ETSI SMG2 UMTS L2/3 Tdoc TSGR2#2(99) 127
Stockholm
8-11 March 1999
Agenda item: 6.2
Source: Motorola

Comparison of RRC and MAC based methods of dynamic scheduling between users on the uplink

1 Introduction

To date two methods have been proposed for scheduling access to the uplink resource:

1. SRNC based RRC scheduling on DCH's

2. CRNC based MAC scheduling on the Uplink Shared CHannel (USCH).

It is the view of Motorola that the RRC/DCH approach looks more like 'fast circuit' as opposed to a proper packet data solution. In this paper the motivations for performing scheduling between users in the CRNC on the MAC layer are provided.

The key points are as follows:

1. Fast radio resource allocation and de-allocation is necessary in order to exploit the advantages of the 'fat shared pipe'.

2. Messaging overhead is less if this rapid resource allocation is performed on the lowest layer possible - ie the MAC layer.

3. The RRC 'TFC Control' message must be sent in Acknowledged mode if shared access to a fat pipe is to be achieved. With the proposed MAC resource allocation this is not the case, consequently the delay in accessing the resource is decreased and spectral efficiency and QoS are improved.

4. With CRNC/MAC scheduling communication is always between the UE and the RNC which controls access to the resource required (ie. the CRNC). Since the CRNC is sometimes the DRNC the delay is reduced and QoS improved over the SRNC/RRC/DCH alternative.

2 Contents

3 Advantages of rapid resource allocation

Maximum statistical multiplexing gains are achieved when users are multiplexed onto a shared fat data pipe [1]. It is therefore important to ensure that it is possible to rapidly allocate and de-allocate capacity between different packet data users. For the system to be operated efficiently the capability should be provided to do this on a frame by frame basis.

The advantages of rapid allocation and re-allocation of capacity onto the shared channels using the MAC layer as opposed to using the alternative of RRC scheduling on DCH's were listed in [2] and were as follows:

Spectral efficiency and packet call completion times are improved

There is a well known traffic theory result which states that it is more efficient to have one high rate server (single fat data pipe or more rigorously M/M/1) as opposed to multiple low rate servers (thin pipes or M/M/n). The RRC/DCH approach has NOT been designed with fast resource allocation in mind and therefore implicitly assumes a multiple thin pipe (circuit like) scenario.

Spectral efficiency and packet call completion times are much less dependent on inaccurate and unpredictable traffic generation parameters supplied at packet call set-up

With the RRC/DCH approach there is a concern about how the data rate of the DCH is to be chosen. How does the RRC know what rate DCH to assign given that very little information is available at the start of a packet call ? (the first thing that RRC knows about the packet call is when packets start to arrive from the higher layer). If RRC assigns a DCH rate which is too large then capacity will go wasted. If it assigns a rate which is too small then packet call completion times will be unnecessarily increased. By only allowing 'short leases' on the radio resources (ie. a frame at a time) as is the case with the shared channels then the system is very responsive to changes in conditions and no assumptions have to be made about unpredictable traffic characteristics.

Highest priority packets are always served first

In deciding which packet to send next the shared channel MAC can consider the priority of ALL packets queued in all UE's/ destined to all UE's, and in each frame will always serve the highest priority packet first. Therefore the situation never arises where high priority packet calls are queued, waiting for DCH's carrying lower priority data to clear down.

Capacity is not wasted whilst DCH's are being released and re-assigned

Likewise, with the RRC/ DCH approach the situation will arise whereby capacity is wasted whilst the RRC waits to see if any more packets are going to arrive before de-allocating the DCH. During this period whilst the RRC is waiting for its timers to expire, radio resource is wasted. In the shared channel approach, the capacity is fully packed in every frame (assuming of course that there is some data to send).

More efficient capacity packing - Better spectrum efficiency

The size of the shared channel can be dynamically modified to exploit even very rapid changes in conditions, ie which might occur on a frame by frame basis. The shared channel can thereby soak up excess radio resource in a way that is not possible with the DCH approach.

4 Advantages of scheduling on MAC as opposed to RRC

Light weight L2 protocol

By only granting short leases of the radio resource to packet users it is possible to be more responsive to changing traffic patterns. With the shared channel concept, allocations just last for one frame and this means that it is possible to exploit a light-weight Layer 2 protocol. Since the allocation messages will be sent frequently, it is important that the messages are as compact and as efficient as possible (in order that the signalling overhead is minimised). In our view this means that the scheduling and the allocation messages should be sent at the lowest layer possible (the MAC layer). If the messages are sent from the RRC layer then the protocol overhead will be greater.

