Scheduling* (SPS) may be used. SPS enables radio resources to be semi-statically configured and allocated to a UE for a longer time period than one subframe, avoiding the need for specific downlink assignment messages or uplink grant messages over the PDCCH for each subframe. It is useful for services such as VoIP for which the data packets are small, periodic and semi-static in size. For this kind of service the timing and amount of radio resources needed are predictable. Thus the overhead of the PDCCH is significantly reduced compared to the case of dynamic scheduling.

For the configuration of SPS, RRC signalling indicates the interval at which the radio resources are periodically assigned. Specific transmission resource allocations in the frequency domain, and transmission attributes such as the MCS, are signalled using the PDCCH. The actual transmission timing of the PDCCH messages is used as the reference timing to which the SPS interval applies. It is necessary to distinguish the PDCCH messages which apply to SPS from those used for dynamic scheduling; hence for SPS a special identity is used, known as the Semi-Persistent Scheduling C-RNTI (SPS-C-RNTI), which for each UE is different from the C-RNTI used for dynamic scheduling messages. The SPS-C-RNTI is used both for the configuration of SPS resources and for the indication of HARQ retransmissions of semi-persistently scheduled TBs. To differentiate these two cases, the NDI is used: for the configuration of the SPS resources, the SPS-C-RNTI with NDI set to 0 is used; for HARQ retransmissions, the SPS-C-RNTI with NDI set to 1 is used.

Reconfiguration of resources used for SPS can be performed for cases such as transitions between silent periods and talk spurts, or when the codec rate changes. For example, when the codec rate for a VoIP service is increased, a new downlink assignment message or uplink grant message can be transmitted to configure a larger semi-persistently scheduled radio resource for the support of bigger VoIP packets.

Allocated resources for SPS can be cancelled by an explicit scheduling message on the PDCCH using the SPS-C-RNTI indicating *SPS release*.[15] However, because there is a risk that scheduling messages can be lost in transmission, or the eNodeB's decision to release the resources may be late, an implicit mechanism to release the allocated radio resources is also specified. In the implicit mechanism, when a certain number of MAC PDUs not containing any MAC SDUs has been transmitted, the UE releases the radio resources.

Resources allocated for SPS can be temporarily overridden in a specific subframe by a scheduling message using the C-RNTI. For example, if the semi-persistently scheduled resources collide with resources configured for the Physical Random Access CHannel (PRACH) in a certain subframe, the eNodeB may choose to allocate other resources for the SPS in that subframe in order to avoid a collision with the PRACH.

Other factors potentially affecting the scheduling are uplink *Transmission Time Interval (TTI) bundling*, designed to improve uplink coverage (see Section 14.3.2), and the configuration of measurement gaps during which the UE tunes its receiver to other frequencies (see Section 22.2.1.2). In the latter case, whenever a subframe for a given HARQ process collides with a configured measurement gap, the UE can neither receive from nor transmit to a serving cell in that subframe. In such a case, if the UE cannot receive HARQ ACK/NACK feedback for an uplink TB, the UE considers that HARQ ACK is received TB and does not autonomously start a HARQ retransmission at the next transmission opportunity; to resume

[15] Explicit SPS resource release messages are positively acknowledged by the UE if they relate to downlink SPS, but not for uplink SPS.

HARQ operation, the UE has to receive a new scheduling message. If an uplink TB cannot be transmitted due to a measurement gap, the UE considers that HARQ NACK is received for that TB and transmits the TB at the next opportunity.

4.4.2.2 Scheduling Information Transfer

Buffer Status Reports (BSRs) from the UE to the eNodeB are used to assist the eNodeB's allocation of uplink radio resources. The basic assumption underlying scheduling in LTE is that radio resources are only allocated for transmission to or from a UE if data is available to be sent or received. In the downlink direction, the scheduler in the eNodeB is obviously aware of the amount of data to be delivered to each UE; however, in the uplink direction, because the scheduling decisions are performed in the eNodeB and the buffer for the data is located in the UE, BSRs have to be sent from the UE to the eNodeB to indicate the amount of data in the UE that needs to be transmitted over the UL-SCH.[16]

Two types of BSR are defined in LTE: a long BSR and a short BSR; which one is transmitted depends on the amount of available uplink transmission resources for sending the BSR, on how many groups of logical channels have non-empty buffers, and on whether a specific event is triggered at the UE. The long BSR reports the amount of data for four logical channel groups, whereas the short BSR reports the amount of data for only one logical channel group. Although the UE might actually have more than four logical channels configured, the overhead would be large if the amount of data in the UE were to be reported for every logical channel individually. Thus, grouping the logical channels into four groups for reporting purposes represents a compromise between efficiency and accuracy.

