3GPP TSG RAN WG2#50 **R2-060125**
Sophia Antipolis, France Jan. 9-13, 2006

Agenda Item: **8.3**

Souce: **Motorola**

Title: **Native Packet Transmission**

Document for: **Discussion**

Introduction

Some of the primary goals of LTE are to provide for reduced latency and increased spectral efficiency. The targets for the latter are especially challenging – 3X over UTRA. It is therefore important to look at different techniques to improve the radio efficiency.

Discussion

Significant improvements have been made to Rel-99 for data applications with the adoption of HSDPA and HSUPA. Primarily the improvements result from the usage of channel condition based modulation and coding rate adaptation and HARQ. However, especially in the context of HSDPA, due to the need for supporting channel switching the RLC layer is still resident in the RNC. To accommodate the round trip delays between the RNC and Node B among other reasons this requires the use of sufficiently long sequence numbers for the RLC header. Furthermore, there is the hit from the MAC-d header. Thus in the end there is a significant impact on overall throughput due to protocol headers of the individual packets that are concatenated into one MAC-hs PDU. It would therefore be very beneficial to look at mechanisms that avoid the transport of expendable information such as headers of individual protocol layers.

To a significant extent this is accomplished through the movement of the RLC(upper ARQ) and MAC layers into the edge node. While the full extent of the RLC layer functionality is still FFS, one can hope that the overhead from this layer can be significantly reduced.

In case of HSDPA, a MAC-hs PDU consists of multiple MAC-d PDUs which are further comprised of RLC PDUs based on the segmentation length. Each of these has it's own header and the overall overhead for a single MAC-hs PDU can be significant. To get the full benefits of the shift of functionality to the Node B and eliminate and unnecessary overhead, it would be useful to be able to send a complete IP packet without fragmentation with (H)ARQ operating on the encoded IP packet-sized payload (it is assumed that the IP header has been compressed). This would also assist in achieving low latency through the exploitation of fat-pipe scheduling techniques wherein a user in a good channel condition can be provided the entire bandwidth of the channel for data transfer. Hence maximum spectral efficiency can be achieved through the combination of the following techniques:

1. Identifying users in good channel conditions
2. Link layer assisted adaptive modulation and coding and HARQ
3. Removal of all unnecessary user-plane protocol overhead – IP header, RLC, MAC, etc.
4. Terminating all ARQ mechanisms in the edge.

One potential mechanism to further strengthen #3 above is the encoding of native packets as opposed to encoding of (RLC) fragments of the native packets. This helps reduce the number of higher layer protocol headers embedded in the lower protocol layer packet while still ensuring end-to-end functionality.

It is well known that for advanced coding schemes such as turbo codes, LDPC codes, etc. the FER performance improves with the frame size (Fig. 1). This improvement in Es/No requirements translates to increased throughput accruing from the coding gain and removal of unnecessary overhead.

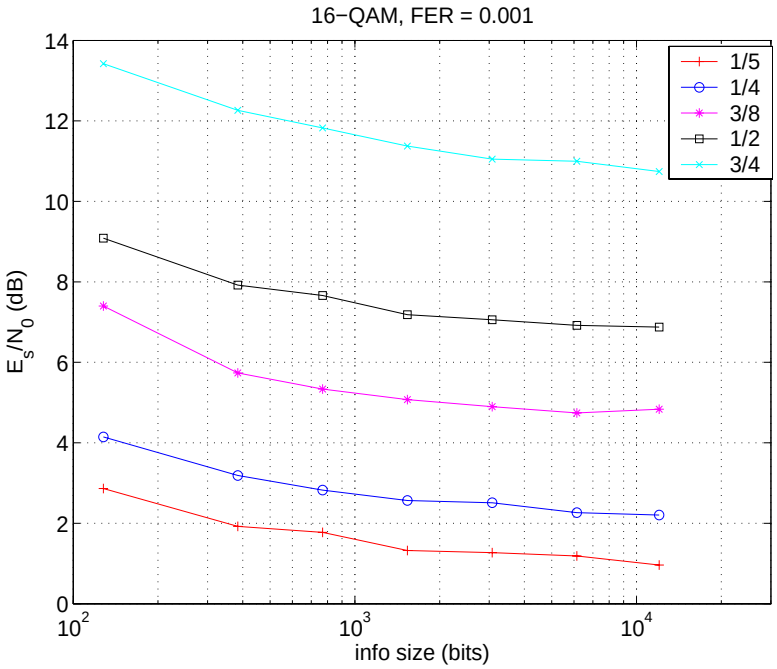

Fig. 1: Coding Gain vs. Packet Size

In LTE the elemental physical frame is of 0.5 ms duration. For the various combinations of modulation and coding rates the Table 1 below provides the maximum bit rate per frame that can be achieved. It assumes that a maximum of 5 symbols are available over 0.5 ms duration in case of 5MHz bandwidth. In case of uplink transmission a maximum of 6 symbols are available for transmission of data; note however that in the uplink 64-QAM would not be expected to be used. With MIMO the data rates that can be achieved are even higher.