Problems with RRC Transport Format Combination Control procedure

If RRC does not signal allocation messages regularly eg. in every frame, and if allocations do not last for a pre-defined interval (neither of which is currently proposed) then we believe that in order to support shared access to a fat pipe then the RRC Transport Format Combination Control message would have to be sent in acknowledged mode. Acknowledged mode would be needed since the percentage of the capacity allocated to a user will be quite high, it is therefore very important that the network knows that the user has received the TFC Control message (before the network can re-assign the capacity to another user). A missed TFC Control message could prove disastrous and destroy all uplink communication in the cell. The use of acknowledgements on the RRC layer will reduce the speed with which scheduling and resource re-assignment can be achieved.

When frame by frame messaging on the MAC layer is used then this problem does not arise, the network never sends resource release messages and only allocation messages are sent. If the UE does not receive an allocation message then the only effect is that the allocated resource goes unused for a frame. The performance of other UE's on the system is not endangered.

Less scope for taking pre-defined actions in the event that a message is received in error

Another problem with RRC scheduling stems from the fact that there are many different types of message which can be sent by RRC. With the MAC/CRNC scheme the UE / network knows implicitly, from the channel that it is listening to, that the message is a MAC message, if it doesn't receive the message correctly then it can take pre-defined actions. For example, if a message assigning uplink resource allocation is received in error then the UE can assume the worst case, ie. that no assignment was made. However, when an RRC message is received incorrectly the UE /network could not take a pre-defined action because it doesn't know the message type a priori.

RRC terminates in SRNC, MAC-sh terminates wherever the channel terminates (which can be the SRNC OR the DRNC)

In the current model the RRC always terminates in the SRNC (the RNC which has the Iu connection to the core network for the UE in question). The choice of the SRNC as a termination point for all resource requests, therefore seems somewhat arbitrary from a resource management point of view and is not always the most efficient alternative. Generally, it is more efficient to facilitate direct communication of resource requests between the UE and the CRNC to which the resource request is directed (which it can be noted might be located in the DRNC). This point is discussed further in the next section.

5 Advantages of scheduling in CRNC as opposed to SRNC

There will sometimes be unnecessary extra delay incurred with RRC scheduling at the SRNC as opposed to the alternative where the RNC which controls the resource also makes the decision as to whether the resource can be allocated (ie MAC scheduling at the CRNC).

In this section we consider an example scenario which demonstrates this point for the more interesting case where there are two RNC's involved (an SRNC and a DRNC). We consider the scenario of a UE which carries downlink circuit data (at 32 kbit/s) plus uplink packet data (at 64 kbit/s).

A UE might be operating with this kind of configuration when using, for example, a video on demand service. Downlink video would be streamed at a constant rate, uplink commands (fast forward, freeze frame) etc. would be carried in packet mode and preferably at the highest data rate possible in order to minimise the delay in system response.

We consider the cases where:

1. Downlink circuit data is carried on a DCH and uplink packet data is carried on a code multiplexed USCH.

2. Downlink and uplink data are carried on separate code multiplexed DCH's (Each DCH has its own individual bi-directional DPCCH's carrying power control bits etc. on the return path).

This example assumes code multiplexing, about which there has been little discussion. In order to demonstrate that we are considering a realistic example it is worth making a few comments on the chosen example and the code multiplexing assumption, this is provided in Appendix A.

5.1 Requirement to support different active sets for uplink/downlink or for different services

One can envisage a number of scenario's where the uplink and downlink active sets are different or where a UE carries multiple services via code multiplexing, each service having a different active set. By considering these cases one of the short-comings with the RRC based scheduling approach is high-lighted.

Figure 1 shows a scenario where the DCH is operated in SHO, but the USCH is operated in HHO. This situation might result where the uplink and downlink coverage areas are (temporarily[1]) different. In the example considered, we assume that there is much more uplink interference at BS#1 than at BS#2. This may result for any number of reasons (different inter and intra-cellular interference levels, different adjacent channel interference, different RACH loading etc etc.). This means that at the given point in time the uplink 64 kbit/s coverage region is less at BS#1 than at BS#2. In the downlink, we assume that at the given point in time the propagation loss is the same from both BS's and that downlink power from each BS is the same. Thus the downlink DCH should be operated in SHO (both BS's are equally good). However, it would not be worth operating the USCH in SHO since network resources (BS receivers, land-line capacity) would be consumed for no benefit. It is assumed that the magnitude of the received signal at BS#1 is small enough such that it is not necessary to schedule the transmission in BS#1.