A BSR can be triggered in the following situations:

- whenever data arrives for a logical channel which has a higher priority than the logical channels whose buffers previously contained data (this is known as a Regular BSR);
- whenever data becomes available for any logical channel when there was previously no data available for transmission (a Regular BSR);
- whenever a 'retxBSR' timer expires and there is data available for transmission (a Regular BSR);
- whenever a 'periodicBSR' timer[17] expires (a Periodic BSR);
- whenever spare space in a MAC PDU can accommodate a BSR (a Padding BSR).

The 'retxBSR' timer provides a mechanism to recover from situations where a BSR is transmitted but not received. For example, if the eNodeB fails to decode a MAC PDU containing a BSR and returns a HARQ NACK, but the UE erroneously decodes the NACK as ACK, the UE will think that transmission of the BSR was successful even though it was not received by the eNodeB. In such a case, a long delay would be incurred while the UE

[16] Note that, unlike High Speed Uplink Packet Access (HSUPA), there is no possibility in LTE for a UE to transmit autonomously in the uplink by means of a transmission grant for non-scheduled transmissions. This is because the uplink transmissions from different UEs in LTE are orthogonal in time and frequency, and therefore if an uplink resource is allocated but unused, it cannot be accessed by another UE; by contrast, in HSUPA, if a UE does not use its transmission grant for non-scheduled transmissions, the resulting reduction in uplink interference can benefit other UEs. Furthermore, the short subframe length in LTE enables uplink transmission resources to be dynamically allocated more quickly than in HSUPA.

[17] Periodic BSR timer is used by RRC to control BSR reporting.

waited for an uplink resource grant that would not be forthcoming. To avoid this, the retxBSR timer is restarted whenever a uplink grant message is received; if no uplink grant is received before the timer expires, the UE transmits another BSR.

If a UE does not have enough allocated UL-SCH resources to send a BSR when a trigger for a Regular BSR occurs, the UE sends a Scheduling Request (SR) on the Physical Uplink Control CHannel (PUCCH – see Section 16.3.7) if possible; otherwise, the random access procedure (see Section 4.4.2.3) is used to request an allocation of uplink resources for sending a BSR. However, if a periodic or padding BSR is triggered when the UE does not have UL-SCH resources for a new transmission, the SR is not triggered.

Thus LTE provides suitable signalling to ensure that the eNodeB has sufficient information about the data waiting in each UE's uplink transmission buffer to allocate corresponding uplink transmission resources in a timely manner.

4.4.2.3 Random Access Procedure

The random access procedure is used when a UE is not allocated with uplink radio resources but has data to transmit, or when the UE is not time-synchronized in the uplink direction. Control of the random access procedure is an important part of the MAC layer functionality in LTE. The details are explained in Chapter 17.

4.4.2.4 Uplink Timing Alignment

Uplink timing alignment maintenance is controlled by the MAC layer and is important for ensuring that a UE's uplink transmissions arrive in the eNodeB without overlapping with the transmissions from other UEs. The details of the uplink timing advance mechanism used to maintain timing alignment are explained in Section 18.2.

The timing advance mechanism utilizes MAC Control Elements (see Section 4.4.2.7) to update the uplink transmission timing. However, maintaining the uplink synchronization in this way during periods when no data is transferred wastes radio resources and adversely impacts the UE battery life. Therefore, when a UE is inactive for a certain period of time the UE is allowed to lose uplink synchronization even in RRC_CONNECTED state. The random access procedure is then used to regain uplink synchronization when the data transfer resumes in either uplink or downlink.