Modulation	**Coding Rate**	**Information Bits/Frame**	**# Consecutive Sub-frames**
QPSK	1/5	600	20
	1/4	750	16
	1/2	1500	8
	3/4	2250	5.33
16QAM	1/2	3000	4
	3/4	4500	2.67
	5/6	5000	2.4
64QAM	1/2	4500	2.67
	3/4	6750	1.67
	5/6	7500	1.6

The last column indicates how many 0.5 ms sub-frames are needed to transmit a 1500 byte (typical MTU) IP packet. The point to be noted with the table is that it is indeed feasible to transmit a 1500 byte packet using between 2 and 3 sub-frames in areas of moderate to high geometry i.e. relatively good channel conditions. Note that Table 1 was generated for a 5 MHz bandwidth. For larger bandwidths, (e.g., 20 MHz), even some QPSK users can transmit a 1500 byte packet using a small number of frames. For most terminal speeds under consideration, the channel can be considered to be invariant over this period.

In the current Rel-6 scheme the following applies (again considering HSDPA as baseline for simplicity):

1. The MCS is chosen without taking into consideration upper layer (L3) packet boundaries. Thus an IP packet may be encoded over multiple sub-frames each with it's own CRC. Hence in the case of a single IP packet that is fragmented over multiple 2 ms HSDPA TTIs, there is a CRC associated with each 2ms transmission and there is an associated ACK/NACK for each transmission.

2. The re-transmissions are associated with the corresponding transmission as opposed to the packet as a whole. i.e. in the case of retransmission, the additional parity bits sent during retransmission are associated with the MAC fragment that was NACKed since each fragment is coded distinctly. The individual fragment transmissions are soft combined and after successful decoding are delivered to the upper layer for concatenation and regeneration of the original IP packet.

Native Packet Transmission

Figure 2 below illustrates the steps involved in the native packet transmission mechanism being proposed in this contribution.