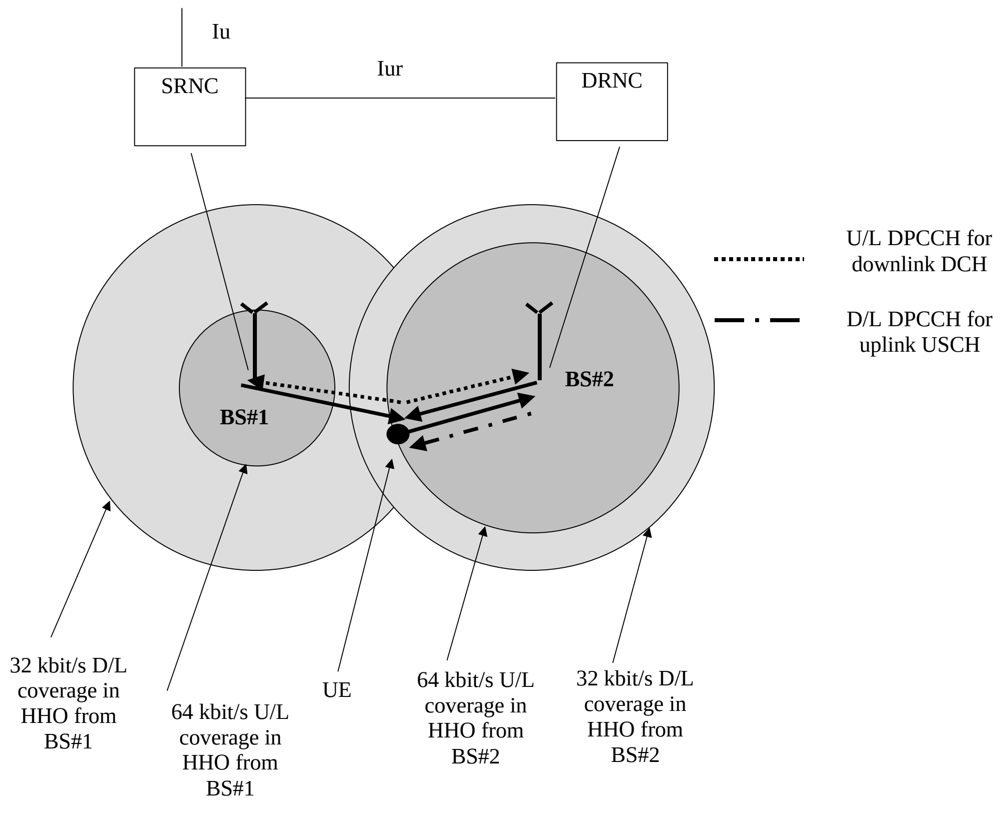

Figure 1) Example hand-off scenario for a UE which uses a DCH for conveyance of D/L circuit data and a USCH for conveyance of uplink packets

[1] Note that at any instant in time the coverage areas may be different to those planned by the operator due to prevailing (dynamic) changes in conditions.

Similar situations can be envisaged when a UE carries a voice service on a bi-directional DCH and an uplink data service on another code (DCH or USCH).

5.2 Signalling flow: Downlink circuit data on DCH and uplink packet data on DCH

Given this example scenario (as shown in Figure 1), Figure 2 shows the resource request procedure where the uplink packet data is carried on a DCH. Although the resource is only required from the DRNC the messages have to be sent to the SRNC where the RRC protocol terminates.