4.4.2.5 Discontinuous Reception (DRX)

DRX functionality can be configured for an 'RRC_CONNECTED' UE[18] so that it does not always need to monitor the downlink channels. A DRX cycle consists of an 'On Duration' during which the UE should monitor the PDCCH and a 'DRX period' during which a UE can skip reception of downlink channels for battery saving purposes.

The parameterization of the DRX cycle involves a trade-off between battery saving and latency. On the one hand, a long DRX period is beneficial for lengthening the UE's battery life. For example, in the case of a web browsing service, it is usually a waste of resources for a UE continuously to receive downlink channels while the user is reading a downloaded

[18]Different DRX functionality applies to UEs which are in 'RRC_IDLE'. These RRC states are discussed in Chapter 3.

web page. On the other hand, a shorter DRX period is better for faster response when data transfer is resumed – for example when a user requests another web page.

To meet these conflicting requirements, two DRX cycles – a short cycle and a long cycle – can be configured for each UE, with the aim of providing a similar degree of power saving for the UE in RRC_CONNECTED as in RRC_IDLE. The transition between the short DRX cycle, the long DRX cycle and continuous reception is controlled either by a timer or by explicit commands from the eNodeB. In some sense, the short DRX cycle can be considered as a confirmation period in case a late packet arrives, before the UE enters the long DRX cycle – if data arrives at the eNodeB while the UE is in the short DRX cycle, the data is scheduled for transmission at the next wake-up time and the UE then resumes continuous reception. On the other hand, if no data arrives at the eNodeB during the short DRX cycle, the UE enters the long DRX cycle, assuming that the packet activity is finished for the time being.

Figure 4.19 shows an example of DRX operation. The UE checks for scheduling messages (indicated by its C-RNTI on the PDCCH) during the 'On Duration' period of either the long DRX cycle or the short DRX cycle depending on the currently active cycle. When a scheduling message is received during an 'On Duration', the UE starts a 'DRX Inactivity Timer' and monitors the PDCCH in every subframe while the DRX Inactivity Timer is running. During this period, the UE can be regarded as being in a continuous reception mode. Whenever a scheduling message is received while the DRX Inactivity Timer is running, the UE restarts the DRX Inactivity Timer, and when it expires the UE moves into a short DRX cycle and starts a 'DRX Short Cycle Timer'. The short DRX cycle may also be initiated by means of a MAC Control Element (see Section 4.4.2.7). When the 'DRX Short Cycle Timer' expires, the UE moves into a long DRX cycle.

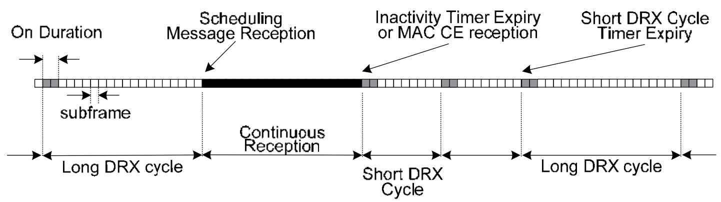

Figure 4.19: The two-level DRX procedure.

In addition to this DRX behaviour, a 'HARQ RTT (Round Trip Time) timer' is defined per downlink HARQ process with the aim of allowing the UE to sleep during the HARQ RTT. When decoding of a downlink TB for one HARQ process fails, the UE can assume that the next retransmission of the TB will occur after at least 'HARQ RTT' subframes. While the HARQ RTT timer is running, the UE does not need to monitor the PDCCH (provided that there is no other reason to be monitoring it). At the expiry of the HARQ RTT timer, if received data for a downlink HARQ process is not correctly decoded, the UE starts a 'DRX Retransmission Timer' for that HARQ process. While the timer is running, the UE monitors the PDCCH for HARQ retransmissions. The length of the DRX Retransmission Timer is related to the degree of flexibility desired for the eNodeB's scheduler. For optimal UE battery consumption, it is desirable that eNodeB schedules a HARQ retransmission as soon as the HARQ RTT timer expires. However, this requires that eNodeB always reserve some radio resources for this, and therefore the DRX Retransmission timer can be used to relax this scheduling limitation while limiting the amount of time for which the UE has to monitor the PDCCH. The HARQ RTT is illustrated in Section 10.3.2.5.