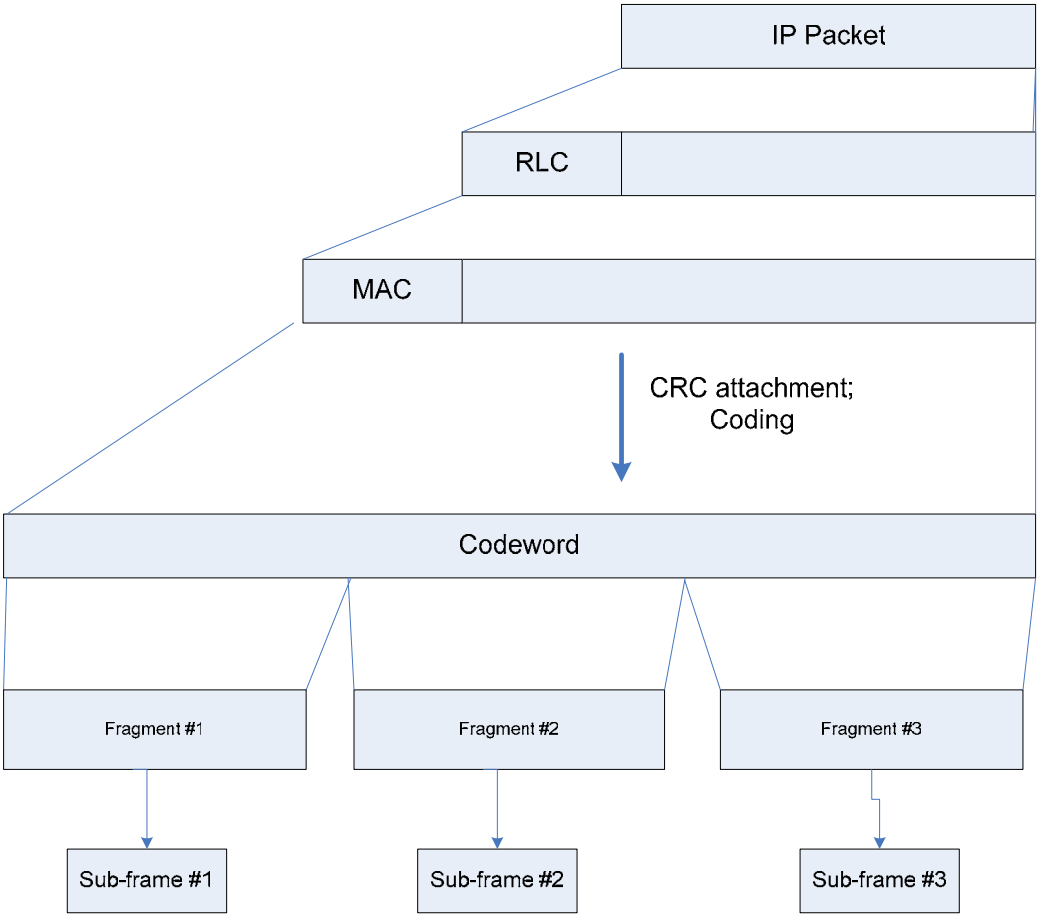

Figure 2: Native Packet Transmission

The incoming IP packet undergoes header compression and the output forms the payload for the RLC layer. However, in contrast to the current operation, RLC does not perform any fragmentation.. The MAC layer chooses a scheduling occasion for the user's packet and selects appropriate resources and modulation, coding rate and transport block size. The resource allocation may span multiple physical sub-frames in order to fit the initial transmission of the coded packet. The output of the MAC layer forms the transport block and is processed by L1 for transmission. Several IP packets can be aggregated at RLC/MAC especially in the case of small IP packets.

The modifications to the current transmission scheme are as follows:

1. There is a single CRC for the entire associated L3 (IP) packet; the entire L3 packet is transported as a single transport block.
2. After CRC attachment there is no transport block concatenation. The transport block with the CRC is input to the code block. The coded bits are then fed to a radio frame segmentation block that matches the number of bits to the bits assigned over a sequence of contiguous 0.5

ms sub-frames. Any fragmentation, if necessary, is performed on the coded packet at the physical layer.

3. The HARQ process is therefore associated with a TTI which potentially spans multiple physical sub-frames in contrast to the current scheme.
4. There is only one L1 ACK/NACK for the entire L3 packet.
5. Retransmissions on reception of a NACK are not associated with any particular fragment transmitted over a given sub-frame associated with this scheduling occasion but carry incremental coded bits associated with the entire packet and thus all fragments transmitted. It may be beneficial to be able to change the TTI for the retransmissions in order to ensure efficient usage of radio resources and reduce latency, especially in case of asynchronous HARQ schemes. This would permit faster rescheduling of retransmissions and new transmissions.

Due to the coding gain achieved through this scheme and the avoidance of fragmentation of the original packet, there is a reduction in overall latency. In addition, the link adaptation inherent in this mechanism permits the allocation of the saved resources to improve the performance for users in lower geometry conditions, e.g. cell-edge, and thereby increase the overall throughput and latency of the system.

The transmission scheme illustrated above can be further optimized by allowing non-contiguous transmission of the coded fragments in order to allow for transmission of higher priority packets. This may require definition of new frame formats providing for gaps in transmission for higher priority packets. However, it is also possible to support resource reservation while allowing for contiguous transmission through the use of extended TTIs.

Benefits of Native Packet Transmission

The native packet transmission scheme proposed here provides benefits in the following ways:

- It avoids unnecessary fragmentation of packets in upper L2 protocol layers due to L1 transmission constraints and consequentially also reduces L2 control signaling/feedback.
- The coding gain derived from the scheme leads to reduced latency due to optimal resource utilization and assists with further exploitation of the fat-pipe scheduling concept in good channel conditions.
- Better resource utilization for the good geometry and high bandwidth users permits better resource allocation for users in low geometry conditions thereby increasing cell-edge performance and overall throughput and hence spectral efficiency.

There is no impact to any of the Layer 2 functionality such as re-ordering, ciphering, etc.

From Table 1 above it can be seen that the scheme is practical in the case of high and moderate geometry users. For users at the cell edge also, the potential gain of using the native packet transmission scheme can be substantial. Note however that for smaller bandwidths, the transmission duration of a 1500 byte packet can be quite long.

Summary

Meeting the spectral efficiency targets that have been set for E-UTRA is a significant challenge. This requires the study and adoption of different mechanisms that can improve the throughput. By increasing the peak data rates achievable in moderate and high geometry regions, it is possible to increase the overall spectral efficiency of EUTRA and achieve reduction in latency. The proposal here provides an alternate mechanism for transmission of data that avoids un-necessary L2 fragmentation that would have been otherwise needed, to match the resources available in a sub-frame, through re-definition of the transmission time interval as a dynamic attribute of the transport channel. The TTI may also be allowed to vary for a scheduling occasion due to the different requirements of resources for retransmissions compared to the initial transmissions.

It is proposed that the transmission scheme proposed here be considered for adoption by supporting the modifications to the transport channel attributes and HARQ protocol as outlined above.