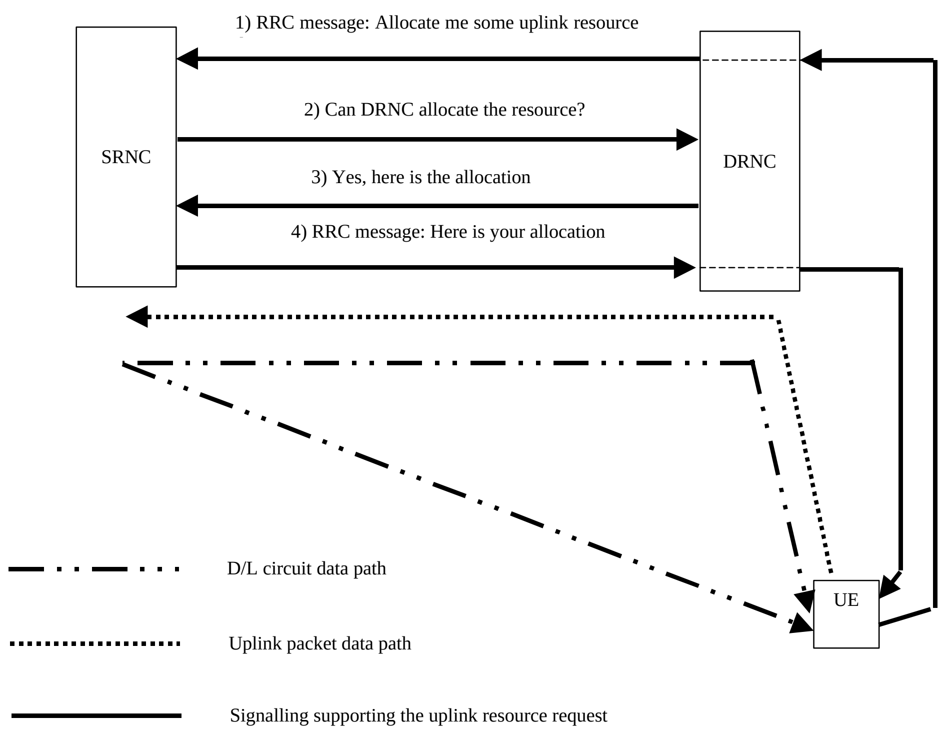

Figure 2) Signalling exchange when UE requests uplink resource for transmission of information on an uplink DCH

<u>Observations:</u>

1. Note the large number of signalling flows across the Iur which will contribute significantly to the delay before the UE is assigned uplink resource for conveyance of the packet data.

2. Note also that this is a best case scenario, delays could be even worse if:
   a) The DRNC could not have accommodated the resource request, since in this case there would have been further negotiation between SRNC and DRNC across the Iur.
   b) The messages had been sent in acknowledged mode.

5.3 Signalling flow: Downlink circuit data on a DCH and uplink packet data on USCH

In Figure 3 the UE requests uplink resource for the data connection on the USCH. This time the UE communicates directly with the relevant CRNC. No communication is required across the Iur interface.

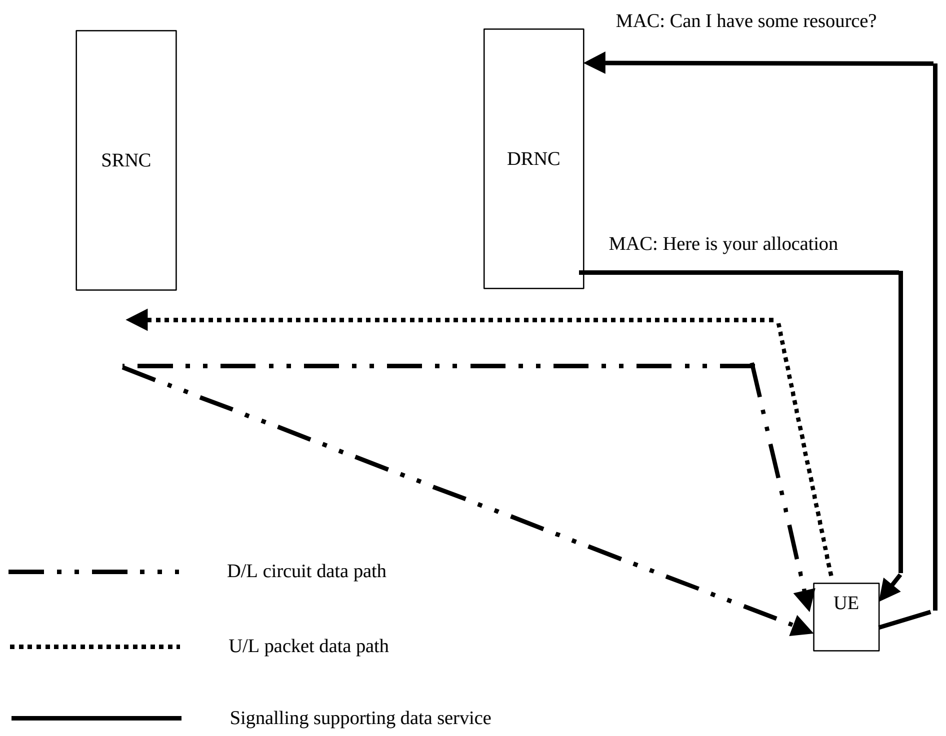

Figure 3) Signalling exchange when UE requests uplink resource on the USCH

Observation

1. There is no signalling exchange across the Iur. The UE can receive its allocation with minimal delay.

5.4 Deficiencies of the current uplink resource assignment model

Figure 1, Figure 2 and Figure 3 demonstrate an important deficiency with the currently proposed models for requesting and assigning resource.