4.4.2.6 Multiplexing and Logical Channel Prioritization

Unlike the downlink, where the multiplexing and logical channel prioritization is left to the eNodeB implementation, for the uplink the process by which a UE creates a MAC PDU to transmit using the allocated radio resources is fully standardized; this is designed to ensure that the UE satisfies the QoS of each configured radio bearer in a way which is optimal and consistent between different UE implementations. Based on the uplink transmission resource grant message signalled on the PDCCH, the UE has to decide on the amount of data for each logical channel to be included in the new MAC PDU, and, if necessary, also to allocate space for a MAC Control Element.

One simple way to meet this purpose is to serve radio bearers in order of their priority. Following this principle, the data from the logical channel of the highest priority is the first to be included into the MAC PDU, followed by data from the logical channel of the next highest priority, continuing until the MAC PDU size allocated by the eNodeB is completely filled or there is no more data to transmit.

Although this kind of priority-based multiplexing is simple and favours the highest priorities, it sometimes leads to starvation of low-priority bearers. Starvation occurs when the logical channels of the lower priority cannot transmit any data because the data from higher priority logical channels always takes up all the allocated radio resources.

To avoid starvation, while still serving the logical channels according to their priorities, in LTE a Prioritized Bit Rate (PBR) is configured by the eNodeB for each logical channel. The PBR is the data rate provided to one logical channel before allocating any resource to a lower-priority logical channel.

In order to take into account both the PBR and the priority, each logical channel is served in decreasing order of priority, but the amount of data from each logical channel included into the MAC PDU is initially limited to the amount corresponding to the configured PBR. Only when all logical channels have been served up to their PBR, then if there is still room left in the MAC PDU, each logical channel is served again in decreasing order of priority. In this second round, each logical channel is served only if all logical channels of higher priority have no more data for transmission.

In most cases, a MAC Control Element has higher priority than any other logical channel because it controls the operation of a MAC entity. Thus, when a MAC PDU is composed and there is a MAC Control Element to send, the MAC Control Element is generally included first and the remaining space is used to include data from logical channels. However, since the padding BSR is used to fill up remaining space in a MAC PDU, it is included into a MAC PDU after other logical channels. Among the various types of MAC Control Element (see Section 4.4.2.7) and logical channel, the 'C-RNTI' MAC Control Element and CCCH (CCCH) have the highest priority because they are used for either contention resolution or RRC connection management. For example, the 'RRCConnectionReestablishmentRequest' message (see Section 3.2.3.5) on the uplink CCCH is used to recover a lost RRC connection, and it is more important to complete the connection reestablishment procedure as soon as possible than to inform the eNodeB of the UE's buffer status; otherwise, the data transfer interruption time would be longer and the probability of call failure would increase due to the delayed signalling. Likewise, a BSR of an unknown user is useless until the eNodeB knows which UE transmitted the BSR. Thus, the C-RNTI MAC Control Element has higher priority than the BSR MAC Control Element.

Figure 4.20 illustrates the LTE MAC multiplexing by way of example. First, channel 1 is served up to its PBR, channel 2 up to its PBR and then channel 3 with as much data as is available (since in this example the amount of data available is less than would be permitted by the PBR configured for that channel). After that, the remaining space in the MAC PDU is filled with data from the channel 1 which is of the highest priority until there is no further room in the MAC PDU or there is no further data from channel 1. If there is still a room after serving the channel 1, channel 2 is served in a similar way.

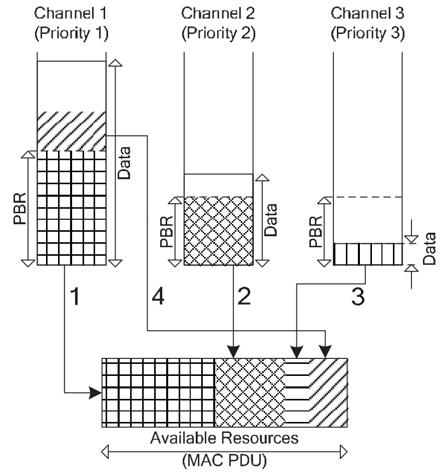

Figure 4.20: Example of MAC multiplexing.