In the existing model there is only one RRC connection per UE and this is terminated in the SRNC (the RNC in which the Iu connection for the UE is terminated). This is a problem because currently the only way of requesting uplink resource is via the RRC protocol. This means that even when the DRNC is in complete control of the resource required (for the uplink packet data transmission in the example considered) the UE is still forced to communicate with the SRNC. This greatly contributes to the delay experienced without achieving any benefits.

From Figure 2 it can be seen that this will be a problem (independent of the USCH concept), ie. it will potentially be a problem whenever two DCH's are conveyed by code multiplexing.

5.5 MAC/CRNC solution to the problem

Clearly, it is more efficient if the UE is able to communicate directly with the RNC which controls access to the resource which it requires. Therefore when the UE has to make an uplink data transmission it should communicate with the relevant CRNC, irrespective of whether this RNC is the serving or drift RNC.

With the USCH proposal this is exactly what is achieved.

6 Management of the radio resource for UE's situated on the edge between two RNC cell-sets

UE's situated on the edge between two cells will sometimes be operated in SHO and hence require resource from both cells. If these cells are controlled by different RNC's then in order to schedule an uplink transmission from the UE it may be necessary for the RNC's to negotiate with each other.

In the following sections we consider a number of ways in which this might be done, firstly, for the case of SRNC/RRC scheduling onto the DCH and then for the case of CRNC/MAC scheduling onto the USCH.

It is concluded that it can be useful to provide the facility for MAC_sh entities to communicate with one another across the Iur.

In order to elaborate on the issues we consider the case where the only service requirement of the UE is to convey uplink packet data and we make the assumption that it is necessary to convey the uplink packet data in SHO. We consider two scenario's:

1. The data is conveyed on a DCH using SRNC/RRC scheduling

2. The data is conveyed on the USCH using CRNC/MAC scheduling

6.1 SRNC/RRC scheduling onto a DCH

6.1.1 First: Some definitions

It proves useful if we break down the operation of the RNC into a number of different functions:

1. UE specific - Functions and entities in the RNC related to one UE.
2. Cell specific - Functions and entities in the RNC related to the management of resources in one radio cell.
3. Channel specific - Functions and entities in the RNC related to one particular channel within one particular cell.

6.1.2 Inter-RNC resource arbitration solution #1

This solution is shown in Figure 4 in which the UE-specific RRC entity acts as a centralised 'resource arbitrator'. In Figure 5 the related signalling process is shown (this defines the steps which are refered to in Figure 4).