4.4.2.7 MAC PDU Formats

When the multiplexing is done, the MAC PDU itself can be composed. The general MAC PDU format is shown in Figure 4.21. A MAC PDU primarily consists of the MAC header and the MAC payload. The MAC header is further composed of MAC subheaders, while the MAC payload is composed of MAC control elements, MAC SDUs and padding.

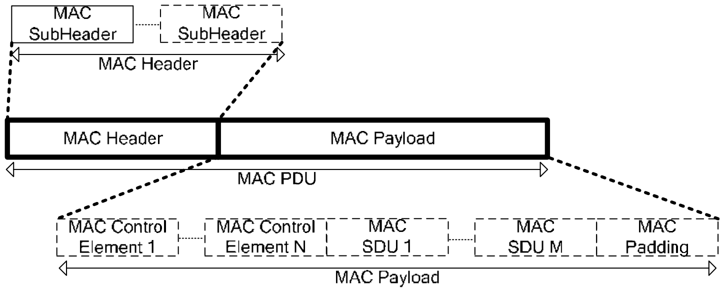

Figure 4.21: General MAC PDU format.

Each MAC subheader consists of a Logical Channel ID (LCID) and a Length (L) field. The LCID indicates whether the corresponding part of the MAC payload is a MAC Control Element, and if not, to which logical channel the related MAC SDU belongs. The L field indicates the size of the related MAC SDU or MAC Control Element.

MAC Control Elements are used for MAC-level peer-to-peer Signalling. The available types of MAC Control Element include the following:

- **Buffer Status Report MAC Control Element** for delivery of BSR information (see Section 4.4.2.2);
- **Power Headroom MAC Control Element** for the UE to report available power headroom (see Section 18.3.3);
- **DRX Command MAC Control Element** to transmit the downlink DRX commands to the UEs (see Section 4.4.2.5);
- **Timing Advance Command MAC Control Element** to transmit timing advance commands to the UEs for uplink timing alignment (see Sections 4.4.2.4 and 18.2.2);
- **C-RNTI MAC Control Element** for the UE to transmit its own C-RNTI during the random access procedure for the purpose of contention resolution (see Section 17.3.1);
- **UE Contention Resolution Identity MAC Control Element** for the eNodeB to transmit the uplink CCCH SDU that the UE has sent during the random access procedure for the purpose of contention resolution when the UE has no C-RNTI (see Section 17.3.1);
- **MBMS Dynamic Scheduling Information MAC Control Element** transmitted for each MCH to inform MBMS-capable UEs about scheduling of data transmissions on MTCH (see Section 13.6).

For each type of MAC Control Element, one special LCID is allocated.

When a MAC PDU is used to transport data from the PCCH or BCCH , the MAC PDU includes data from only one logical channel. In this case, because multiplexing is not applied, there is no need to include the LCID field in the header. In addition, if there is a one-to-one correspondence between a MAC SDU and a MAC PDU, the size of the MAC SDU can be known implicitly from the TB size. Thus, for these cases a headerless MAC PDU format is used as a transparent MAC PDU.

When a MAC PDU is used to transport the Random Access Response (RAR – see Section 17.3.1.2), a special MAC PDU format is applied with a MAC header and zero or more RARs. The MAC header consists of one or more MAC subheaders which include either a random access preamble identifier or a backoff indicator. Each MAC subheader including the Random Access Preamble IDentifier (RAPID) corresponds to one RAR in the MAC PDU (see Section 17.3.1).

4.5 Summary of the User Plane Protocols

The LTE Layer 2 protocol stack, consisting of the PDCP, RLC and MAC sublayers, acts as the interface between the radio access technology-agnostic sources of packet data traffic and the LTE physical layer. By providing functionality such as IP packet header compression, security, handover support, segmentation/concatenation, retransmission and reordering of packets, and transmission scheduling, the protocol stack enables the physical layer to be used efficiently for packet data traffic.

References[19]

[1] 3GPP Technical Specification 36.323, ‘Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) Specification’, www.3gpp.org.

[2] 3GPP Technical Specification 36.322, ‘Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) Protocol Specification’, www.3gpp.org.

[3] 3GPP Technical Specification 36.321, ‘Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification’, www.3gpp.org.

[19] All web sites confirmed 1st March 2011.