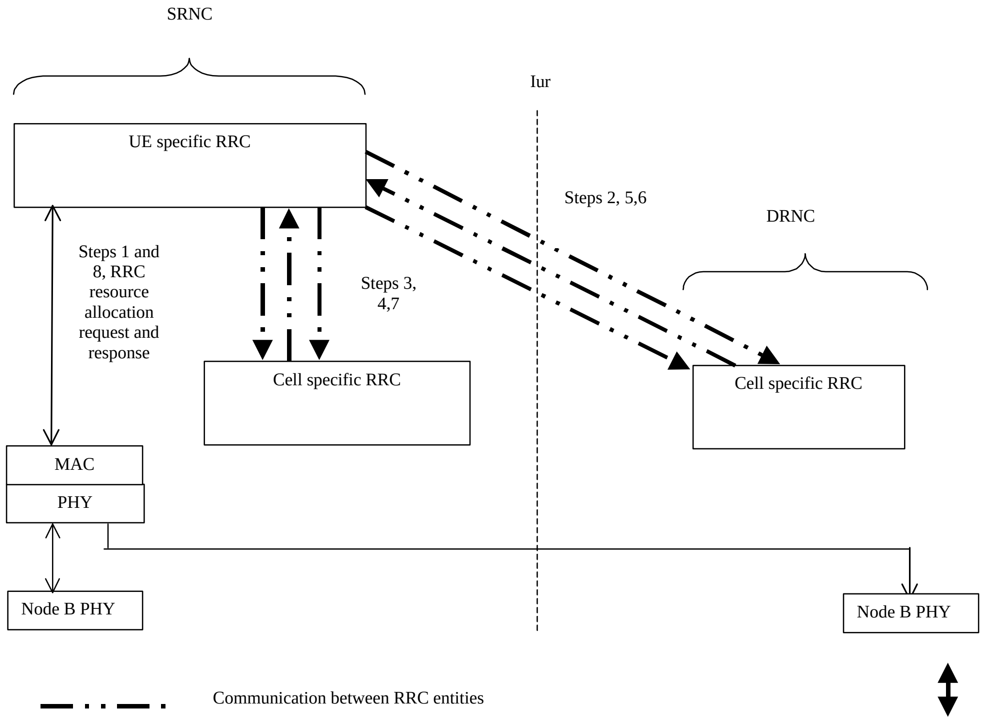

Figure 4) Inter-RNC resource arbitration solution #1

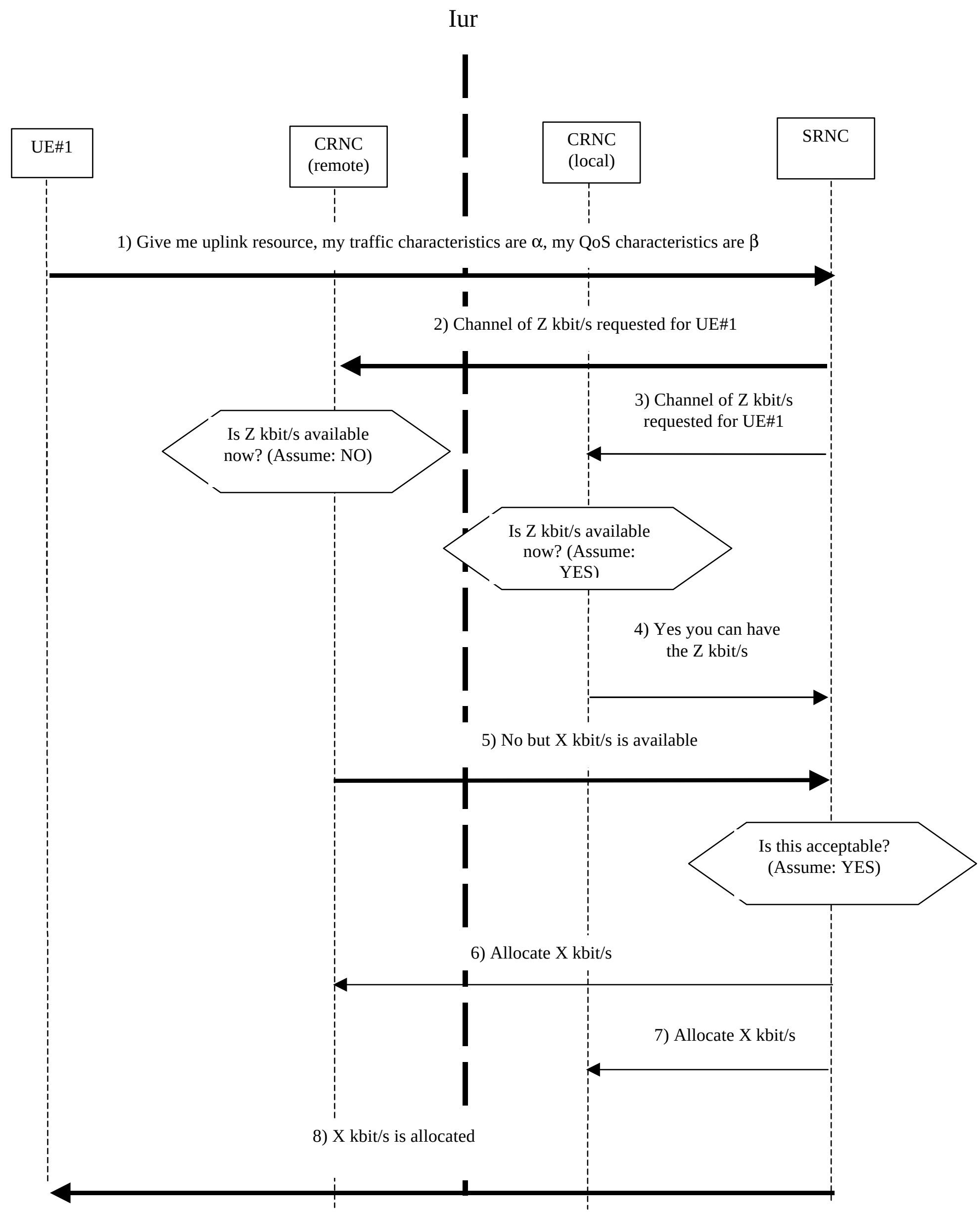

Figure 5) Example of signalling sequence for RRC/DCH uplink scheduling where SRNC acts as a 'resource arbitrator'

We can make some observations about this solution:

1. This is really the best case scenario, yet still at least 4 messages have to go across the Iur which are in the critical delay path (shown in bold).

2. In the event that X kbit/s is not acceptable it is unclear how the procedure would proceed. Certainly, further signalling exchanges across the Iur would be necessary.

3. Another possibility, is that when the resource request arrives at the CRNC's the resource might not currently be available but it is possible that the new resource request could be of higher priority than the transmissions of users which are currently being served. In this instance, it would be necessary for the CRNC's to force the (low priority) UE's which are currently being served to release their resource. This would have to be done in acknowledged mode and this would again add to the delay of resource re-assignment.

6.1.3 Inter-RNC resource arbitration solution #2

In Figure 6 an alternative solution is shown. This time the responsibility for arbitrating between the RNC's is delegated to the cell specific RRC (CRNC) entity in the SRNC. The advantage of this is that the Cell specific RRC entity has the knowledge of exactly what bandwidth is available and also has knowledge of all the resource requests and associated priorities for all UE's on the network. It can be assumed that the cell specific RRC would therefore be in a better position to perform a more efficient arbitration than the UE specific RRC.

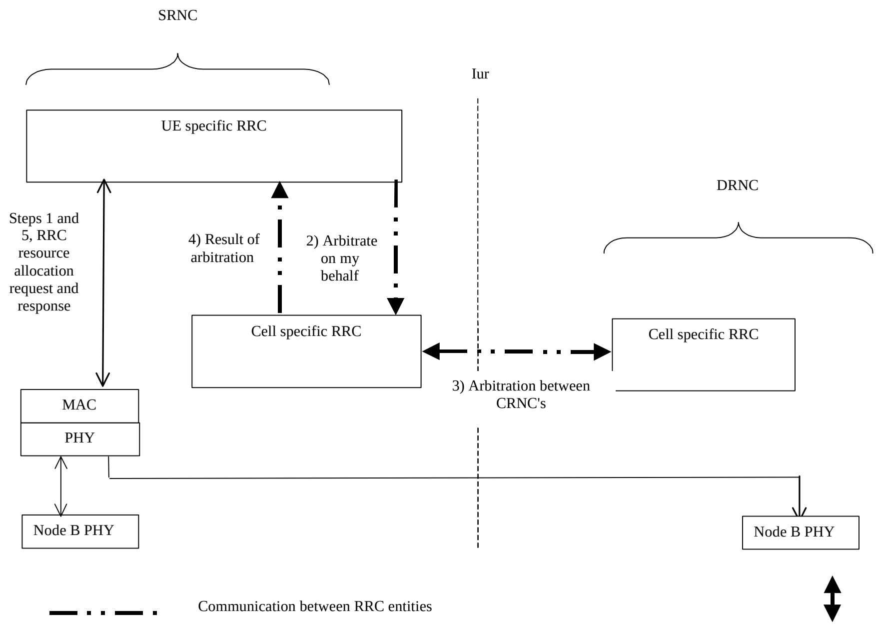

Figure 6) Inter-RNC resource arbitration solution #2

6.2 CRNC/MAC scheduling onto the USCH

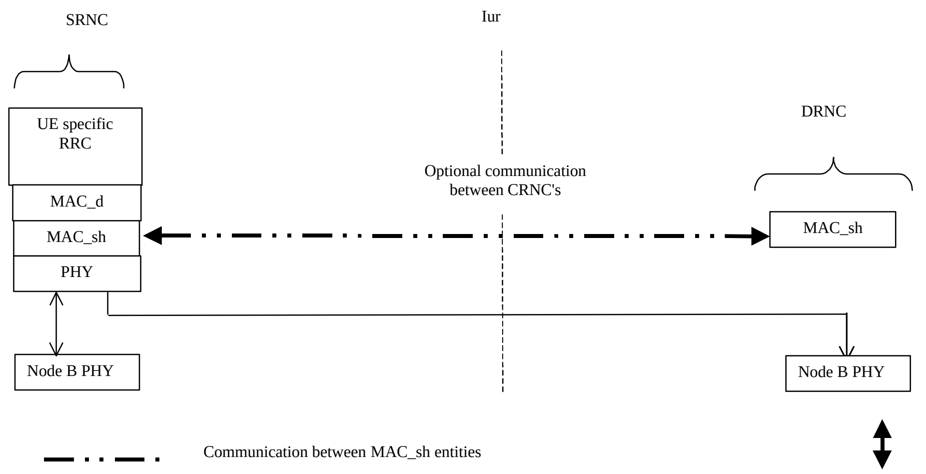

Figure 7) CRNC/MAC_sh based arbitration between RNC's

Figure 7 shows the method for managing resource when using CRNC/MAC_sh scheduling. It is assumed that each UE selects one MAC_sh entity (in either the SRNC or the DRNC) as being its 'scheduling controller'. A very light-weight signalling exchange across the Iur may be used between CRNC MAC_sh scheduling entities to improve the efficiency of the SHO scheduling.

The option should be provided for sending resource allocations/resource requests from/to only the scheduling controller CRNC in order to minimise further signalling across the Iur (ie. MAC messages can be sent/read before splitting/combining is done).

Further discussion on the various signalling options is provided in a companion paper [3].

We recommend that UMTS provide the capability for communication between MAC_sh entities over the Iur. This is in line with the conclusion drawn from Figure 6 in which it was shown that it will be more efficient to enable the cell-specific RRC (ie the CRNC's) to communicate with one another rather than to enable the SRNC-RRC to be the 'resource arbitrator' (Figure 4).

7 Conclusion

1. Fast radio resource allocation and de-allocation is necessary in order to exploit the advantages of the 'fat shared pipe'.

2. Messaging overhead is less if this rapid resource allocation is performed on the lowest layer possible - ie the MAC layer.

3. The RRC 'TFC Control' message must be sent in Acknowledged mode if shared access to a fat pipe is to be achieved. With the proposed MAC resource allocation this is not the case, consequently the delay in accessing the resource is decreased and spectral efficiency and QoS are improved.

4. There is less scope for taking pre-defined actions in the event that an RRC message is received in error compared to the case where a MAC message is received in error. This will impact the efficiency of radio resource management.

5. With CRNC/MAC scheduling, communication is always between the UE and the RNC which controls access to the resource required (ie. the CRNC). Since the CRNC is sometimes the DRNC the delay is reduced and QoS improved over the SRNC/RRC/DCH alternative.

6. UMTS should provide the capability for communication between MAC_sh entities over the Iur.

8 References

[1] Motorola, "Channel Bandwidth Allocation Strategy", SMG2 UMTS L23 534/98

[2] Motorola, "Benefits of the Uplink Shared Channel," 3GPP RAN WG2 Tdoc 032/99

[3] Motorola, 'Operating the USCH in Soft Hand-Off', 3GPP RAN WG2 Tdoc TSGR2#2(99) 126

9 Appendix A - Code multiplexing

In general, there are a number of reasons why it can be beneficial to use code multiplexing:

1. Code multiplexing is often a requirement when the base site framing is not synchronised and multiple services are being carried to/from one UE. Consider the case of a UE engaged in a voice call. It's initial framing will be defined by that of the cell to which it initially gains access. If the UE travels into a new cell then in order to prevent the loss of any speech frames, the framing used

will have to be the same as that of the originating cell (and hence different to that of the new cell). Now let's imagine that the UE wants to set up a HHO connection on the USCH in the new cell. In order to maximise efficient use of the shared uplink resource then allocations on the USCH should be synchronised to a common framing (eg the ACCH). Since the framing of the USCH and the DCH may be different, it is necessary to use code multiplexing (if DCH frames are not to be lost).

A similar situation could occur with the example of the video on demand service considered in this section.

2. The curves of Eb/No vs BER for different services will change differently as conditions change. It proves more efficient to have different outer loop targets for the different services, this can only be achieved with code multiplexing.

   Equally, in terms of the video on demand example considered herein, the BER target of the downlink video may well be different to that of the uplink packet data. It can be useful to set separate outer loop targets for each.

3. Different services have different requirements in terms of hand-off operation. For example, for a particular UE at a particular point in time it will sometimes be more efficient to operate a voice service in SHO and the data service in HHO. This is easily achieved with code multiplexing.

   In terms of the video on demand example, under some conditions it may be more efficient to operate the downlink video circuit in SHO with the high speed uplink packet transmission operated in HHO (this is discussed in more detail in Section 5.1). Again, code multiplexing can prove useful in facilitating this.

4. We have proposed that the uplink packet transmission should be made at a high rate (to facilitate fat pipe multiplexing). If the data rate of the uplink packet connection is much higher than the data rate of any other DCH's (eg. a voice DCH, or U/L DPCCH (in the case of the example considered)) then the UE peak to average transmit power of the aggregated services will be dominated by that of the data connection. Hence the impact on UE transmit linearity requirements need not be so severe. [However, it is worth noting that in some circumstances code multiplexing can lead to linearity problems in a UE's PA and therefore is not always to be recommended].